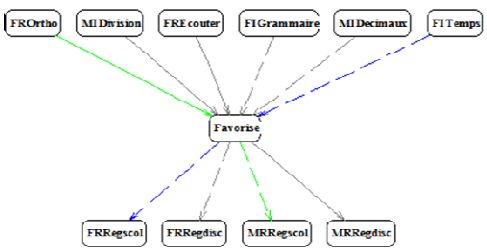
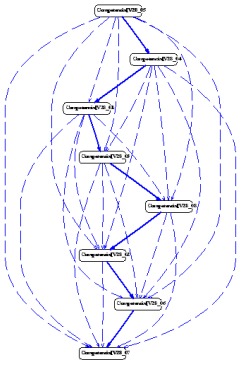
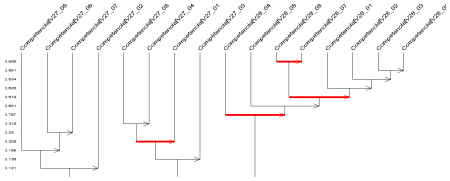






Analyse Statistique Implicative.

Des sciences dures aux sciences humaines et sociales

Statistical Implicative Analysis

From the exact sciences to human and social sciences

Editeurs: Jean-Claude Régnier, Yahia Slimani, Régis Gras, & Association ARSA

REMERCIEMENTS

Nous tenons à remercier les différents organismes qui nous ont apporté un soutien pour la tenue de ces Journées sur les plans **scientifique, financier et logistique** :

- ISET – Institut Supérieur des Études Technologiques de Radès
- ARSA – Association de Recherche en Statistique Appliquée
- Université Lumière Lyon2 (Lyon France)
- Laboratoire de recherche : Unité Mixte de Recherche UMR 5191 – ICAR (Lyon France)
- Association pour la Recherche en Didactique des Mathématiques (ARDM France),

ou leur parrainage scientifique *:*

- Association Extraction et Gestion des Connaissances (EGC) (France),
- École Polytechnique de l'Université de Nantes (France),
- Laboratoire d'Informatique de Nantes (LINA France),
- Société française de Statistique (SFdS France),
- Société Francophone de Classification (SFC France),
- G.R.I.M. (Gruppo di Ricerca sull'Insegnamento delle Matematiche), Dipartimento di Matematica, Università di Palermo (Italie)
- Revue des Sciences de l'éducation (Montréal Québec Canada)

Nous tenons à remercier tous les membres du comité d'organisation dont l'implication logistique a rendu possible la réalisation de ce 8ème colloque sur l'Analyse Statistique Implicative.

Président du comité scientifique : Jean-Claude Régnier

Vice-Président du comité scientifique : Yahya Slimani

Président d'honneur : Régis Gras

Président du comité d'organisation: Ahmed Dhouibi

Vice-Président du comité d'organisation: Makram Ben Jeddou

INTRODUCTION DE L'OUVRAGE

L'ANALYSE STATISTIQUE IMPLICATIVE : DES SCIENCES DURES AUX SCIENCES HUMAINES ET SOCIALES

Jean-Claude REGNIER[1], Régis GRAS[2]

Le huitième colloque A.S.I. 8 s'est déroulé dans un contexte d'espérance stimulée pour la démocratie tunisienne avec l'attribution du Prix Nobel de la Paix pour 2015. Comme on a pu le lire dans les journaux :

« Le comité Nobel norvégien a décidé de récompenser, vendredi 9 octobre 2015, le quartet menant le dialogue national en Tunisie, qui s'est distingué pour « *sa contribution décisive dans la construction d'une démocratie pluraliste en Tunisie après la "révolution du jasmin" de 2011* » ». (Le Monde[3])

Un tel contexte est de toute évidence plus propice à la créativité scientifique, à l'exercice de la pensée critique qui fonde les dimensions épistémologique et méthodologique des champs scientifiques. Évidemment, nous plaçons le champ de l'Analyse Statistique Implicative dans ces perspectives scientifiques.

A.S.I. – Analyse statistique implicative : une fois encore et encore, de quoi s'agit-il ?

Au risque de nous répéter, nous définissons l'Analyse Statistique Implicative comme un cadre théorique d'analyse de données fondée sur une relation non symétrique. Il s'agit d'« … un champ théorique centré sur le concept d'implication statistique ou plus précisément sur le concept de quasi-implication pour le distinguer de celui d'implication logique des domaines de la logique et des mathématiques. L'étude de ce concept de quasi-implication en tant qu'objet mathématique, dans les champs des probabilités et de la statistique, a permis de construire des outils théoriques qui instrumentent une méthode d'analyse de données. »[4] (Gras, Régnier, 2009 p.12).

Le présent ouvrage est constitué des articles issus d'un appel à contributions lancé dans le cadre du huitième colloque sur l'Analyse Statistique Implicative – ASI8 - organisé à Radès en novembre 2015. Ces articles ont été soumis à la lecture critique[5] des membres du comité international scientifique qui en ont assuré la qualité scientifique. Nous rapportons ici la liste en leur adressant nos remerciements les plus chaleureux pour leur travail.

- Nadja **Acioly-Régnier**, EAM 4128 SIS - Université Lyon 1 (France)

---

[1] Président du comité scientifique de ASI8 - UMR 5191 ICAR, Université de Lyon - Lyon 2 (France)

[2] Président d'honneur du comité scientifique de ASI8 - LINA, Université de Nantes (France)

[3] http://www.lemonde.fr/prix-nobel/article/2015/10/09/le-prix-nobel-de-la-paix-attribue-au-dialogue-national-tunisien_4786205_1772031.html

[4] Gras R., Régnier J.-C., Guillet F. (Eds) (2009) *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. RNTI-E-16 Toulouse Cépaduès Editions

[5] Chaque contribution a été anonymement soumise à trois relecteurs dont les expertises ont été retournées anonymement aux auteurs.

- Saddo **Ag Almouloud**, PPG Educação Matematica - Pontifícia universidade Católica de São Paulo (Brésil)
- Marc **Bailleul**, EA 965 CERSE, Université de Caen (France)
- Makram **Ben Jeddou**, ISET de Radès, Association ARSA et Université virtuelle UVT (Tunisie)
- Sadok **Ben Yahia**, Département des Sciences de l'Informatique, Faculté des Sciences de Tunis (Tunisie)
- Younes **Boujelbene** Faculté des Sciences économiques et de gestion de Sfax (Tunisie)
- Guy **Brousseau**, DAEST - Université Bordeaux 2 (France)
- Raphaël **Couturier**, FEMTO-ST département DISC Université de Franche-Comté (France)
- Benedetto **Di Paola**, Département de mathématiques - Université di Palermo (Italie)
- Iliada **Elia**, Département de Sciences de l'Éducation - University of Cyprus (Chypre)
- Athanasios **Gagatsis**, Département de Sciences de l'Éducation - University of Cyprus (Chypre)
- Pablo **Gregori**, IMAC - Universitat Jaume I de Castellón de la Plana (Espagne)
- Fabrice **Guillet**, LINA CNRS 6241, Equipe COD - École Polytechnique de l'Université de Nantes (France)
- Pascale **Kuntz**, LINA CNRS 6241, Equipe COD - École Polytechnique de l'Université de Nantes (France)
- Dominique **Lahanier-Reuter**, Equipe Théodile CIREL - Université Lille 3 (France)
- Chiraz **Latiri**, Université de la Manouba (Tunisie)
- Philippe **Lenca**, DECIDE CID UMR 6285 Lab-STICC - Telecom Bretagne - (France)
- Dhafer **Malouche**, ESSAI - Ecole Supérieure de la Statistique et de l'Analyse de l'Information - Tunis - (Tunisie)
- Paraskevi **Michael-Chrysanthou**, Department of Education University of Cyprus (Chypre)
- Abdelmajid **Naceur** Université virtuelle de Tunis, (Tunisie)
- Maria del Pilar **Orus Baguena**, IMAC - Universitat Jaume I de Castellón de la Plana (Espagne)
- François **Pluvinage**, Depto. Matemática Educativa, CINVESTAV-IPN México and France
- Gilbert **Ritschard**, Institut d'études démographiques et des parcours de vie - Pôle de recherche national LIVES - Université de Genève (Suisse)
- Gilbert **Saporta**, Laboratoire CEDRIC - CNAM, Paris (France)
- Gérard **Vergnaud**, CNRS, Université Paris VIII (France)
- Miguel R.**Wilhelmi**, Departamento de matemáticas, Universidad Pública de Navarra (Espagne)
- Djamel Abdelkader **Zighed**, Laboratoire ERIC - Université Lyon2 (France)

A.S.I. – Analyse statistique implicative : pour mémoire, quelques-uns de ses lieux de débats et de construction.

Pour mémoire, le premier colloque ASI1 qui s'était tenu à l'Institut de Formation des Maîtres de Caen (France) les 23-24 Juin 2000 à Caen et avait été organisé par Marc Bailleul et Régis Gras, avait une thématique générale centrée sur « La fouille dans les données par la Méthode Statistique Implicative ». Ce choix correspondait à une orientation prometteuse quand on considère le développement et les activités de l'association EGC (**E**xtraction et **G**estion des **C**onnaissances). Rappelons que la « fouille dans des données » (encore appelée **K**nowledge **D**iscovery in **D**atabases ou encore "Data Mining" dans la littérature anglo-saxonne) part, en général, du croisement de sujets (ou objets) et de variables (propriétés ou attributs) binaires, ordinales ou numériques. Son objectif majeur consiste à conjecturer des modèles basés sur des relations quantitatives ou qualitatives et des structures induites à partir des données. Différentes méthodes, comme l'Analyse Factorielle des Correspondances (A.F.C.), la Classification Ascendante Hiérarchique (C.A.H.), sont communément utilisées pour de telles fouilles dans des données. Parmi elles, l'Analyse Statistique Implicative (A.S.I.) vise l'extraction de connaissances, d'invariants, de règles inductives non symétriques consistantes, et accorde une mesure à des propositions du type « quand a est choisi, on a tendance à choisir b ».

Le deuxième colloque ASI2 a été organisé à l'Université PUC de São Paulo (Brésil) les 9-11 Juillet 2003, par Saddo Ag Almouloud.sur la thématique globale « O metodo estatístico implicativo utilizado em estudos qualitativos de Régis Gras de associação. Contribuição à pesquisa em Educação »

Le troisième colloque ASI3 a été organisé par Filippo Spagnolo de l'Université de Palerme à Palerme (Italie) les 6-8 octobre 2005.

Le quatrième colloque ASI4 s'est tenu à l'Université Jaume I de Castellón de la Plana (Espagne) du 18 au 21 octobre 2007 et fut organisé par Pilar Orus et Pablo Gregori.

Le cinquième colloque ASI5 s'était tenu à Palerme (Italie) les 5-7 novembre 2010, et a été l'occasion d'entendre près de 30 communications qui avaient été retenues par le comité scientifique parmi celles qui avaient été soumises, et ainsi de poursuivre le débat. Il est possible d'accéder aux publications :

Sur la site ASI 5 - http://sites.univ-lyon2.fr/asi5/?page=15

Sur le site de la Revue QRDM - "QUADERNI DI RICERCA IN DIDATTICA" - G.R.I.M. http://math.unipa.it/~grim/QRDM_20_Suppl_1.htm

Le sixième colloque ASI6 s'était tenu à Caen (France) les 7-10 novembre 2012, et a été un fois de plus le lieu de riches échanges, l'occasion d'entendre une conférence de Guy Brousseau, et d'émettre une pensée chaleureuse à la mémoire de Filippo Spagnolo, président du comité d'organisation de ASI 5 qui nous a quittés peu de temps après, au début de 2011. Il est possible d'accéder aux publications :

Sur la site ASI 6 - http://sites.univ-lyon2.fr/asi6/?page=15

Sur le site de la Revue QRDM - "QUADERNI DI RICERCA IN DIDATTICA" - G.R.I.M. http://math.unipa.it/~grim/QRDM_22_Suppl_2.htm

Le septième colloque ASI7 s'était tenu à São Paulo (Brésil) les 27-30 novembre 2013. Il est possible d'accéder aux publications :

Sur la site ASI 7 - http://sites.univ-lyon2.fr/asi7/?page=16

Rappelons qu'en 2009 et 2013, deux ouvrages majeurs francophones ont été publiés chez Cépaduès à Toulouse : Analyse Statistique Implicative. Méthode exploratoire et confirmatoire pour la recherche de causalités[6] et Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités[7].

Ils permettent de faire un point global sur l'actualisation significative de l'ouvrage de 1996, *L'implication statistique*, édité par La Pensée Sauvage Editeur (Grenoble France). Une autre publication majeure anglophone a été réalisée avec l'ouvrage[8] de 2008, *Statistical Implicative Analysis* diffusé par Springer-Verlag, (Berlin-Heidelberg, Allemagne). Enfin une troisième en espagnol a aussi été produite avec l'ouvrage[9] de 2009 *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, diffusé par Universitat Jaume-1, (Castellon Espagne).

Dans ce huitième colloque A.S.I.8 de 2015 à Radès, les dimensions internationale et interculturelle, initiée dès ASI5 par l'introduction de cinq idiomes de travail : anglais, espagnol, français, italien et portugais, a été maintenue, certes non sans difficultés et obstacles à dépasser, mais parce que cela nous semble fondamental pour enrichir nos champs de recherche.

Les thèmes privilégiés furent :

- Concepts fondamentaux en ASI : modèles statistiques, types de variables, variables principales et supplémentaires
- Avancées nouvelles en cours, stabilité des indices, intensité d'implication entropique ; extension à de nouveaux types de variables, espace des sujets continu ; règles d'exception; dualité espace des sujets - espace des règles, structure métrique et topologie de l'espace des sujets induites par leur contribution ou leur typicalité, analyse vectorielle ;
- Comparaison critique des démarches, des modèles, des représentations et des résultats de l'ASI avec ceux d`autres méthodes d'analyse de données (treillis de Galois, réseaux bayesiens, arbres d'induction, analyses factorielles, etc.)
- Pratique du logiciel CHIC, les développements actuels et attendus
- Applications traitées par l'ASI et comparativement avec d'autres méthodes, dans les domaines de la didactique, des sciences de l'éducation, de la psychologie, de la sociologie, de l'économie, de l'histoire de l'art, de la biologie, de la médecine, de l'archéologie, etc.

---

[6] GRAS, R., REGNIER, JC., MARINICA, C., GUILLET, F. (Eds) (2013**)** Analyse Statistique Implicative. Méthode exploratoire et confirmatoire pour la recherche de causalités. Toulouse : Cépadues

[7] GRAS R., RÉGNIER J.-C., GUILLET F. (Eds) (2009) Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités. RNTI-E-16 Toulouse : Cépaduès

[8] Gras R., Suzuki E., Guillet F.., Spagnolo F. (Eds) (2008) *Statistical Implicative Analysis*, Berlin-Heidelberg : Springer-Verlag, ISBN : 978-3-540-78982-6

[9] Orus P., .Zemora L., Gregori P. (Eds) (2009) *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, Castellon : Universitat Jaume-1,, ISBN : 978-84-692-3925-4

- Présentations graphiques et numériques des résultats applicatifs, aides à l'interprétation de ces résultats, rôles respectifs et critiques des types de variables, des variables principales et supplémentaires choisies
- Spécificités de la formation à l'ASI : usage du logiciel CHIC, interprétation des représentations graphiques (graphe implicatif ; arbre de la hiérarchie cohésitive)
- Problématiques de didactique de l'ASI

Mais ce colloque était aussi le lieu de confrontation des débats autour des défis qui ont été lancés à l'issue de A.S.I. 7

Défis 2015

- Défi 1 : A chaque couple de variables de VxV, associer le vecteur q, indice centré-réduit d'implication. Étude du champ de gradient dq
- Défi 2 : Variables continues de loi uniforme ou quelconque. Exemples attendus.
- Défi 3 : Notion de variable " attracteur ", sommet du cône implicatif. Exploitation sémantique. Exemples attendus.
- Défi 4 : Relation entre l'intensité d'implication et la confiance (probabilité conditionnelle). Exemples attendus.
- Défi 5 : Exploiter la métaphore ensembliste pour définir une intensité d'inclusion prenant en compte l'implication directe et la contraposée
- Défi 6 : Dans le cadre de la conférence Genèse et développement de l'Analyse Statistique Implicative ; rétrospective intuitive parue dans les actes de ASI 7, une étude de l'organisation conceptuelle hiérarchisée de la cognition (par ex. schèmes, procédures, conception, méthode, ..) mériterait d'être faite par le psychologue ou le didacticien s'appuyant sur les classes plus ou moins emboîtées d'une hiérarchie cohésitive. Exemples attendus.

Délibérément le programme de ce colloque A.S.I. 8 s'est organisé, dans la continuité, autour d'une alternance équilibrée entre des communications portant sur des approches théoriques ou sur des mises en application, et des travaux pratiques sur le logiciel CHIC et sa nouvelle version RCHIC.

A.S.I. – Analyse statistique implicative : des sciences exactes aux sciences sociales et humaines.

Le présent ouvrage donne à voir une partie de l'activité de production scientifique suscitée par l'appel à communication du 8ème colloque A.S.I.8. Les contributions constituent les chapitres successifs de ce livre. Mais auparavant nous présentons les textes des conférences réalisées au cours de ce colloque.

La première donnée par Régis GRAS aborde un retour sur le sens de la construction du cadre théorique : ***Un survol paradigmatique de l'analyse statistique implicative***

La seconde, par Yahya SLIMANI, s'intéresse à la question des rapports entre Data Mining, Big Data et A.S.I. : ***Big Data et Datamining: entre mythes et réalités***

La troisième, par François PLUVINAGE s'intéresse à la question: Regards sur les rapports entre statistique et enseignement mathématique

Contributions retenues et publiées par le comité scientifique

A – Développement du cadre théorique A.S.I.

Notion de champ implicatif en analyse statistique implicative
Régis Gras, Pascale Kuntz, Nicolas Greffard

Un mariage arrangé entre l'implication et la confiance ?
Régis Gras, Raphaël Couturier, Pablo Gregori

Analyse d'un questionnaire « enseignants de mathématiques » par différentes méthodes
Régis Gras et Antoine Bodin

Variable nodale et cône implicatif
Dominique Lahanier-Reuter, Régis Gras, Marc Bailleul

Hiérarchie de règles en A.S.I. et conceptualisation
Régis Gras, Nadja Maria Acioly-Régnier

Ajout de la confiance au graphe implicatif
Souhila Ghanem et Raphaël Couturier

Extension de l'analyse statistique implicative au cas des variables continues quelconques
Régis Gras, Jean-Claude Régnier

Classifying objective interestingness measures based on the tendency of value variation
Nghia Quoc Phan, Hiep Xuan Huynh, Fabrice Guillet, Régis Gras

Medidas alternativas de los grados de adhesión de individuos a la implicación y similaridad para variables modales
Larisa Zamora Y Pablo Gregori

Variability of GRAS classical implication index and other rule quality measures under small size sampling from bivariate binary processes
Pablo Gregori, Raphaël Couturier

Pourquoi et comment transformer des variables quantitatives en qualitatives ? Application à la mélodie de la langue française
Martine Cadot, Anne Bonneau

Apport de la combinaison de la méthode d'analyse statistique implicative (ASI) avec la théorie de réponses aux items (IRT)
Hayette Khaled, Raphaël Couturier

B – Application aux recherches en didactiques disciplinaire et professionnelle

La coordinación de los procesos de aproximación en la comprensión del límite de una función en un punto. Una aproximación a través del análisis implicativo
Joan Pons, Julia Valls, Salvador Llinares

Les supports d'enseignement dans la représentation du métier chez des professeurs des écoles débutants.
Marc Bailleul, Laurence Leroyer

Uma análise das diferentes praxeologias no ensino das equações de segundo grau. um olhar a partir da quadro teorico da A.S.I
Marcus Bessa De Menezes, Marcelo Câmara Dos Santos

Professores que ensinam matemática nos anos iniciais do ensino fundamental: Análise das Tendências de Pesquisa no Brasil (2006-2014)
Luciana Silva Dos Santos, Marcelo Câmara Dos Santos,
Nadja Maria Acioly-Régnier, Edênia Maria Ribeiro Do Amaral

L'approche socioconstructiviste dans les situations d'enseignement-apprentissage de la biologie en Tunisie : le cas de la reproduction humaine
Fadhila Farhane Horrigue

Étude des significations données à la notion de fraction par des élèves de CM1 et de CM2 de l'École primaire en France
Abdul Aziz Alahmadati

Análise das pesquisas didáticas sobre função afim no Ensino Fundamental e Médio no quadro da análise estatística implicativa
Aveilson José de Santana, Vladimir Lira Veras Xavier De Andrade,
Jean-Claude Régnier

Représentations et valeurs implicites des enseignants gabonais sur l'éducation à la santé
Laurence Ndong

L'enseignement de la résolution de problèmes mathématiques à l'école primaire au Maroc : représentations des enseignants à l'égard de leurs pratiques

Brahim El Mekaoui, Fadhila Farhane Horrigue

Mestrado profissional em ensino de matemática: articulação de competências

Silvia Maria De Aguiar Isaia, Eleni Bisognin, Vanilde Bisognin, Jean-Claude Regnier, Nadja Acioly-Régnier, Andréia Cardoso Silveira

C – Application aux recherches du champ de la psychologie

Le traitement statistique de données et la complexité de l'expression et du développement différentiels du sujet psychologique

Christian Pellois

Le sujet psychologique : la complexité différentielles de son expression et de son développement au regard du traitement statistique de données

Christian Pellois

Bem-estar subjetivo de estudantes brasileiros da área de ciências exatas: abordagem dos dados no quadro da análise estatística implicativa

Fabiana Ferreira, Céline Faure, Sofiane Bouzid

D – Application aux recherches du champ des TICE

Les potentiels pédagogiques de l'usage du tableau numérique interactif à l'Ecole primaire en France à la lumière de l'analyse statistique implicative

Hassan Alcheghri

Mapeamento de conhecimentos de professores sobre Tecnologias de Informação e Comunicação e seus usos didático-pedagógicos

Saddo Ag Almouloud, Cileda De Queiroz E Silva Coutinho Maria José Ferreira Da Silva

E – Application aux recherches en Histoire de l'art

Divers motifs dans les représentations de l'Ascension du Christ en occident entre le IXe et le XIIIe siècle. Étude de liens quasi implicatifs

Magali Guenot, Jean-Claude Régnier

F – Meta-analyse des publications contributives à l'A.S.I.

A formação humana dos educadores atraves das comunicações nos colóquios internacionais de Análise Estatística Implicativa (A.S.I.): casos dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7

Djailton Pereira Da Cunha, Nadja Maria Acioly-Régnier, Aurino Lima Ferreira

G – Application aux recherches du champ de la sociologie de l'éducation

Expectativas de inserção no mercado de trabalho dos estudantes cotistas e não cotistas da universidade federal da Bahia-Brasil: um estudo no quadro da análise estatística implicativa

Andréia Cardoso Silveira, Jean-Claude Régnier, Nadja Maria Acioly-Régnier, Robinson Moreira Tenório

A questão da permanência e desistência dos estudantes de licenciatura em ciências e matemática no brasil. Estudo exploratório no CFP/UFCG abordado pelo quadro da análise estatística implicativa

Valéria Maria De Lima Borba, Anna Paula De Avelar Brito Lima, Jean-Claude Régnier

UN SURVOL PARADIGMATIQUE DE L'ANALYSE STATISTIQUE IMPLICATIVE

Régis GRAS [1]

A PARADIGMATIC SURVEY OF THE IMPLICATIVE STATISTICAL ANALYSIS

RESUME.

Cette conférence présente tout d'abord, l'origine de la situation fondamentale dans laquelle la nécessité d'organiser des comportements de réponse d'élèves à un test de didactique des mathématiques est apparue, en respectant la complexité a priori d'exercices. Cela a conduit à la création d'un indice d'implication entre items de réponses, pour évaluer des règles comme : *« si a alors généralement b »*. puis à des représentations du préordre partiel obtenu entre les réponses. La théorie, dénommée alors Analyse Statistique Implicative, s'est ensuite développée, sous la poussée des applications variées rencontrées, par extension de la nature des variables de comportement à des variables non binaires. Enfin, une relation topologique duale a été établie entre les sujets et les variables. Ce développement est de type paradigmatique car il associe un ensemble de problèmes à étudier et de techniques propres à leur étude..

Mots-clés *: taxonomie, rapport non-linéaire tout/partie, dialectique, système dynamique, stabilité structurelle, propriété émergente, règle, métarègle, intensité d'implication, graphe implicatif, hiérarchie cohésitive, variables binaire, modale, numérique, intervalle, floue, vectorielle, supplémentaire, structure topologique duale*

1 Introduction

L'A.S.I. est le dernier plat sorti de la marmite dans laquelle j'ai, au cours de mes 62 ans de service dans l'Éducation Nationale (de la maternelle à l'université), ajouté, mélangé un nombre important d'ingrédients en un ensemble jubilatoire et passionné.

Je me propose, avec une telle composition de la marmite, de vous raconter l'histoire de l'Analyse Statistique Implicative, sa raison d'être, son fondement épistémologique, ses développements, sans rentrer dans les détails techniques, les formules mathématiques, bref en ne faisant appel qu'à votre bon sens et votre intuition[2].

2 Problématique d'ordre psycho-didactique

Au cours des années 70, dans le cadre des Instituts de Recherche sur l'Enseignement des Mathématiques de France, j'ai fréquenté une nouvelle fois des classes du

[1] Ecole Polytechnique de l'Université de Nantes, Equipe DUKE, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, E-mail : regisgra@club-internet.fr,

Site : http://math.unipa.it/~grim/homegras_03.htm

[2] « … il n'y a de connaissance vraie que par l'intuition, c'est-à-dire par un acte singulier de l'intelligence pure et attentive, et par la déduction, qui lie entre elles les évidences », M.Foucault, « Les mots et les choses », p. 103

secondaire, particulièrement du 1er cycle (11 à 15 ans) où j'y ai conduit et évalué une expérience nationale tout en participant à la formation continue et aux recherches des enseignants de ces classes. J'ai donc été le témoin des difficultés d'apprentissage des élèves et quelquefois l'acteur des tentatives de remédiation aux obstacles qui s'opposent à l'assimilation des notions enseignées. Ces obstacles n'étaient pas toujours rencontrés par les élèves mais certains étaient récurrents et relativement partagés. Leur nature relevait de la didactique mais aussi souvent de l'épistémologie en s'opposant à l'acquisition de leurs connaissances[3]. Je pouvais observer ces problèmes d'apprentissage aussi bien directement par l'attitude ou l'expression orale des élèves mais également à travers des questionnaires ou des travaux écrits. Faisant l'hypothèse que les attitudes ou les comportements de réponse étaient globalement identifiables, je disposais de données constituées de traces laissées par les élèves. A l'occasion de résolution d'exercices de mathématiques ou de problèmes, une certaine hiérarchie de difficultés segmentait l'ensemble des élèves interrogés. Plus la difficulté s'accroissait, plus le nombre de réussites diminuait, ce qui peut sembler une tautologie : on peut en effet s'attendre à ce que tout élève qui réussit une épreuve jugée difficile, dans un contexte qui serait comparable, réussirait a fortiori ce qui était facile. Ce qui pourrait **étonner et le contester**, ce sont les incohérences par rapport à cet attendu. L'idée que la difficulté a priori soit définissable objectivement par ma propre pratique ne tenait plus en tant que prédicteur, même si elle était le plus souvent respectée. Ainsi, c'est la relation stable et relativement prévisible entre réussites et échecs, entre comportements de réponse qui m'intéressait plutôt que la réussite ou l'échec à un item donné. Ce qui rejoint l'opinion de H. Poincaré qui dit « *que les mathématiciens n'étudient pas les objets mais les relations entre les objets* ».

D'où l'idée, afin d'aider les enseignants dans l'évaluation d'un niveau d'acquisition d'un concept mathématique donné et dans un projet de construction d'épreuves calibrées, de modéliser des niveaux d'acquisition, en une **taxonomie d'objectifs cognitifs** Celle-ci, à l'instar de celle de Bloom, la plus connue, visait à organiser a priori, selon un ordre de complexité croissante et à travers une analyse des tâches, la maîtrise ou l'appropriation d'un concept (et non pas les moments de son apprentissage). Par exemple, un objectif s'exprimant en termes d'utilisation d'un algorithme serait considéré de complexité inférieure à celle d'un objectif exigeant la construction d'un contre-exemple. Une relation de type causal sous-tendrait cette hypothèse : les outils cognitifs d'un objectif supérieur seraient suffisants à ceux que mobilise l'élève pour un objectif de niveau inférieur, comme une « conséquence » ou un « effet » serait le fruit d'une « cause ». Dit autrement : « résoudre un exercice complexe » impliquerait « résoudre un exercice moins complexe » et sa réussite en serait un **bon prédicteur**.

Relativement à un questionnaire constitué d'items spécifiant chacun des objectifs cognitifs de la taxonomie, en théorie on aurait pu attendre des réussites organisées linéairement selon la complexité a priori. Ce qui n'a pas été observé. A l'ordre total présumé s'est substitué un **préordre partiel**, comme sont définis les stades différentiels de développement de l'enfant chez Piaget. Nous abordons cette question avec plus

[3] « *C'est en termes d'obstacles qu'il faut poser le problème de la connaissance scientifique* » écrit G. Bachelard dans « La formation de l'esprit scientifique »

d'attention dans la communication ASI 8 : « Hiérarchie de règles et conceptualisation ». L'observation de ce préordre partiel signifie que des élèves pouvaient dans certains cas et pour quelques-uns d'entre eux, réussir à un item *a* jugé difficile tout en échouant à un item *b* jugé plus facile. Et ceci sans remettre en cause l'affirmation que «généralement la réussite à *a* s'accompagne de la réussite à *b* » et sans que sa réciproque ne soit nécessairement vraie. Mon intérêt va alors porter sur ce type de relation non symétrique, tenter de pondérer la qualité de son caractère approximatif et d'organiser si possible l'ensemble des couples de variables-items en jeu de ce préordre partiel.

Quels **outils statistiques** étaient alors à ma disposition pour qualifier et quantifier cette relation non symétrique entre deux variables ? Un **test paramétrique** non symétrique ? mais pour réfuter quelle hypothèse ? que les élèves qui ont répondu à *a* ont aussi répondu à b ? que faire de la réfutation ? la ranger sagement ? Établir une liste de cas réfutés ou acceptés ? non, pas de structure globale attendue d'une telle liste ; utiliser la mesure de liaison entre deux variables sur la base de leur **corrélation** ? Mais cette mesure quantifie la qualité des co-occurrences et est donc symétrique. Utiliser **une méthode multidimensionnelle d'analyse de donnée**s afin d'organiser les relations en un tout ? J'avais alors entretenu des collaborations avec J.P. Benzecri sur **l'A.F.C.** et I.C. Lerman sur la **classification hiérarchique** et, très souvent, enseigné et utilisé leurs méthodes d'analyse. Mais leurs fondements théoriques sont essentiellement **symétriques**. Ma réserve était la même que pour la corrélation. D'ailleurs, soulignant bien la différence fondamentale de points de vue, la métaphore ensembliste suivante éclairait la différence et facilitait la compréhension intuitive de la problématique de quasi-implication : parmi la population de sujets concernés par l'étude, le sous-ensemble A des sujets qui satisfont *a* est presque contenu dans le sous-ensemble B des sujets qui satisfont *b* . Restait la **théorie Bayesienne** qui offre le moyen de calculer ce que l'on appelle la « probabilités des causes ». Elle est d'une grande efficacité mais - je l'ai souligné dans un article et Martine Cadot en a étudié la comparaison avec l'ASI – sans nier cette efficacité, elle me semble présenter moins de sensibilité aux effectifs des échantillons (gênant en statistique) et écrase quelque peu les cas rares. Circonstances qu'évitera l'ASI et qu'Yves Kodratoff a exprimées par la recherche « *des pépites de connaissances* ».

Point de départ épistémologique, il me fallait donc établir un **indice**, compris entre 0 et 1 par exemple, capable de rendre compte de l'écart entre **prédiction** et **contingence**, c'est-à-dire entre ce qui était attendu de l'ordre a priori [4](A est inclus dans B dans la métaphore ensembliste) et ce qui était effectivement observé, à savoir une **règle de quasi-implication** « **si *a* alors généralement *b*** ». La stratégie que j'ai alors utilisée, en 1978, a consisté à prendre plutôt en considération la non-satisfaction de l'implication « **si *a* alors *b*** » qui, comme on le sait, apparaît dès lors que *a* étant vrai, *b* est faux. Ce sont donc les **contre-exemples** sur lesquels va porter mon attention. Comme je lui souhaitais une vertu inductive, je devais prendre en compte les effectifs des populations concernées : effectif total des élèves, nombre de ceux qui satisfont *a* et ne satisfont pas *b*. De plus, sans information sur la population, ni sur l'existence d'une relation entre les variables étudiées, j'ai fait l'hypothèse d'une **absence de liaison a**

[4] « Contrairement à l'opinion commune, la grande affaire de la science est moins la production de vérités absolues et universelles ou la reconnaissance d'erreurs rédhibitoires, que la délimitation des conditions de validité d'énoncés… » (J.-M. Lévy-Leblond, « Aux contraires », p.35)

priori entre elles, comme le faisait I.-C. Lerman, comme il est fréquent dans des tests non paramétriques et comme l'exprime H. Atlan[5]. L'objectif est, si possible, de la conserver et la qualifier en mettant en évidence la faible probabilité de la dérogation à la règle sur des bases statistiques. Belle illustration de l'adage : « **l'exception confirme la règle** ».

3 Une mesure statistique de la quasi-implication. Représentations associées

La stratégie a donc été la suivante : si les variables réussite-échec, **booléennes** en l'occurrence, étaient indépendantes, le nombre de contre-exemples aléatoires à la règle de quasi-implication suivrait une certaine loi de probabilité, définissable sur la base des effectifs des échantillons. Si le nombre de contre-exemples attendus avec la probabilité p est supérieur à celui des contre-exemples observés, on admet la règle assortie **d'une mesure de qualité** p. Cette probabilité a été appelée **intensité d'implication**, En tant que telle, elle s'identifie à une échelle de probabilité, propriété que ne possèdent pas d'autres indices numériques comme l'échelle de Guttman, les indices de Loevinger ou de Shapiro, par exemple. Elle représente une sorte d'étonnement ***statistique***, *donc de nature anthropologique*[6], devant le faible nombre de contre-exemples par rapport à ceux qui étaient attendus dans la théorie. Voici une illustration de son caractère subjectif :

- sur un ensemble de 10 individus, des attributs *a* et *b* sont vérifiés respectivement 6 et 8 fois, sans contre-exemple à la règle « si *a* alors *b* ». Celle-ci est logiquement acceptable ; la fréquence de *b* sachant *a* est 1,
- sur un ensemble de 1000 individus, des attributs *c* et *d* sont vérifiés respectivement 600 et 800 fois, avec un seul contre-exemple à la règle « si *c* alors *d* ». La règle logique n'est plus acceptable.

Laquelle vous étonnerait ? A laquelle accorderiez-vous la meilleure qualité prédictive ? Dans le premier cas, la règle est stricte mais la confiance en elle est fragile. Dans le second cas, c'est le contraire, l'étonnement est plus grand en dépit de la moindre valeur de la fréquence conditionnelle. Ce paradoxe relatif à l'acceptabilité de la règle est soluble dans la subjectivité. La statistique ASI va en restaurer une composante objective. Certes, l'intensité d'implication est d'autant plus proche de 1 que la qualité pressentie de l'implication est grande. La relation qu'elle établit entre variables s'exprime fréquemment en termes de causalité. Mais cette *explicabilité causale* d'une variable par une ou plusieurs autres (J.-M. Levy Leblond parle de « **champ causal** »

[5] « …[en accord avec Jung] si la fréquence des coïncidences n'excède pas de façon significative la probabilité qu'on peut leur calculer en les attribuant au seul hasard à l'exclusion de relations causales cachées, nous n'avons certes aucune raison de supposer l'existence de telles relations. », Atlan H., (1986) à tort et à raison. Intercritique de la science et du mythe, Paris : Seuil, p.160

[6] C'est aussi ce qu'affirme René Thom (« Paraboles et catastrophes », 1980, p.130) : « …le problème n'est pas de décrire la réalité, le problème consiste bien plus à repérer en elle ce qui a de sens pour nous, ce qui est surprenant dans l'ensemble des faits. Si les faits ne nous surprennent pas, ils n'apportent aucun élément nouveau pour la compréhension de l'univers : autant donc les ignorer » et plus loin : « … ce qui n'est pas possible si l'on ne dispose pas déjà d'une théorie ».

dans « Aux contraires ») que l'intensité d'implication évalue, n'est en rien **déterministe.** D'ailleurs, elle n'est pas **transitive** comme nous l'avons formalisé par l'étude des **règles d'exception** où les règles a =>b et b=>c ne s'accompagnent pas de la règle a =>c. Elle ne relève pas non plus d'un **déterminisme probabiliste** comme elle est souvent interprétée à tort : si 0.95 est une intensité d'implication de *a* sur *b*, cela ne signifie pas que *b* se réalise avec la probabilité 0.95 si *a* est réalisée. Ainsi, partis de recherche de règles strictes, respectant la logique platonicienne, j'ai transigé et cherché des quasi-règles, ne respectant plus la logique formelle en raison de ses contre-exemples. Cette démarche illustre -nous y reviendrons- le mode de pensée que l'on dit **dialectique**[7] car il accepte les contradictions et les intègre pragmatiquement tout en enrichissant la connaissance. Supposons donc effectué le calcul de l'intensité d'implication pour chaque paire de variables de la situation expérimentale. On ne conserve pour chaque paire que celle relative au couple conduisant à la plus grande intensité d'implication. Que faire du tableau de toutes ces valeurs ? Comment en dégager les lignes de force qui les structureraient comme le fait un plan factoriel ? Disposant d'un ensemble de relations binaires valuées, j'ai fait le choix de le représenter par un **graphe orienté pondéré et sans cycle,** image plus facilement appréhendable par l'utilisateur expert du domaine, par exemple, la didactique, la psychologie, la sociologie, la médecine, etc.. En général, il ne se réduit pas à un chemin linéaire puisque à une même « cause » peuvent être associés plusieurs « effets » et un « effet » peut être le fruit de plusieurs « causes ». Le problème de représentation d'**ontologies** est alors compatible avec ce cadre (travaux en collaboration avec **Jérôme David**). Les graphes suivants illustrent ces deux situations.

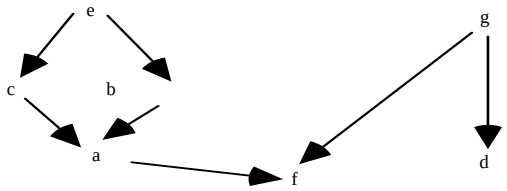


Fig.1

L'arc c →a représente la règle « si la variable c est choisie alors généralement la variable a l'est aussi » ou c=> a. e → c →a est un chemin implicatif.

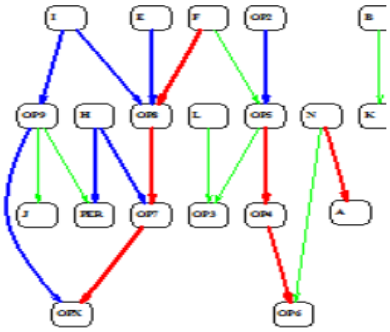


Fig. 2

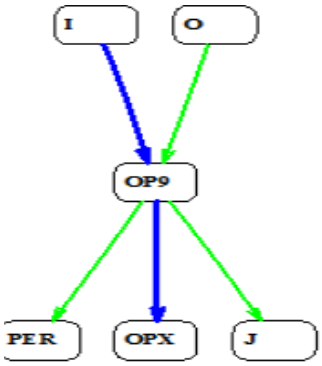


Fig. 3

[7] « …la dialectique n'entrant en scène que pour examiner et résoudre les difficultés logiques de niveau supérieur ». (« Emergence, complexité et dialectique », L. Sève et al, 2005, p. 86).

L'expert analyse et interprète alors les différents chemins en termes conceptuels en donnant du sens dans son domaine d'expertise à des **chemins** du graphe, à des **réseaux** comme dit Marc Bailleul, à un cône ascendants-descendants d'une variable au rôle d**'attracteur** qu'elle joue (Fig. 3)(voir la communication « Variable modale et cône implicatif » de D.Lahanier-Reuter et al), mais du sens aussi aux connexités ou à leur absence. Le graphe relatif au questionnaire construit selon la taxonomie cognitive l'a quasiment validée[8] en organisant généralement dans le préordre attendu les 5 classes et la vingtaine de sous-classes de cette taxonomie.

En psychologie du développement selon Piaget, la notion **d'abstraction réfléchissante** décrit le passage d'un niveau de conceptualisation à un niveau supérieur, chacun des niveaux étant constitué de règles portant sur des objets, puis d'opérations sur ces objets, puis sur des opérations sur ces opérations, etc... (par ex. schèmes, procédures, conception, méthode, ..). On retrouve, d'ailleurs, dans une théorie mathématique ces mêmes élargissements lorsque l'on passe d'un théorème à un corollaire ou, par exemple, dans l'étude des fonctions à celle, en analyse fonctionnelle, de fonction de fonctions. D'où l'idée de construire un second plan de relations implicatives, celui de **règles de règles** selon une **hiérarchie dite cohésitive** en raison de l'indice de **cohésion** qui permet d'engendrer des classes orientées de règles. A l'instar de certaines méthodes d'analyse de données, l'intérêt de passer, non linéairement, du niveau relationnel à celui de hiérarchie m'est apparu et nous avons étendu notre recherche de règles à celui de règles de règles ou méta-règles ou règles généralisées.

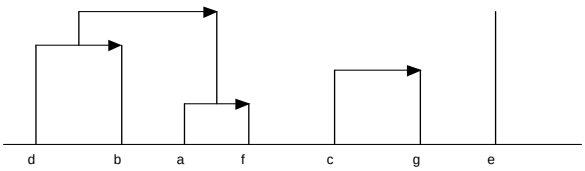


Fig. 4 - Graphe de la hiérarchie orientée

La règle $(d \Rightarrow b) \Rightarrow (a \Rightarrow f)$, illustrée par Fig.4 peut s'exprimer par la phrase : le « théorème » $d \Rightarrow b$ a généralement pour conséquence le « théorème » $a \Rightarrow f$. Par ex, de théorèmes comportementaux, on peut dégager **un trait** ou une **conception** (ex : conceptions du hasard par Dominique Lahanier-Reuter). Un indice statistique permet en outre de repérer les niveaux hiérarchiques correspondant à une **significativité des classes** formées à ces niveaux (en rouge sur la Fig.5). Nous construisons et détaillons le modèle mathématique qui fonde la hiérarchie cohésitive dans la communication « Hiérarchie de règles et conceptualisation ».

[8] Elle est d'ailleurs utilisée pour des évaluations d'acquisitions et pour des constructions de tests mathématiques en France et dans quelques pays francophones.

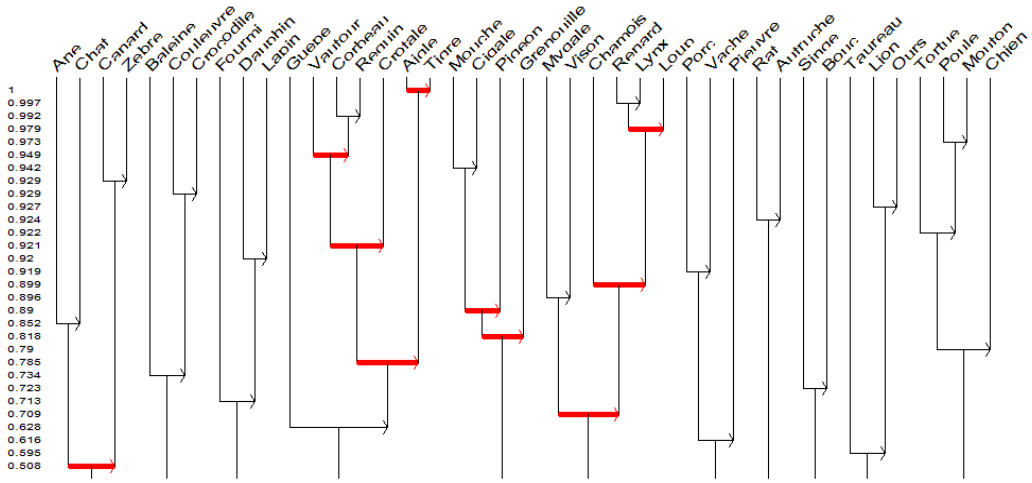


FIG.5 - Graphe de la hiérarchie orientée

4 Un détour par la philosophie des sciences.

A travers ces deux types de représentations de règles simples ou généralisées, qui mettent au jour deux types de structures dans l'ensemble des variables, nous répondons à la philosophie structuraliste qui estime que le **« tout » est plus riche que la somme de ses « parties ».** Je cite, à ce sujet, deux extraits du livre de L. Sève (ib. p. 58) « *…le **tout** ne se compose de rien d'autre que de ses **parties**, et pourtant il présente, en tant que tout, des propriétés n'appartenant à aucune de ses parties. Autrement dit, dans le passage **non additif**, **non linéaire** des parties au tout, il y a apparition de propriétés qui ne sont d'aucune manière précontenues dans les parties et ne peuvent donc s'expliquer par elles* » et de poursuivre *« Tout se passe donc comme si se produisait une génération spontanée de propriétés du tout…C'est le paradoxe de **l'émergence**».*

J'insiste sur cette propriété spécifique de l'ASI que la hiérarchie cohésitive satisfait de façon originale à travers l'extension des relations entre variables. C'est par un saut qualitatif, produit généralement par un effet de seuil dans la quantité (ex. : psychologie de groupe, vaporisation de l'eau…), que le tout, au prix d'une synthèse, prend son sens. Celui-ci s'extrait de la lecture véritablement dialectique du **rapport non-linéaire tout/partie**[9] (non-proportionnalité et non-additivité de la cause sur l'effet). Avec l'ASI, il y a alors **paradoxe** entre des contraires, lien et absence de lien, car le tout, constitué (on devrait dire *organisé*) de parties en un **système dynamique**, possède des propriétés que ne possèdent pas les parties et qui sont généralement de **niveau supérieur**. De la même façon, et métaphoriquement, en linguistique, la signification d'une phrase ne se fait pas par l'analyse de chacun de ses mots mais par l'interaction de ceux-ci[10]. La

[9].. comme le montrent les équations différentielles non-linéaires de l'indice fondamental par rapport aux paramètres cardinaux des observations, contrairement à ce qui est observable avec d'autres indices concurrents.. « La société n'est pas constituée d'individus, mais exprime la somme des relations, des rapports où ces individus se situent les uns par rapport aux autres » (K. Marx, « Manuscrits de 1857-1858). Une rue n'est pas la somme des maisons qui y figurent. De même une ville n'est pas somme de ses rues, etc.

[10] « Il n'y a ni additivité ni proportionnalité entre le sens des unités (mots) et celui de la phrase. On

logique qui sous-tend ce rapport tout/parties est **dialectique** (et non pas dichotomique) car elle concilie interactivement des contradictions : règle et non-règle, **instabilité d'un système dynamique et stabilité structurelle**. Elle se fonde en règle sur l'inexistence importante du contre-exemple et en métarègles sur la négation de l'entropie, du désordre. En cela, la logique dialectique s'oppose à la logique stricte (du mathématicien) sans, bien sûr, être un sophisme. La fécondité et l'originalité de l'ASI tiennent à ce caractère, particulièrement dans l'analyse hiérarchique manifestement non-linéaire où le tout fonde son sens, non par addition des propriétés de ses parties (sous-classes) mais par la synthèse des interactions inférentielles. C'est par la notion de **niveau significatif** que nous pouvons mettre en évidence le phénomène de **propriété émergente**. En ce sens, l'A.S.I. apparaît comme une sorte d'avatar de l'apprentissage non linéaire des connaissances, apprentissage fait de ruptures et de reconstruction dialectique (cf. l'épistémologie de G. Bachelard ou de Lev Vygotsky et notre communication déjà citée sur la conceptualisation). A l'opposé de l'A.S.I., le rapport tout-parties serait **linéaire** dans le cas d'emboîtements de classes comme en clustering fondée sur la similitude jusqu'à son nœud terminal.

Des premières applications, en didactique des mathématiques et sur des problèmes de gestion humaine, il est apparu que le premier indice, l'intensité d'implication, pêchait dans sa discrimination lorsque le nombre de sujets s'accroissait. D'où la nécessité, pour cette raison et pour approcher au plus près de l'interprétation causale des règles, d'intégrer à cette intensité l'information apportée par la **contraposée de l'implication**. Ainsi, non seulement la mesure de l'implication va prendre en compte la relation «si *a* alors généralement *b* » mais également, dialectiquement, sa contraposée « **si *non b* alors généralement *non a*** ». Cet indice nouveau est basé sur la notion d'**entropie,** donc **d'information,** des expériences évaluant les deux règles. Ce choix permet d'agir plus en profondeur pour accéder aux « pépites de connaissances », règles qui seraient rejetées ou ignorées par une méthode basée sur le support et la confiance (l'algorithme a priori d'Agrawal relève, semble-t-il, de la pensée linéaire).

Bien que nous recherchions des relations entre variables par des règles non symétriques et que ces relations puissent souvent s'exprimer en termes de causalité, nous ne prétendons pas, comme je l'ai dit plus haut, qu'elles soient déterministes, mais simplement qu'elles ont la capacité de permettre d'émettre des hypothèses quantifiables sur leur prédictibilité.

5 Un logiciel de traitement informatique de l'A.S.I.

Lorsque le nombre de sujets et de variables croît, il est difficile d'effectuer les calculs associés et surtout de fournir les deux représentations : graphe et hiérarchie. Au début, les deux tâches étaient faites à la main. La plaie ! Je l'ai fait pour ma thèse. J'ai

---

voit se dessinée une topologie du sens » (F. Gaudin dans « Emergence, complexité et dialectique »).et « …le mot isolé de la langue chinoise n'a en vérité ni signification ni existence à part, chacun ne reçoit sa signification que du parler même (de l'intonation, etc….), pris isolément il a dix, voire quarante significations,… ; si nous soustrayons ce mot à la totalité, il se perd dans une creuse infinité. « (F.-W. Schelling, « Philosophie de la mythologie », p.361).

alors écrit un premier programme en Basic (!) qui effectuait les calculs et construisait la **hiérarchie**. Un doctorant de Lerman -H.Rostam- construisit à son tour le **graphe implicatif** à l'aide d'un programme plus sophistiqué dont j'avais construit l'algorithme. Puis, deux de mes doctorants, Saddo Ag Almouloud et Harrisson Ratsimba-Rajohn, intégrèrent l'ensemble en un seul logiciel que nous avons dénommé **Classification Hiérarchique Implicative et Cohésitive** sous l'acronyme « **C.H.I.C**. ». Enfin, depuis la fin des années 90, Raphaël Couturier a unifié le traitement complet des calculs et des représentations tout en lui apportant continûment les améliorations dues au développement de la théorie sous-jacente et de la variété des applications. En particulier, une option permet de faire apparaître les différents éventuels prédicteurs et les descendants d'une variable, mais également de déterminer la conjonction optimale de ceux-ci au sens de leur originalité. La possibilité de modifier le seuil de construction du graphe implicatif met en lumière la plasticité de la structure des variables, tout en conciliant, dialectiquement l'instabilité d'un système dynamique et sa stabilité structurelle. Une version écrite en R est présentée par Raphaël Couturier lors du colloque ASI 8.

Ce logiciel est opérationnel à travers le monde puisque 26 pays, gréco-latins pour la plupart, le possèdent et l'utilisent[11]. Il y joue, pour la recherche, le double rôle de **révélateur et d'analyseur**. Si bien que des questions sur les possibilités et les limites de l'ASI se font jour à travers des questions d'ordre général mais aussi spécifiques du fait des traditions et des cultures différentes. Outre le lien épistolaire, elles se renforcent au fil de nos rencontres internationales sur l'ASI, de ASI 1 à ASI 7 en 2013 et maintenant en 2015 en Tunisie. Ce sont ces questions qui poussent la théorie et son outil vers des développements continus comme nous allons le voir. Pour illustrer ceci, je citerai Anne Lauvergeon dans « La femme qui résiste » (Plon, 2012) : « … *lorsque l'on produit, on finit également par concevoir* ».

6 Extension de la méthode à d'autres variables

C'est donc à travers les différentes situations rencontrées que la limitation aux variables binaires, ayant servi à donner un sens ensembliste à l'implication, est apparue contraignante. Celui qui a tiré le premier est Marc Bailleul qui a voulu rechercher des relations de type préférences entre des assertions d'enseignants. Des **variables modales** (« un peu », « beaucoup », « pas du tout », …), devaient être traitées. Il a proposé un premier indice satisfaisant à la mesure d'expression du type : « **si pour *a* la modalité « *un peu* » est choisie alors généralement pour *b* une modalité supérieure ou égale est choisie »**. Dans sa thèse, Marc Bailleul a obtenu d'excellents résultats, dont certains imprévisibles, relativement à 4 conceptions de l'enseignement des mathématiques sous le regard des enseignants. Nous utiliserons son étude dans plusieurs communications de ASI 8. Afin de procéder, relativement à la même problématique, par extension de l'intensité d'implication entre variables booléennes, avec J.-B. Lagrange, nous avons alors défini une nouvelle mesure portant sur des **variables numériques** et modales permettant d'attribuer une valeur à des énoncés comme : « **si $a = \alpha$ alors généralement $b \geq \beta$** ». Jean-Claude Régnier, de son côté, a ramené la problématique de la recherche de

[11] A l'heure actuelle, Pablo Gregori (Castellon) et Ruben Rodriguez (Quito) œuvrent pour traduire CHIC en une version informatique sous R

relations entre préférences à celle de **variables de rangs** qui a fourni une autre extension de l'A.S.I...

De là, suite à une question d'E. Diday se plaçant dans le cadre des **variables symboliques** et par une extension des variables numériques, nous avons affecté des valeurs à des expressions « si ***a*** **prend ses valeurs dans l'intervalle I_a alors généralement *b* prend ses valeurs dans I_b** ». L'idée maîtresse a consisté à partitionner l'étendue des valeurs prises par chaque variable en sous-intervalles maximisant leur variance interclasse. L'intérêt de cette nouvelle catégorie de variables, **dites intervalles**, se manifestait dans l'enseignement, pour comparer hiérarchiquement des performances dans des disciplines différentes et dans la recherche de transfert de compétences spécifiques vers d'autres compétences. On voit alors que de ces **variables-intervalles** traitées en collaboration avec E. Diday et Pascale Kuntz, nous parvenons naturellement à **des variables floues** comme nous les rencontrons dans l'étude de relations du type : « si la tension *a* d'un sujet est plutôt élevée alors généralement son rythme cardiaque *b* l'est aussi ». On pressent les applications pratiques qui peuvent en découler dans des problèmes de construction ou de détections de pannes. Grâce à la collaboration de Fabrice Guillet, Maurice Bernardet, Raphaël Couturier, Filippo Spagnolo, cela m'a permis de modéliser la notion de variable floue dans le cadre des variables-intervalles et d'en faire une présentation dans un congrès sur la logique floue. Un problème est survenu un jour dans le traitement implicatif d'un ensemble trop large de variables au point de rendre illisibles les graphes obtenus. Avec Raphaël Couturier, Fabrice Guillet et Robin Gras, nous avons défini une relation d'équivalence entre les variables, basée sur leurs comportements implicatifs voisins, qui a conduit à substituer à un paquet de variables un représentant leader de ce paquet. Cette **réduction** s'est avérée efficace dans de nombreuses autres situations, par exemple dans la thèse de Laurence Ndong.

Un petit arrêt pour parler des collaborations dans l'équipe DUKE, toujours appuyées sur des choix épistémologiques en réponse à des attendus sémantiques. L'explosion d'une multitude de questions applicatives ou théoriques, toujours non symétriques, a conduit à des travaux communs entrepris, depuis 1990, avec certains membres de l'équipe du DUKE actuel : je citerai par exemple, avec Pascale Kuntz sur **les hiérarchies de règles** (ah ! la démonstration de l'ultramétricité de la hiérarchie ! le modèle algébrique de la hiérarchie !), sur les **règles d'exception** avec Einoshin Suzuki, la **redondance de règles** avec Pascale, avec Julien Blanchard sur **l'analyse entropique**, les **variables temporelles**, **l'expression de gènes en bio-informatique** avec Gérard Ramstein, etc.. Par exemple, comment la nécessité de variables temporelles est-elle apparue ? Eh bien, pour rendre compte de l'évolution des relations entre variables économiques, sociales ou cognitives. En effet, peut-on expliquer dans l'enseignement, l'implication entre variables lorsque l'on procède à des interventions en cours d'année sur l'apprentissage ou encore en socio-psychologie par des interventions successives, par des entretiens ou par des stages (en collaboration avec D.Pasquier). Nous avons alors formalisé ces variables indexées par le temps en **variable vectorielle,** où une variable est modélisée par un vecteur paramétré par le temps. Puis, Julien Blanchard, en collaboration avec Fabrice Guillet et moi-même, a, de façon différente, défini des **variables séquentielles** modélisées par un processus de Poisson. Autant de nouveaux concepts et de nouveaux champs d'application nés de problématiques variées et

d'extensions attendues. **Des extensions généralisantes** de l'ASI ont vu le jour récemment :

- d'une part, à un ensemble continu de sujets, par exemple des couleurs, des opinions, muni d'une loi de distribution donnée, généralisation puisque jusqu'alors le modèle se limitait à des ensembles discrets et finis ;
- d'autre part, à l'ensemble des valeurs prises par les différents types de variables dans des espaces continus, par exemple des champs, munis de lois données ;
- enfin, lors du colloque ASI 8, nous présentons un prolongement à des variables continues, ce qui représente une avancée théorique reconnue majeure en analyse de données.

J'insiste sur une remarque susceptible de satisfaire tout cartésien : lors de chaque extension, nous nous sommes efforcés de prouver et nous y sommes parvenus, que **la restriction au cas fondamental de variables binaires et d'espaces discrets était toujours satisfaite**. L'emboîtement mathématique est original et rigoureux.

7 Rôle explicatif de variables supplémentaires

Une autre question m'a taraudé très tôt. Y a-t-il dans la population de sujets concernés par une étude, une structure interne qui expliquerait la structure des variables révélée par l'ASI ? Supposant une structure implicative de variables obtenue par un graphe ou une hiérarchie, est-il possible de désigner les sujets et les catégories de sujets plus ou moins responsables des éléments de ces structures ? Par exemple, si nous observons que dans l'ensemble de classes scolaires une certaine conception de notions géométriques entraîne certaines conduites de réponses, à quel type d'élèves peut-on plus spécifiquement l'attribuer ? Inversement quels sont les élèves qui y seraient réfractaires ? Avec les thèses de Harrisson Ratsimba-Rajohn et Marc Bailleul, nous avons formalisé et exploité, dans l'A.S.I., la notion de variable supplémentaire (ou **exogènes** par opposition aux variables **endogènes** analysées) et sa **contribution** (on parle aussi de **typicalité**) à des éléments de structure, des **réseaux** ou des classes cohésitives par exemple, sémantiquement expliquées. Mieux encore, nous avons pu définir de façon rétroactive une **structure topologique duale** sur l'ensemble des sujets. Ainsi, par exemple, une conception « emblématique de l'école républicaine » est renforcée dans un milieu scolaire favorisé où la distance inférée entre les élèves par un élément des structures de règles est la plus faible (travaux avec Dominique Lahanier-Reuter). Mais puisque j'ai cité au fil des pages les apports et le rôle d'aiguillons de collaborateurs pour le développement de l'ASI, en particulier de l'équipe DUKE, je dois mettre en exergue le rôle critique, poil à gratter mais constructif que joue Jean-Claude Régnier tant à l'égard de la théorie et ses applications qu'à l'égard de la diffusion, la popularisation de l'ASI puisqu'il a pris en charge la présidence des manifestations internationales ASI. Combien de lapsus, d'erreurs, de maladresses, d'excès de précipitation, Pascale Kuntz et Jean-Claude ont-ils redressés ! Je n'oublie pas non plus le dévouement fidèle et les compétences informatiques de Raphaël Couturier sans lequel l'ASI, privée de CHIC, et maintenant de RCHIC, ne serait qu'une construction théorique et peut-être spéculative à contempler. Il a comblé, avec CHIC, mon handicap en informatique comme Pascale a bien voulu le faire pour l'anglais. Je crains d'ailleurs

que cette dernière carence ait coûté une meilleure audience de l'ASI sur le plan international. Cependant, d'autres équipes internationales participent également aux travaux sur le développement et les applications de l'ASI en utilisant le logiciel CHIC. Je cite, en particulier, étant donnée leur régularité de participation :

- conduite activement par Pilar Orus (Université de Castellon en Espagne), une diversité de noyaux hispanisants en Espagne (Pablo Gregori, Eduardo Lacasta, Miguel Wilhelmi, …), à Cuba (Larisa Zamora, …), au Chili (Guzman-Retamal, …) ; en Argentine (Pablo Carranza, …), en Equateur (Ruben Rodriguez) ont créé des liens permanents ;
- en Italie, Filippo Spagnolo, récemment décédé, avait créé une équipe qui se resserre maintenant autour de Benedetto Di Paola ;
- au Brésil, autour de Saddo Ag Almouloud et grâce au renforcement apporté par des conventions avec Jean-Claude Régnier et l'université Lyon II,
- à Chypre autour de Athanasios Gagtasis,
- en Slovaquie avec Lucia Rumanova.
- de façon plus épisodique, des chercheurs de Belgique, de Suisse, d'Allemagne, de Grèce, de Roumanie, de Tchéquie, du Japon, du Vietnam, du Canada, du Mexique, du Gabon, de Tunisie, et j'en oublie, utilisent méthode et outil de l'ASI.

8 Les spécificités de l'ASI et conclusion

L'ASI, mesurée à d'autres méthodes d'analyse de données, présente des caractères originaux importants. Je les résume :

- **modèles successifs** de variables répondant à des contraintes épistémologiques explicites compatibles avec la sémantique des situations à modéliser ;
- **non-symétrie** de la méthode ;
- extension progressive de la nature des variables traitées tout en conservant les propriétés de plongement ;
- **capacités pédagogiques et ergonomiques** des représentations, en particulier pour l'examen des règles généralisées ;
- **dualité structurelle** des deux espaces en jeu : sujets et variables avec les notions de contribution et de typicalité aux structures ;
- **extension du discret fini au continu** tant pour les variables que pour les sujets ;
- **originalité du raisonnement dialectique** à la base de la définition des règles simples et généralisées ;
- la **simplicité du modèle mathématique** sous-jacent lui assurant accessibilité plasticité et fécondité utiles pour répondre à des attentes applicatives dans de larges domaines.

D'aucuns diront que l'ASI, par l'amplitude de ses champs d'application et par l'homogénéité de ses propriétés analytiques et graphiques présente une **nature paradigmatique** originale. Doit-on le reconnaître ? Djamel Zighed, Directeur de

l'Institut des Sciences de l'Homme de Lyon exprime cette idée de la façon suivante : « *l'ASI n'est pas une méthode mais un cadre théorique, large, dans lequel se traitent des problèmes modernes de l'extraction des connaissances à partir des données. C'est une théorie générale dans le domaine de la causalité parce qu'elle répond à des faiblesses d'autres théories, elle apporte un outillage formel et des méthodes pratiques de résolution de problèmes. Ses applications sont multiples... ».* Pour terminer, je voudrais remercier tous les participants à ce 8ème colloque scientifique international sur l'Analyse Statistique Implicative organisé sous la direction scientifique et organisationnelle de Jean-Claude Régnier, Yahya Slimani, Makram Ben Jeddou, Ahmed Dhouibi et Inès Ben Tarbout, pour leurs contributions au développement du cadre théorique, méthodologique et applicatif au travers de la mise en œuvre des concepts et des outils qui le constituent, ainsi que par la confrontation aux défis théoriques traduits par des problématiques nouvelles.

Références

[1] *L'implication statistique. Nouvelle méthode exploratoire de donnée,* sous la direction de R.Gras, et la collaboration de S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn, A.Totohasina, La Pensée Sauvage, Grenoble. ISBN : 2.85919.129.1 (1996)

[2] *Mesures de Qualité pour la Fouille de Données,* H. Briand, M. Sebag, R. Gras et F.Guillet (eds), RNTI-E-1, Cépaduès, 2.85428.646.4 (2004)

[3] *Quality Measures in Data Mining* , F. Guillet et H. Hamilton (eds), Springer, ISBN : 3.540.44911.6. (2007)

[4] *Statistical Implicative Analysis, Theory and Applications,* R. Gras, E. Suzuki, F. Guillet, F. Spagnolo, (eds), Springer, ISBN : 978.3.540.78982.6 (2008)

[5] *Analyse Statistique implicative. Une méthode d'analyse de données pour la recherche de causalités,* sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, ISBN : 978.2.85428.8971. (2009)

[6] *Teoria y Aplicaciones del Analisis Estadistico Implicativo,* Eds : P.Orus, L.Zemora, P.Gregori, Universitat Jaume-1, Castellon (Espagne), ISBN : 978-84-692-3925-4, (2009).

[7] *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire.* J.C. Régnier, Marc Bailleul, Régis Gras, (Eds) Université de Caen, ISBN : 978-2-7466-5256-9, (2012)

[8] *L'Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités* sous la direction de direction de Régis Gras, (Eds.) R. Gras, J.C. Régnier, C. Marinica, F. Guillet, Cépaduès Ed. Toulouse, 201, ISBN : 978.2.36493.056.8. (2013)

BIG DATA ET DATAMINING: ENTRE MYTHES ET REALITES

Yahya SLIMANI [12]

L'évolution de l'informatique a été et reste toujours marquée par l'apparition de concepts, de technologies, etc. dont certaines disparaissent au bout d'un certain temps (systèmes experts, téléinformatique) alors que d'autres sont plus ou moins pérennes (processus, technologie objet, etc.). Avec le développement du Web qui permet de générer de la connaissance collective et l'apparition des réseaux sociaux, nous assistons de plus en plus à l'apparition de termes qui relèvent parfois du *buzz* et ont une résonnance beaucoup plus commerciale que technique. Cette conférence a pour objectifs de :

- Démystifier certaines idées et concepts qui sont essentiellement utilisés à travers des stratégies commerciales pour attirer de plus en plus de sociétés et de clients vers des « technologies informatiques » qui s'avèrent parfois très coûteuses et inefficaces pour les entreprises, sans leur apporter une plus value dans leurs systèmes d'informations.
- Présenter le domaine du datamining, dans lequel les statistiques et l'analyse de données, de manière générale, sont très largement utilisées.
- Présenter la notion de Big Data qui est parfois utilisée à mauvais escient dans le seul but d'inciter les sociétés et les entreprises à investir dans cette nouvelle « technologie », sans se soucier des bouleversements qu'elle peut apporter au niveau de leurs systèmes d'informations et de leurs exploitations.
- Présenter les liens qui existent entre datamining et Big Data pour l'exploitation de larges volumes de données.

[12] Institut Supérieur des Arts Multimédia (ISAMM) Université de la Manouba (Tunisie) http://www.yahya-slimani.net/

REGARDS SUR LES RAPPORTS ENTRE STATISTIQUE ET ENSEIGNEMENT MATHEMATIQUE

Francois PLUVINAGE[1]

Un éclairage de nature socioépistémologique fait bien voir comment les statistiques ont pénétré les mathématiques, tardivement dans l'histoire (XIXe siècle). Jusqu'au XVIIe siècle, les pratiques de référence étaient la prise de données par écrit, et les pratiques sociales touchaient à l'anticipation, par exemple la quantité de blé à attendre de la prochaine récolte, et à des décisions, telles celles des impôts à prélever.

Die
Staatsklugheit
nach ihren ersten
Grundsätzen
entworfen
von
Gottfried Achenwall
ordentlichem Lehrer des Natur- und Völkerrechts wie auch der Politick auf der Universität zu Göttingen und aufserord. Mitgliede der Königl. Gesellschaft der Wissenschaften daselbst.
Göttingen
in Verlag der Witwe Vandenhoeck 1761.

Certains textes, comme celui que donne la Brockhaus Enzyklopädie, prétendent que le terme allemand *Statistik* a été introduit par Gottfried Achenwall (1719-1762). Il semble cependant que le mot, construit sur le latin, soit apparu plus tôt. La figure ci-contre reproduit la page de couverture du livre « *Die Staatsklugheit* » (la sagesse de l'Etat, en allemand) d'Achenwall, ouvrage dans lequel le mot *Statistik* apparaît.

La pénétration des statistiques dans le champ des mathématiques s'est produite lorsque la saisie de données confiée depuis l'antiquité à quelques spécialistes, tels les scribes égyptiens, s'est combinée avec les probabilités au service de prédictions prenant en compte des risques. Dans un premier temps, c'est-à-dire avant que les mathématiques n'incluent la statistique comme l'une de ses branches, les pratiques de référence se sont enrichies de la représentation des données et les pratiques sociales de la formalisation et de la prédiction argumentée. Ensuite se sont glissées dans les pratiques de référence la modélisation et les techniques d'enquête, incluant l'échantillonnage, et dans les pratiques sociales les évaluations de risques.

[1] CINVESTAV-IPN Mexico (Mexique) email: fpluvinage@cinvestav.mx

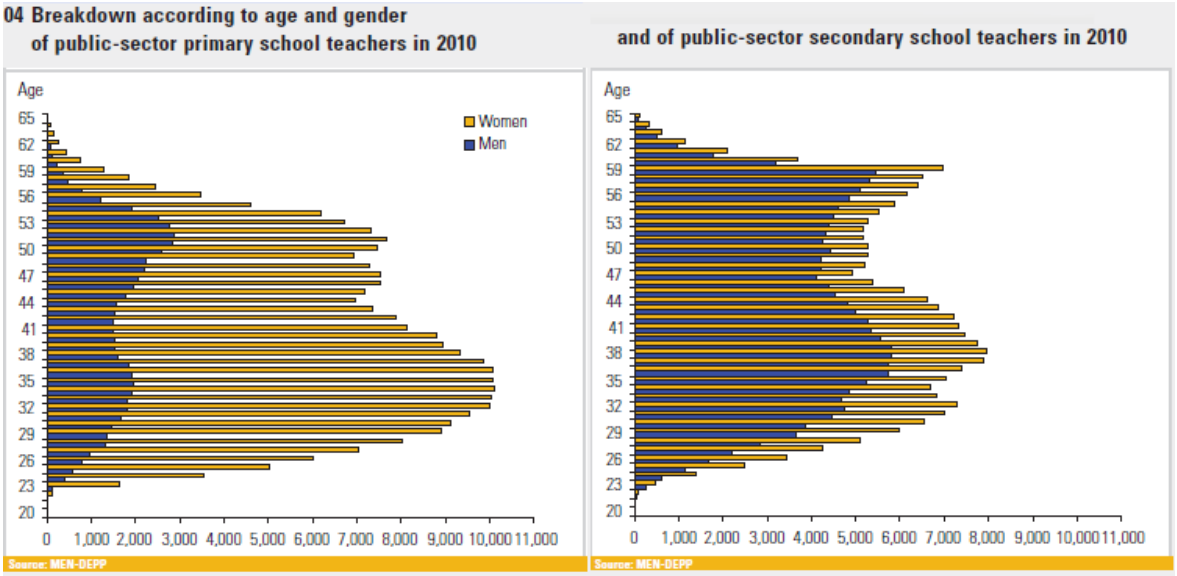


Figure 1: Une représentation qui interpelle : Pyramides des âges par genres des enseignants français des premier et second degrés.

Au service de l'institution, les statistiques d'une part alimentent le système éducatif en données factuelles et d'autre part fournissent des indicateurs de la qualité de l'enseignement. Ce dernier rôle est plus controversé au sein de la communauté éducative, allant jusqu'à être générateur de tricheries et même de fraudes importantes. C'est que les objectifs d'explications de phénomènes touchant l'enseignement et les plans d'améliorations ne coïncident pas nécessairement quand ils se situent au niveau d'un établissement ou d'une population expérimentale et quand ils se situent au niveau de tout un système éducatif.

Probabilités d'accès au lycée en fonction du niveau scolaire au CP (première année d'école primaire) et du milieu social d'origine

Niveau	Faible	Moyen	Bon
Enfants de cadres supérieurs	63,4	80,6	90,8
Enfants de prof. intermédiaires	33,2	54,3	74,0
Enfants d'ouvriers	15,8	31,0	51,9

Source : Duru-Bellat, M., Jarousse, J.-P. et Mingat, A. (1993)

Tableau 1

Il est clair qu'aujourd'hui la mise en œuvre de statistiques avec les moyens de traitement fournis par l'informatique peut donner lieu à des usages sans intérêt, par exemple en appliquant un quelconque test à une population puis en passant ses résultats à la moulinette d'un logiciel d'analyses statistiques. Mais en sciences de l'éducation, touchant en particulier l'enseignement des mathématiques, nombreuses sont les questions de recherche qui demandent de s'appuyer sur l'usage d'analyses statistiques pour leur étude.

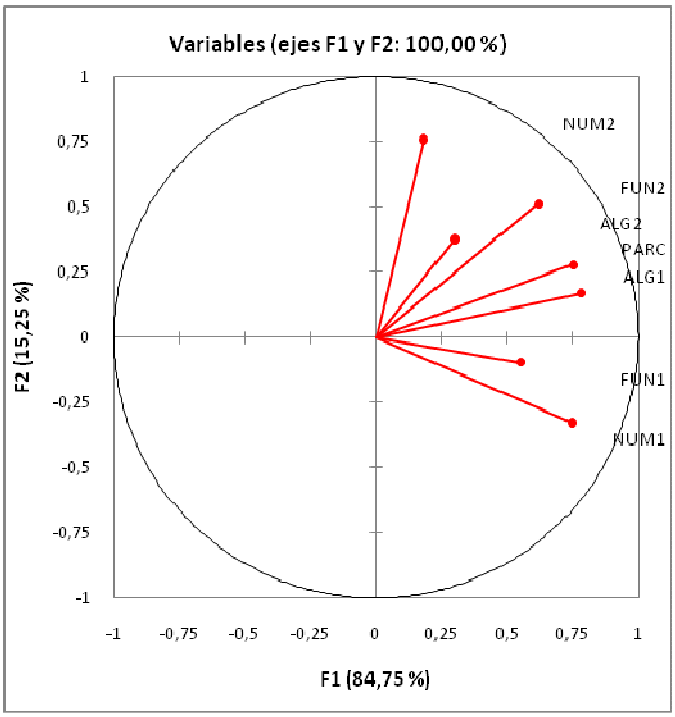


Figure 2: Sortie d'analyse discriminante extraite de Cuevas, Martinez & Pluvinage (2011)

Nous illustrerons ce propos par des exemples, utilisant pour certains des données accessibles au public et s'appuyant pour d'autres sur des résultats d'enquêtes. Au service plus spécifiquement de la didactique des mathématiques, nous verrons que les statistiques sont bienvenues dans des études comparatives et que les techniques d'analyse statistique, comme l'analyse implicative, peuvent s'imposer en contexte expérimental. En rapport avec le dernier volet de notre exposé, la participation de professeurs en exercice dans de telles études nous semble éminemment souhaitable.

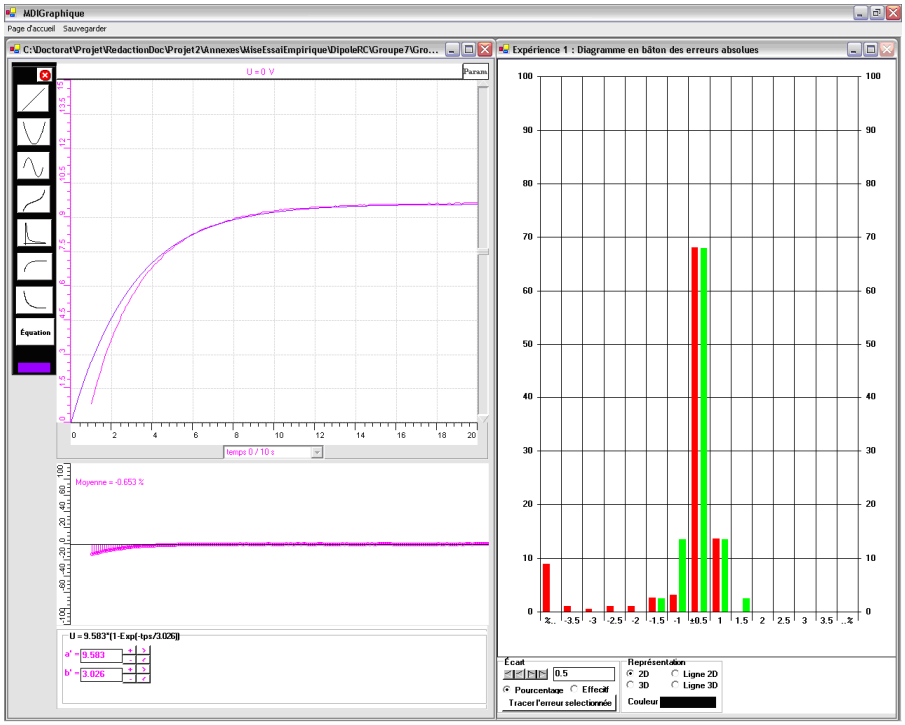


Figure 3: Modélisation avec recours à la statistique, extrait de Touma (2009)

Codage : Réussite forte ou faible, un problème ouvert pour l'ASI

Un problème que nous rencontrons dans des études expérimentales et pour lequel des recherches sont encore à conduire est tout bêtement celui du codage de réponses quand celles-ci donnent lieu, ce qui est fréquent, à des niveaux de prise en compte. Le cas le plus simple est celui d'un intermédiaire entre échec et réussite sur critère mathématique : des réponses non entachées d'erreur, mais incomplètes. On peut parler alors de réussite stricte et réussite large, et on a le choix entre une analyse qui distingue Réussite stricte et l'oppose aux autres réponses (Echec et Réussite large) et une analyse qui regroupe Réussites stricte et large, les opposant aux seuls échecs. Dans un exemple que nous avons proposé pour être analysé dans un atelier que dirigeait Maria Trigueros à Mexico, les arbres de similarités de l'une et l'autre analyse faisaient apparaître deux grands groupes, mais la composition des groupes différait entre les deux cas.

L'enseignement du domaine Probabilités et Statistique

Le dernier volet de notre présentation concerne l'enseignement mathématique de la branche *probabilités et statistique*, et la place que les enseignants lui accordent. Des situations qui nous semblent paradigmatiques et que nous commenterons sont celles de décision en présence d'incertitude et celles de modélisation avec contrôle statistique. Les unes et les autres justifient l'introduction des paramètres statistiques de position et dispersion, des lois usuelles et des outils d'analyse. Leur orchestration en classe sera d'autant plus efficace qu'elle sera conduite par un enseignant qui en a eu l'expérience personnelle dans son vécu.

Etudes citées

Cuevas Vallejo, C. A., Martínez Reyes, M. & Pluvinage, F. (2011). La promoción del pensamiento funcional en la enseñanza del cálculo: un experimento con el uso de tecnologías digitales y sus resultados. *Annales de Didactique et de Sciences Cognitives*, Vol. 17, 137-168

Duru-Bellat, M., Jarousse & J.-P., Mingat A. (1993). Les scolarités de la maternelle au lycée. Etapes et processus dans la production des inégalités sociales. *Revue française de sociologie*. Vol. 34-1. 43-60. Consulté en ligne à

http://www.persee.fr/web/revues/home/prescript/article/rfsoc_0035-2969_1993_num_34_1_4218

Touma, G. (2009). Une étude sémiotique sur l'activité cognitive d'interprétation. *Annales de Didactique et de Sciences Cognitives*, Vol. 14. 79-101

Quelques sources officielles

Online Education Database

http://www.oecd.org/document/54/0,3343,en_2649_39263238_38082166_1_1_1_37455,00.html

Panorama de l'éducation 2010 au niveau international (OCDE)

http://browse.oecdbookshop.org/oecd/pdfs/browseit/9610074E.PDF

Sites français consultés

Données internationales http://www.ciep.fr/sitographie/ries44.php
Données sur la France http://media.education.gouv.fr/file/07/0/3070.pdf
http://media.education.gouv.fr/file/etat20/41/7/The_state_of_education_168417.pdf

NOTION DE CHAMP IMPLICATIF EN ANALYSE STATISTIQUE IMPLICATIVE

Régis GRAS, Pascale KUNTZ et Nicolas GREFFARD[1]

NOTION OF IMPLICATIVE FIELD IN STATISTICAL IMPLICATIVE ANALYSIS

RÉSUMÉ

Dans le cadre de la théorie de l'Analyse Statistique Implicative, la problématique de la stabilité de l'indice qui permet de définir et évaluer la qualité de l'indice d'implication est posée par l'utilisateur qui renouvelle ses expériences dans un domaine particulier. Dans cet article, nous étudions ce problème en invoquant les concepts différentiels de l'analyse mathématique. Nous examinons un à un les paramètres intervenant dans la formule donnant l'indice d'implication. Nous comparons les variations de ces paramètres avec ceux d'autres indices classiques utilisés en fouille de données. Nous étendons cette étude par celle de la structure de l'espace vectoriel qu'ils engendrent et en centrant cette étude sur la notion de gradient implicatif. De là, nous illustrons par une représentation géométrique la problématique de l'équilibre de l'indice via une discrétisation des surfaces équipotentielles.

Mots-clés : *analyse statistique implicative, intensité d'implication, indice d'implication, différentielle, gradient, surface équipotentielle*

ABSTRACT

In the context of the theory of implicative statistical analysis, a user repeating some experimentation in a specific domain is faced with the issue of the robustness of the metric appraising the quality of the implicative index. In this paper, we address this problem through a differential analysis instead of bootstrapping. We study each individual parameter involved in the implicative index equation. And we compare their variations with those of other indices from the data-mining literature. Furthermore, we study the structure of the vector field they span by focusing on the notion of implicative gradient. From there, a geometrical representation is used to illustrate the index equilibrium problematic through a series of figures of equipotential surfaces.

Keywords *: statistical implicative analysis, implication intensity, measure of implication, differential, equipotential surface.*

1 Introduction

Le chercheur en sciences et particulièrement en sciences humaines, expérimentateur ou non, vise, à travers ses interrogations, à construire ou conforter des connaissances dans son domaine. Il est confronté à des situations, vécues par une population de sujets dans lesquelles apparaissent des phénomènes, par exemple des attributs, conduisant à de nombreuses données d'observation. De celles-ci, parmi d'autres informations, il veut extraire des relations d'association entre les variables observées. Puis, à partir de ces relations, il cherche à constituer des savoirs stables et partagés dans son domaine. Parmi

[1] École Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, Site de la Chantrerie, rue C. Pauc, BP 44306, Nantes cedex 3, e-mail : regisgra@club-internet.fr, pascale.kuntz@univ-nantes.fr et nicolas.greffard@univ-nantes.fr.

celles-ci, les relations causales ou règles non symétriques -« cause => effet (ou conséquence) » - figurent en bonne place pour leur intérêt de découvertes à valeur prédictive. Mais rares sont les réponses respectant les principes métaphysiques de la logique formelle c'est-à-dire s'exprimant en terme de « vrai » ou de « faux ». Ces réponses seraient plus acceptables évaluées dans une logique **dialectique,** « logique des contraires », logique qui permet de dépasser les contradictions. C'est cette composante épistémologique qui guide, de façon originale, la méthode d'analyse de données, l'Analyse Statistique Implicative -d'acronyme ASI-, que nous avons élaborée afin que prévalent les **quasi-règles**, c'est-à-dire celles qui acceptent des contre-exemples sans effacer leur plausibilité. En effet, une règle non strictement satisfaite y trouve sa place en même temps qu'elle serait logiquement réfutable ; autrement dit, simultanément, règle et sa négation ont droit de coexister. La contradiction est provisoirement levée en ASI par la donnée d'un seuil probabiliste variable dont le chercheur contrôle la flexibilité : au-delà, la quasi-règle est acceptée, en-deçà elle est mise à l'écart. Un certain déterminisme probabiliste confère à la quasi-règle une prédictibilité statistique. Autrement dit, comme l'énonce le philosophe Lucien Sève (Sève, 2005, p.104), « on perd en rigueur ce que l'on le gagne en richesse » et, ajouterons-nous, en fécondité.

La méthodologie que nous développons dans l'ASI et que nous implémentons dans le logiciel Classification Hiérarchique Implicative et Cohésitive, d'acronyme CHIC, consiste, en partant de la contingence des données, issues d'un ensemble répétitif de phénomènes, à analyser le singulier, le « surprenant ». Celui-ci qui constitue la quasi-règle, débouche dialectiquement sur une interprétation, une signification et donc une esquisse de savoir général, admis provisoirement comme une réalité indépendante, platonicienne. Car, comme le dit plus loin L. Sève : « il y a toujours de l'universel dans le singulier » (p. 200). Pour ce faire, nous construisons un modèle mathématique qui permet la comparaison statistique du contingent et du théorique aux ordres du modèle. Au sujet de son application possible en biologie, Jean-Claude Ameisen déclare dans « Sur les épaules de Darwin, tome 2» (p. 163) : « *A la recherche des régularités, des mécanismes et des relations de causalité qui rendent compte des formes et des comportements du monde vivant, la formalisation mathématique jouera un rôle de plus en plus important en biologie* ».

Plus précisément, plaçons-nous donc dans le cadre de la théorie de l'Analyse Statistique Implicative, où l'on croise des sujets ou des objets avec des variables de natures variées. Restreignons-nous, pour le moment, au cas de variables binaires telles que a et b. La problématique majeure consiste en la recherche d'une mesure qui permettrait de quantifier la règle quasi-implicative de a sur b. Cette mesure, en ASI, se fonde sur le nombre de contre-exemples à l'implication, nombre qui doit être le plus petit possible eu égard aux cardinaux des sous-ensembles de sujets vérifiant respectivement a et b (voir « *L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités* » [2013]) . Nous nous intéresserons à l'indice que retient l'ASI pour signifier ou non si la qualité de la règle est statistiquement acceptable et, qu'en conséquence, si la tendance à la règle implicative a de bonnes raisons d'être ou non retenue en étant *étonnamment* acceptable[2] dans le cadre d'une théorie.

[2] C'est aussi ce qu'affirme René Thom (« Paraboles et catastrophes », 1980, p.130) : « …le problème

Nous examinerons les propriétés de sensibilité de cet indice, dit d'implication, particulièrement aux variations des cardinaux en jeu dans des expériences d'extraction de règles de corpus où les variables sont instanciées sur l'ensemble des sujets. On s'intéressera en particulier au gradient de cet indice significatif de la « vitesse et de l'amplitude » de la croissance ou la décroissance de ces cardinaux. Nous étudierons le champ de vecteurs engendré par les observations de l'indice d'implication et le champ de gradient qui lui est associé. Cette approche diffère donc de celle adoptée par Eugen Barbu dans son mémoire de D.E.A.(Barbu E., 2003) intitulé « Hiérarchie cohésitive » qui, à travers un bootstrap appliqué à des mesures de qualité de règles d'association, a fait varier les paramètres de ces mesures dont celle qui fonde l'implication statistique.

Des modélisations différentes de l'ASI pour la recherche de règles quasi-implicatives se retrouvent sous la forme de treillis de Galois et de réseau Bayesien. Un algorithme dit « A priori » est le plus souvent à la base de la mesure de qualité de règles implicatives. Nous en avons comparé les propriétés face à l'ASI tant sur le plan de leur sensibilité aux variations des instances des variables que sur la représentation des structures des ensembles de règles obtenus. Nous y revenons rapidement dans le § 4. C'est donc plus sur ces phénomènes que nous porterons notre attention et non pas, *dans ce texte*, sur la représentation des règles et méta-règles. Nous rappelons ci-dessous, le processus probabiliste qui en permet l'extraction.

2 Rappel des fondements de l'A.S.I. (Gras, 1979)

Rappelons les fondements premiers de l'ASI dont l'objectif est de rechercher la plausibilité causale derrière une relation quasi-implicative là où le physicien philosophe Bernard d'Espagnat (B.d'Espagnat, 1981) parlerait d'**influence** en ces termes : *« les évènements A influencent appréciablement les évènements B si et seulement si la fréquence avec laquelle les évènements B ont lieu est (appréciablement) différente selon que l'on impose on non aux évènements A d'exister* » (p. 186). Reprenant cette perspective, nous mathématisons cette définition en quantifiant l'expression floue «appréciablement », non pas par une fréquence conditionnelle comme il est fait dans (Agrawal R. et al, 1993) et ses dérivés, par exemple dans les réseaux bayesiens, mais de la façon suivante.

Notons A et B les sous-ensembles respectifs de E d'individus qui vérifient respectivement les variables booléennes a et b (Fig. 1). Depuis 1979, nous avons étendu (cf. Ouvrages cités en référence), de façon appropriée, les propriétés qui suivent à des variables réelles continues ou non.

Pour une règle quelconque $a \rightarrow b$, observée dans E, l'ASI consiste à comparer le nombre de contre-exemples $n_{a \wedge \bar{b}}$ ($\bar{b}$ est la négation de b) à cette règle observés dans l'intersection $A \cap \bar{B}$ ($\bar{B}$ est le complémentaire de B dans E) avec le nombre de contre-exemples qui apparaîtraient lors d'un choix aléatoire et indépendant de deux parties X et Y de E de mêmes cardinaux respectifs que A et B (Fig. 1) (Gras, 1979 ; Lebart et al.,

n'est pas de décrire la réalité, le problème consiste bien plus à repérer en elle ce qui a de sens pour nous, ce qui est surprenant dans l'ensemble des faits. Si les faits ne nous surprennent pas, ils n'apportent aucun élément nouveau pour la compréhension de l'univers : autant donc les ignorer » et plus loin : « … ce qui n'est pas possible si l'on ne dispose pas déjà d'une théorie ».

2006). La variable aléatoire associée est notée $N_{a\wedge\bar{b}}$. Le principe fondamental que nous retenons consiste à comparer la contingence à la théorie par une méthodologie qui relève de la philosophie de l'expérience où sont premières les observations (cf. B. d'Espagnat, 1981, cité plus haut).

La qualité de la règle sera intuitivement d'autant meilleure que Prob[$N_{a\wedge\bar{b}} > n_{a\wedge\bar{b}}$] sera proche de 1 : autrement dit, dans ce cas, on y observe plus de contre-exemples dans des circonstances aléatoires que l'on en a observés dans la contingence, sous l'hypothèse a priori d'indépendances des variables a et b. Dans ce cas, le seul hasard conduit donc, en moyenne, à plus de contre-exemples que ce qui est observé. C'est ainsi que se définit l'ASI en falsifiant la relation implicative par la mesure relative de sa négation exprimée à travers ses contre-exemples.

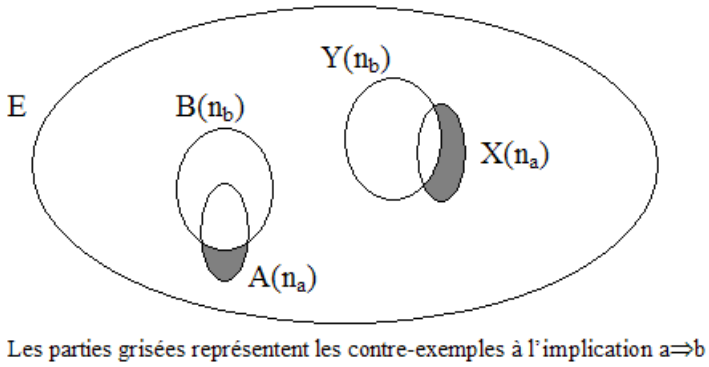


Figure 1. Représentation ensembliste

La méthode de tirage au hasard de X et Y, dans une hypothèse a priori d'indépendance de *a* et *b*, conduit à différentes options pour la loi de la variable aléatoire $N_{a\wedge\bar{b}}$. Deux modélisations de cette variable sont généralement retenues en ASI conduisant à un modèle de Poisson et à un modèle binomial (Gras et al 2009). On centre et on réduit cette variable en la variable $Q(a,\bar{b})$; l'observation contingente (empirique), sa réalisation, est $q(a,\bar{b})$. Par exemple, dans le cas du modèle de Poisson, on obtient l'indice de base pour des variables binaires :

$$Q(a,\bar{b}) = \frac{\mathrm{Card}(X\cap\bar{Y}) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \text{ alors que } q(a,\bar{b}) = \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} \text{ est sa réalisation contingente}$$

L'intensité d'implication est alors définie par $\varphi(a,b)$ = Prob[$Q(a,\bar{b})>q(a,\bar{b})$], dont la valeur gaussienne asymptotique, centrée et réduite est :

$$\varphi(a,b) = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{+\infty} e^{-\frac{t^2}{2}} dt \quad (1)$$

Cette définition (1) de l'intensité d'implication, rappelle au chercheur qu'elle ne présente un intérêt implicatif qu'à condition qu'elle soit supérieure à 0.50, c'est-à-dire

que $q(a,\bar{b})$ soit négatif. Plus $q(a,\bar{b})$ est négatif, meilleure est la qualité de la règle $a \rightarrow b$. Cependant, si $n_b = n$, nous définissons l'intensité d'implication par : $\varphi(a,b) = 0$ qui n'est pas le prolongement par continuité de l'intensité définie par (1) alors que $q(a,\bar{b})$ n'est plus défini pour cette valeur de n_b. Lever son indétermination reviendrait à lui attribuer la valeur 0, et donc $\varphi(a,b) = 0.5$, incompatible avec la sémantique de l'ASI. Cette exception tient au respect de celle-ci puisque l'implication de a sur b est alors devenue triviale, tautologique, donc non informative.

Dans cet article, nous nous intéressons aux propriétés de la fonction q(.) comme fonction des 4 occurrences observées, c'est-à-dire les variables cardinales n, n_a, n_b, $n_{a\wedge\bar{b}}$. Quel rôle jouent ces variations sur celles de l'intensité d'implication qui dépend de q(.) par intégration gaussienne ? Quelle sensibilité observe-t-on lorsque les occurrences varient quelque peu ? En quoi le gradient de q permet-il de définir un champ de gradient sur l'espace des couples de variables ? Nous sommes persuadés que la perception d'un graphe implicatif, passe certes par l'examen de sa structure qui comme l'énergie est de type « intégrale », mais aussi qu'elle se révèle à travers des procédures de dérivation. Sans mésestimer le premier que nous explorerons dans une modélisation originale mécanique de l'ASI, nous consacrerons l'étude qui vient à une approche essentiellement infinitésimale.

3 Variations de l'indice d'implication q en fonction des 4 occurrences

3.1 Stabilité de l'indice d'implication

Étudier la stabilité de l'indice d'implication q, revient à examiner ses petites variations au voisinage des 4 valeurs entières observées (n, n_a, n_b, $n_{a\wedge\bar{b}}$). Pour ce faire, il est possible d'effectuer différentes simulations en croisant ces 4 variables entières dont q dépend (Gras et al., 2013). Mais, considérons ces variables comme variables à valeurs réelles et q comme une fonction continûment différentiable par rapport à ces variables, elles-mêmes contraintes à respecter les inégalités : $0 \leq n_a \leq n_b$ et $n_{a\wedge\bar{b}} \leq \inf\{n_a, n_b\}$ et $\sup\{n_a, n_b\} \leq n$. La fonction q définit alors un champ scalaire et vectoriel sur R^4 en tant qu'espace affine et vectoriel sur lui-même. Dans l'hypothèse vraisemblable d'une évolution d'un processus non chaotique du recueil de données, il suffit alors d'examiner la différentielle de q par rapport à ces variables et d'en conserver la restriction aux valeurs entières des paramètres de la relation $a \Rightarrow b$. La différentielle de q, au sens de la topologie de Fréchet[3], s'exprime de la façon suivante par un produit scalaire :

$$dq = \frac{\partial q}{\partial n} dn + \frac{\partial q}{\partial n_a} dn_a + \frac{\partial q}{\partial n_b} dn_b + \frac{\partial q}{\partial n_{a\wedge\bar{b}}} dn_{a\wedge\bar{b}} = \text{grad q. dM} \quad (2)^4$$

où M est le point de coordonnées (n, n_a, n_b, $n_{a\wedge\bar{b}}$) du champ scalaire C de vecteurs,

[3] La topologie de Fréchet admet comme base de filtres des sections de N, soit des sous-ensembles de naturels de la forme {n, n+1, n+2, ….} alors que la topologie usuelle sur R admet pour filtres des intervalles de réels. Ainsi continuité et dérivabilité sont des concepts parfaitement définis et opératoires selon la topologie de Fréchet au même titre qu'ils le sont avec la topologie usuelle.

[4] Par une métaphore mécaniste, on dira que dq est le travail élémentaire de q pour un déplacement dM.

dM est le vecteur de composantes les accroissements différentiels de ces variables d'occurrences,

et grad q le vecteur de composantes les dérivées partielles de ces variables occurrences.

La différentielle de la fonction q apparaît donc comme le produit scalaire de son gradient et de l'accroissement de q sur la surface représentant les variations de la fonction q(n, n_a, n_b, $n_{a\wedge\bar{b}}$). Ainsi, le gradient de q représente ses propres variations en fonction de celles de ses composantes, les 4 cardinaux des ensembles E, A, B et $A \cap \bar{B}$. Il indique la direction et le sens de croissance ou la décroissance de q dans l'espace de dimension 4. Rappelons qu'il est porté par la normale à la surface de niveau q = cte.

Si l'on veut étudier comment varie q en fonction de $n_{\bar{b}}$, il suffit de remplacer n_b par n-n_b et donc changer le signe de la dérivée de n_b dans la dérivée partielle. En fait, l'intérêt de cette différentielle réside dans l'estimation de l'accroissement (positif ou négatif) de q que nous notons Δq par rapport aux variations respectives Δn, Δn_a, Δn_b, $\Delta n_{\bar{b}}$ et $\Delta n_{a\wedge\bar{b}}$. On a donc :

$$\Delta q = \frac{\partial q}{\partial n}\Delta n + \frac{\partial q}{\partial n_a}\Delta n_a + \frac{\partial q}{\partial n_b}\Delta n_b + \frac{\partial q}{\partial n_{a\wedge\bar{b}}}\Delta n_{a\wedge\bar{b}} + o(\Delta q)$$

où o(Δq) est un infiniment petit du 1er ordre.

Examinons les dérivées partielles de n_b et le nombre de contre-exemples $n_{a\wedge\bar{b}}$. On obtient :

$$\frac{\partial q}{\partial n_b} = \frac{1}{2} n_{a\wedge\bar{b}} \left(\frac{n_a}{n}\right)^{-\frac{1}{2}} (n - n_b)^{-\frac{3}{2}} + \frac{1}{2}\left(\frac{n_a}{n}\right)^{\frac{1}{2}} (n - n_b)^{-\frac{1}{2}} > 0 \ (3)$$

$$\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = \frac{1}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{1}{\sqrt{\frac{n_a (n - n_b)}{n}}} > 0 \ (4)$$

Ainsi, si les accroissements Δn_b et $\Delta n_{a\wedge\bar{b}}$ sont positifs, l'accroissement de $q(a,\bar{b})$ est également positif. Ceci s'interprète ainsi : si le nombre d'exemples de b et celui des contre-exemples de l'implication augmentent alors l'intensité d'implication diminue pour n et n_a constants. Autrement dit, cette intensité d'implication est maximum aux valeurs observées n_b et $n_{a\wedge\bar{b}}$ et minimum aux valeurs n_b+Δn_b. et $n_{a\wedge\bar{b}}$+$\Delta n_{a\wedge\bar{b}}$.

Si nous examinons le cas où n_a varie, nous obtenons la dérivée partielle de q par rapport à n_a qui est :

$$\frac{\partial q}{\partial n_a} = -\frac{1}{2}\frac{n_{a\wedge\bar{b}}}{\sqrt{n_{\bar{b}}/n}} . \left(\frac{n}{n_a}\right)^{\frac{3}{2}} - \frac{1}{2}\sqrt{\frac{n_{\bar{b}}}{n_a}} < 0 \ (5)$$

Ainsi, pour des variations de n_a sur [0, n_b], la fonction indice d'implication $q(a,\bar{b})$ est toujours décroissante (et concave) par rapport à n_a et est donc minimum pour $n_a = n_b$. Par suite, l'intensité d'implication y est croissante et maximum pour $n_a = n_b$.

Notons la dérivée partielle de q par rapport à n :

$$\frac{\partial q}{\partial n} = \frac{1}{2\sqrt{n}}[n_{a\wedge\bar{b}} + \frac{n_a n_{\bar{b}}}{n}]$$

En conséquence, si les 3 autres paramètres sont constants, l'indice d'implication décroît en $\sqrt{n}$. La qualité de l'implication n'en est donc que meilleure, propriété spécifique de l'ASI par rapport à d'autres indicateurs retenus dans la littérature (cf. Gras et Couturier, 2010). Cette propriété est en accord avec les attentes statistiques et sémantiques relatives au crédit accordé à la fréquence des observations. Les dérivées partielles de q (au moins l'une d'entre elles) étant non linéaires selon les paramètres variables en jeu, on a affaire à un système dynamique non linéaire avec toutes les conséquences épistémologiques que nous envisagerons par ailleurs.

3.2 Exemple numérique

Dans une première expérience, on observe les occurrences :

$n = 100$, $n_a = 20$, $n_b = 40$ (d'où $n_{\bar{b}} = 60$), $n_{a\wedge\bar{b}} = 4$.

L'application de la formule (1) donne $q(a,\bar{b}) = -2{,}309$

Dans une 2ème expérience, n et n_a sont inchangées mais les occurrences des b et des contre-exemples $n_{a\wedge\bar{b}}$ s'accroissent d'une unité.

Au point initial de l'espace des 4 variables, les dérivées partielles qui seules nous intéressent (selon n_b et $n_{a\wedge\bar{b}}$) ont respectivement pour valeurs en appliquant les formules (3) et (4)) : $\frac{\partial q}{\partial n_b} = 0{,}0385$ et $\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = 0{,}2887$

Comme Δn_b, $\Delta n_{\bar{b}}$ et $\Delta n_{a\wedge\bar{b}}$ sont égaux à 1, -1 et 1, l, alors Δq est égal à :

$0{,}0385 + 0.2887 + o(\Delta q) = 0{,}3272 + o(\Delta q)$ et la valeur approchée de q lors de la deuxième expérience est $-2{,}309 + 0{,}2887 + o(\Delta q) = -1{,}982 + o(\Delta q)$ en utilisant le développement de q au premier ordre (formule (2)).

Or le calcul du nouvel indice d'implication q au point de la 2ème expérience est, par l'usage de (1) : -1,9795, valeur bien approchée par le développement de q.

3.3 Une première relation différentielle de φ en tant que fonction de la fonction q.

Considérons l'intensité d'implication φ comme fonction de $q(a,\bar{b})$:

$$\varphi(q) = \frac{1}{\sqrt{2\pi}}\int_q^{\infty} e^{-t^2/2}dt$$

On peut alors examiner comment φ(q) varie lorsque q varie au voisinage d'une valeur donnée (a,b), sachant comment q varie lui-même en fonction des 4 paramètres qui le déterminent. Par dérivation de la borne d'intégration, on obtient :

$$\frac{d\varphi}{dq} = -\frac{1}{\sqrt{2\pi}} e^{\frac{-q^2}{2}} < 0 \quad (6)$$

Ce qui confirme bien que l'intensité croît lorsque q décroît, mais la vitesse de croissance est précisée par la formule, ce qui permet d'étudier avec plus de précision les variations de φ. Puisque la dérivée de φ par rapport à q est toujours négative, la fonction φ est décroissante.

3.4 Exemple numérique

Reprenant les valeurs des occurrences observées dans les 2 expériences évoquées plus haut, on trouve pour q = -2,309, la valeur de l'intensité d'implication φ(q) est égale à 0,992. Appliquant la formule (6), la dérivée de φ par rapport à q est : - 0 ,02775 et l'accroissement négatif de l'intensité est alors : - 0,02775 . Δq = - 0,02775. 0,3272. L'intensité approchée au premier ordre est donc : 0, 992 - Δq soit 0,983. Or le calcul réel de cette intensité est, pour q= -1,9795 , φ(q) = 0, 976

4 Examen d'autres indices

Contrairement à l'indice de base q et l'intensité d'implication qui mesure la qualité à travers une probabilité (cf. définition 3), les autres indices les plus courants se veulent eux-mêmes directement des mesures de qualité. Nous examinerons leurs sensibilités respectives aux variations des paramètres retenus dans la définition de ces indices. Nous conservons les notations adoptées au paragraphe 2 et choisissons des indices qui sont rappelés dans (Gras et al, 2004), (Lenca et al., 2005) et (Gras et Couturier, 2010) .

4.1 L'indice de Loevinger

C'est un « ancêtre » des indices d'implication (Loevinger, 1947). Cet indice, noté H(a,b) varie de 1 à - ∞. Il est défini par :

$$H(a,b) = 1 - \frac{n n_{a \wedge \bar{b}}}{n_a n_{\bar{b}}}.$$

Sa dérivée partielle par rapport à la variable nombre de contre-exemples est donc :

$$\frac{\partial H}{\partial n_{a \wedge \bar{b}}} = -\frac{n}{n_a\, n_{\bar{b}}}$$

Ainsi l'indice d'implication est toujours décroissant avec $n_{a \wedge \bar{b}}$. S'il est "proche" de 1, l'implication est "presque" satisfaite. Mais cet indice présente l'inconvénient, ne se référant pas à une échelle de probabilité, de ne pas fournir de seuil de vraisemblance et d'être invariant dans toute dilatation de E, A , B et $A \cap \bar{B}$.

4.2 L'indice Lift

Il s'exprime par : $l = \frac{n.n_{a\wedge b}}{n_a n_b}$. Cette expression, linéaire par rapport aux exemples, peut encore s'écrire pour mettre en évidence le nombre de contre-exemples :

$$l = \frac{n.(n_a - n_{a\wedge\bar{b}})}{n_a.n_b}.$$

Pour étudier la sensibilité de l aux variations des paramètres, nous formons :

$$\frac{\partial l}{\partial n_{a\wedge\bar{b}}} = -\frac{1}{n_a.n_b}$$

Ainsi, la variation de l'indice Lift est indépendante de celle du nombre de contre-exemples. C'est une constante qui ne dépend que des variations des occurrences de a et de b. l décroît donc lorsque le nombre de contre-exemples croît, ce qui sémantiquement, est acceptable mais la vitesse de décroissance ne dépend pas de la vitesse de croissance de $n_{a\wedge\bar{b}}$.

4.3 L'indice MC

Il s'exprime ainsi : $m = \frac{n_a - n_{a\wedge\bar{b}}}{n_b.n_{a\wedge\bar{b}}} n_{\bar{b}}$. Remarquons qu'en étant indépendant de n, il n'a pas de sens statistique aussi intéressant.

Sa dérivé partielle par rapport au nombre de contre-exemples est :

$$\frac{\partial m}{\partial n_{a\wedge\bar{b}}} = -\frac{n_a.n_{\bar{b}}}{n_b}.(\frac{1}{n_{a\wedge\bar{b}}})^2.$$

L'indice m décroît donc lorsque $n_{a\wedge\bar{b}}$ croît et la vitesse de décroissance est même plus rapide qu'avec le Lift et qu'avec l'indice q gaussien basique pour calculer l'intensité d'implication. Il ne résiste pas à l'instabilité du nombre de contre-exemples.

4.4 La confiance

Cet indice est le plus connu et le plus utilisé grâce à la caisse de résonance dont dispose une publication anglo-saxonne (Agrawal et al. 1993). Il est à l'origine de plusieurs autres indices communément employés qui n'en sont que des variantes satisfaisant telle ou telle exigence sémantique.. De plus, il est simple et s'interprète aisément et immédiatement.

$$c = \frac{n_{a\wedge b}}{n_a} \text{ ou } 1 - \frac{n_{a\wedge\bar{b}}}{n_a}$$

La première forme, linéaire par rapport aux exemples, indépendante de n_b, s'interprète comme une fréquence conditionnelle des exemples de b quand a est connu.

La sensibilité de cet indice aux variations des occurrences des contre-exemples se lit à travers la dérivée partielle :

$$\frac{\partial c}{\partial n_{a\wedge\bar{b}}} = -\frac{1}{n_a}$$

Par conséquent, la confiance croît quand $n_{a\wedge\bar{b}}$ décroît, ce qui est sémantiquement acceptable, mais la vitesse de variation est constante, indépendante de la vitesse de décroissance de ce nombre, des variations de n et de n_b. Le gradient de c ne s'exprime que par rapport à $n_{a\wedge\bar{b}}$ et à n_a :

$$\begin{pmatrix} -\frac{1}{n_a} \\ \frac{n_{a\wedge\bar{b}}}{{n_a}^2} \end{pmatrix}$$

Ceci peut apparaître comme une restriction du rôle des paramètres dans l'expression de la sensibilité de l'indice.

5 Champ de gradient, champ implicatif

5.1 Existence d'un champ de gradient

Nous revenons à l'indice implicatif $q(a,\bar{b})$. Considérons l'espace E de dimension 4 où les points M ont pour coordonnées les paramètres relatifs aux variables binaires a et b, soit (n, n_a, n_b, $n_{a\wedge\bar{b}}$). $q(a,\bar{b})$ définit donc un champ scalaire en tant qu'application de R^4 dans R (plongement de N^4 dans R^4).

Pour que le vecteur grad q de composantes les dérivées partielles de q par rapport aux variables n, n_a, n_b, $n_{a\wedge\bar{b}}$ définisse un champ de gradient - champ vectoriel particulier que nous appellerons aussi champ implicatif- il doit respecter le critère de Schwartz d'une différentielle totale exacte à savoir et par exemple :

$$\frac{\delta}{\delta n_{a\wedge\bar{b}}}\left(\frac{\delta q}{\delta n_b}\right) = \frac{\delta}{\delta n_b}\left(\frac{\delta q}{\delta n_{a\wedge\bar{b}}}\right)$$

et idem pour les autres variables prises deux à deux. Or on a bien par les formules (3) et (4) :

$$\frac{\delta}{\delta n_{a\wedge\bar{b}}}\left(\frac{\delta q}{\delta n_b}\right) = \frac{1}{2}\left(\frac{n_a}{n}\right)^{-1/2}\left(\frac{n_{\bar{b}}}{n}\right)^{-3/2} = \frac{\delta}{n_b}\left(\frac{\delta q}{\delta n_{a\wedge\bar{b}}}\right)$$

Ainsi, au champ de vecteurs C = (n, n_a, n_b, $n_{a\wedge\bar{b}}$) de E, dont nous préciserons la nature, correspond un champ de gradient G qui est dit dérivé du **potentiel** q. Le gradient grad q est donc le vecteur qui représente la variation spatiale de l'intensité du champ. Il est dirigé des faibles valeurs du champ aux valeurs plus élevées. En suivant le gradient en chaque point, on suit l'augmentation de l'intensité d'implication du champ dans l'espace et, en quelque sorte la vitesse avec laquelle elle change sous l'effet de la variation d'un ou plusieurs paramètres.

Par exemple, si l'on fixe 3 des paramètres n, n_a, n_b, $n_{a\wedge\bar{b}}$ donnés par la réalisation du couple (a,b), le gradient est un vecteur dont la direction indique la croissance ou la

décroissance de q, donc la décroissance ou la croissance de |q| et par suite de φ en fonction des variations du 4ème paramètre. Nous l'avions indiqué plus haut en interprétant la formule (5).

5.2 Lignes de niveau ou équipotentielles

Une ligne ou une surface équipotentielle (ou de niveau) dans le champ C est une courbe de E le long de laquelle ou sur laquelle un point variable M conserve la même valeur du potentiel q (par ex. lignes isothermiques sur le globe ou lignes de niveau d'une carte IGN). L'équation de cette surface[5] est, bien entendu :

$$q(a,\bar{b}) - \frac{n_{a\wedge\bar{b}} - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = 0$$

Par suite, sur une telle courbe, le produit scalaire grad q. dM est nul. Ce qui s'interprète comme indiquant l'orthogonalité du gradient avec la tangente ou l'hyperplan tangent à la courbe, c'est-à-dire avec la ligne ou la surface équipotentielle. Dans une interprétation cinématique de notre problème, la vitesse de parcours de M sur la surface équipotentielle est orthogonale au gradient en M.

A titre d'illustration, relativement à un potentiel F dépendant de 2 variables seulement, la figure ci-dessous par exemple montre la direction orthogonale du gradient par rapport aux différentes surfaces équipotentielles le long desquelles le potentiel F ne varie pas mais passe de F=7 à F= 10.

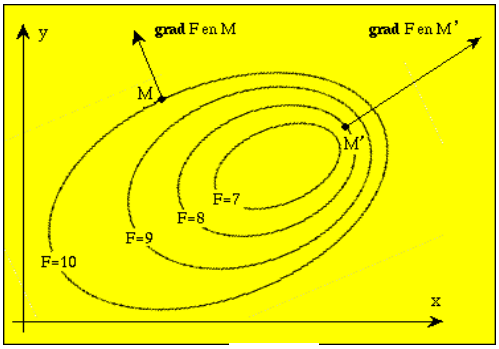


Figure 2

Il est possible dans le cas du potentiel q, de construire des surfaces équipotentielles comme ci-dessus (à deux dimensions pour la facilité de représentation). On peut comprendre que plus le champ est intense plus les surfaces sont serrées. Pour une valeur de q donnée, dans ce cas, on fixe 3 variables, par exemple n, n_a, n_b et une valeur de q compatibles avec les contraintes du champ. Soit : $n = 10^4$; $n_a = 1600 \leq n_b = 3600$ et q =

[5] En géométrie différentielle, on dirait que cette surface est une variété (quasi)différentiable à bord, compacte, homéomorphe au pavé fermé des intervalles de variations des 4 paramètres. Notons que le point dont la composante n_b est égale à n (donc $n_{a\wedge\bar{b}} = 0$) est un point singulier (« catastrophique » au sens de René Thom) de la surface et q, le potentiel, n'est pas différentiable en ce point. Partout ailleurs, la surface est différentiable, les points sont tous réguliers. Si le temps, par exemple, paramètre les observations du processus dont (n, n_a, n_b, $n_{a\wedge\bar{b}}$) est une réalisation, à chaque instant correspond une fibre morphologique du processus représentée par une telle surface dans l'espace-temps..

-2 soit |q| = 2. On trouve alors $n_{a\wedge\bar{b}}$ = 528 en utilisant la formule (1). Mais les points (10^4, 1600, 5100, 728) et (100, 25, 64, 3) appartiennent également à cette surface et la même courbe équipotentielle. Le point (10^4, 1600, 3600, 928) appartient à la courbe équipotentielle q=-3). En fait, sur toute cette surface, on obtient une sorte d'homéostasie de l'intensité d'implication. L'expression de la fonction q de la variable $n_{a\wedge\bar{b}}$ permet de montrer qu'elle est convexe. Cette propriété prouve que le segment des points $t.M_1 + (1-t).M_2$, pour $t \in [0,1]$ qui joint deux points M_1 et M_2 de la même ligne équipotentielle est entièrement contenu dans sa convexité. La figure ci-dessous représente, dans le champ implicatif deux surfaces équipotentielles voisines Σ_1 et Σ_2 correspondant à deux valeurs du potentiel q_1 et q_2. Au point M_1 le champ scalaire prend donc la valeur q_1. M_2 est l'intersection de la normale issue de M_1 avec Σ_2. Etant donné la direction du vecteur normal $\vec{n}$, la différence $\Delta = q_2 - q_1$, variation du champ quand on passe de Σ_1à Σ_2 est alors égale à l'opposé de la norme du gradient de q en M_1 soit $\frac{\partial q}{\partial n}$, si n_a, n_b et $n_{a\wedge\bar{b}}$ sont fixés .

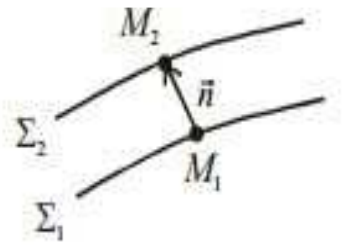


Figure 3

Ainsi, l'espace E peut être comme feuilleté par des surfaces équipotentielles correspondant à des valeurs successives de q relativement aux cardinaux (n, n_a, n_b, $n_{a\wedge\bar{b}}$) que l'on ferait varier. Cette situation correspond à celle qui est envisagée dans la modélisation de l'ASI. Fixant n, n_a et n_b, on considère les ensembles aléatoires X et Y de mêmes cardinaux que A (n_a) et B (n_b) et dont le cardinal de $X \cap \bar{Y}$ suit une loi de Poisson ou une loi binomiale, suivant le choix du modèle. Les différents champs de gradient, véritables « lignes de force », qui leur sont associés sont orthogonaux aux surfaces définies par les valeurs correspondantes de Q. Ceci nous évoque, dans le cadre théorique du potentiel, la métaphore prémonitoire de « flux implicatif » que nous avons exprimée dans (Gras et al. 1996). Derrière cette notion nous pouvons imaginer un transport d'information d'intensité variable dans un univers causal. Nous illustrons cette métaphore avec l'étude des propriétés du cône implicatif à deux nappes (cf. Lahanier-Reuter et al, en publication ASI 8). De plus et intuitivement, l'implication a=>b est d'autant de bonne qualité que la surface équipotentielle C de la contingence recouvre des surfaces équipotentielles aléatoires dépendant de la variable aléatoire $Q(a, \bar{b})$.

Rappelons la relation qui unit le potentiel q à l'intensité d'implication $\varphi(a,b)$ définie par:

$$\varphi(a,b)= \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

Remarque 1

On constate que l'intensité est également invariante sur toute surface équipotentielle de ses propres variations. Les portions de surfaces engendrées par q et par φ sont même en correspondance biunivoque. En termes intuitifs, on peut affirmer que lorsque l'une « enfle » l'autre se « dégonfle.

Remarque 2

Notons une fois encore une particularité de l'intensité d'implication. Alors que les surfaces engendrées par les variations des 4 paramètres des données ne sont pas invariantes par une même dilatation des paramètres, celles associées aux indices citées dans le § 4 sont invariantes et présentent une même forme géométrique indifférenciée.

6 Quelques simulations

6.1 Surfaces équipotentielles

Afin de soutenir l'intuition et l'imagination, nous avons effectué quelques simulations de surfaces équipotentielles. Pour obtenir les figures suivantes avec Matlab, nous avons fixé n = 100, utilisé 3 valeurs de $n_{a\wedge\bar{b}}$: 5, 10 et 20 et avons fait varier n_a selon le 1er axe (dit des « x ») de $n_{a\wedge\bar{b}}$ à n et n_b selon le 2ème axe (dit des « y ») de 1 à $n - n_{a\wedge\bar{b}}$. Les courbes représentent donc les variations de q, indice d'implication selon l'axe vertical (dit des « z »). [6]

Sur ce type de figure en 3D, les valeurs affichées sur les axes, pour des commodités de représentation, peuvent être trompeuses. Sur la première par exemple les valeurs de n_a vont de 0 à 100 alors que n_a n'est bien entendu défini que de 5 à 100 puisque $n_{a\wedge\bar{b}}$ = 5. De même, l'examen utile de la qualité de la règle a => b exige $n_a \leq n_b$.

Les couleurs sont choisies pour signifier une certaine qualité de la règle :

- couleurs chaudes (rouge) pour les situations intéressantes où l'implication est satisfaite et d'autant plus que q est négatif,
- couleurs froides (bleu) où c'est le rejet de la règle a => b, voire la validité de la réciproque qui sont de plus en plus satisfaits et soulignés par l'intensité bleue des valeurs de q.

[6] Faisons une place à l'imagination : dans le cas présent où les variables et les paramètres associés ne prennent que des valeurs entières les points représentant les valeurs de q apparaissent comme des « petits pois » sur la « robe » équipotentielle.

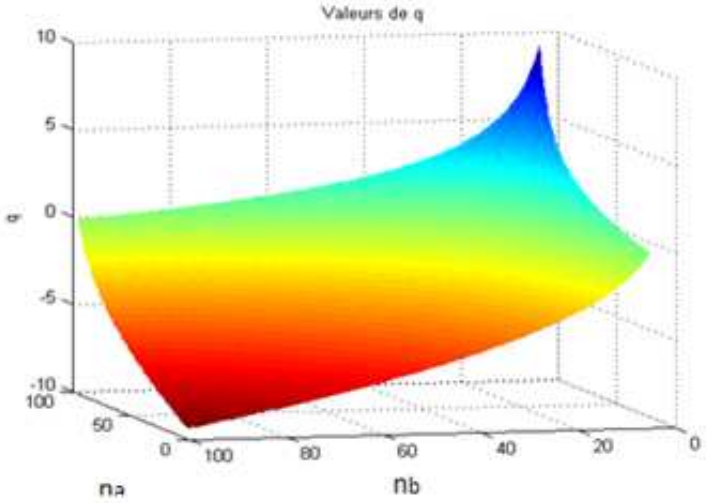


Figure 4 **Courbe 1** : $n_{a\wedge\bar{b}} = 5$

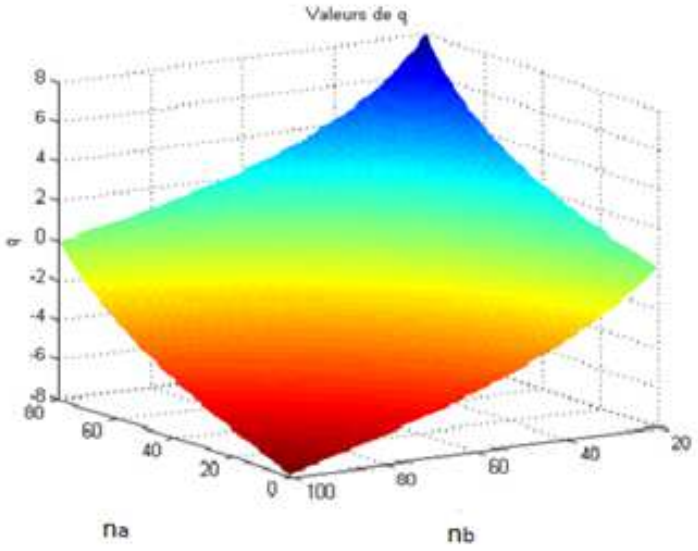


Figure 5 **Courbe 2** : $n_{a\wedge\bar{b}} = 20$

Remarquons l'étendue moins importante de la zone où l'implication est acceptable et, par contre, celle des rejets qui s'est amplifiée ainsi qu'un étalement plus marqué des valeurs indécises pour la règle.

6.2 Courbes équipotentielles de q pour différentes valeurs des contre-exemples

Courbes pour $n_{a\wedge\bar{b}} = 5$

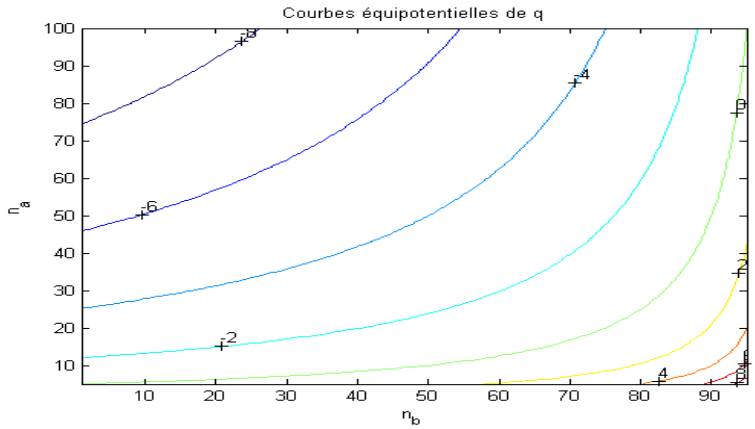


Figure 6

Ces courbes (n_b en abscisse et n_a en ordonnée) sont des projections planes des surfaces précédentes et limitées aux valeurs entières de q de -8 à 0 dans la figure 5 et de -6 à 0 pour la figure 6.

L'indice q est invariant tout le long de chaque arc de courbe. Par exemple, on trouve pour n_a = 45 et n_b = 45 (soit $n_{\bar{b}}$ = 55), q = -3.9 (approximativement courbe q = -4). Mais aussi pour n_a = 25 et n_b = 55 (soit $n_{\bar{b}}$ = 45), q = - 1.86.01$\cong -2$.

On peut remarquer que, pour des valeurs de n_a et de $n_{\bar{b}}$ assez élevées (donc n_b petit) les valeurs prises par q restent fortes plus « longtemps » que pour des valeurs opposées, ce que montre le resserrement des courbes pour des valeurs de q faibles conduisant à des incertitudes au sujet de la règle en raison de la décroissance de q.

Comparaison pour $n_{a\wedge\bar{b}}$ = 5 (courbes rouges) et $n_{a\wedge\bar{b}}$ = 10 (courbes vertes)

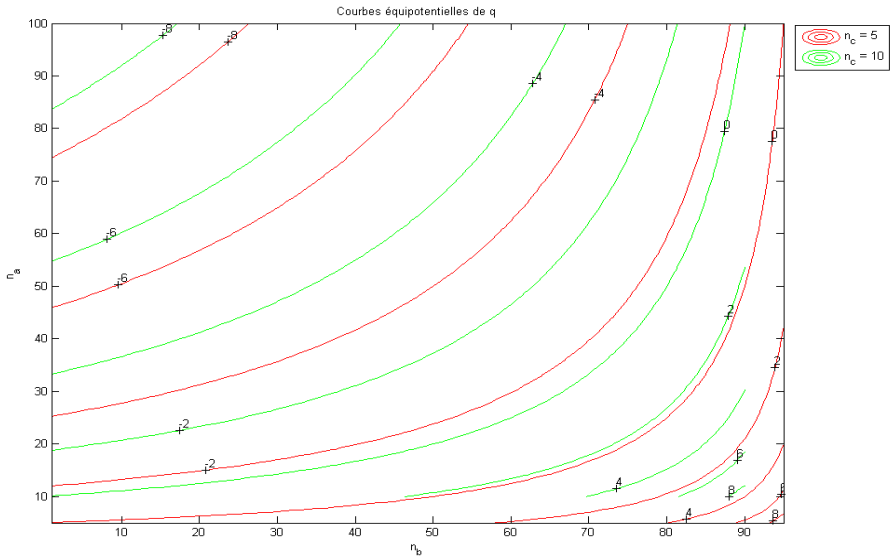


Figure 8

On notera le décalage des lignes équipotentielles rouges par rapport aux vertes où, pour une même valeur du couple (n_a, n_b), q est plus négatif sur les rouges que sur les vertes, donc accompagné d'une intensité d'implication de meilleure qualité.

Mise en évidence des vecteurs « gradient » du champ implicatif.

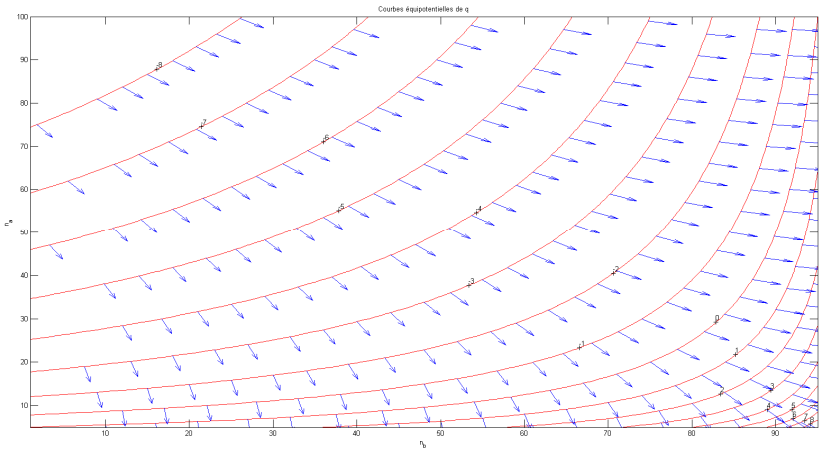


Figure 9

Si l'on fait maintenant apparaître les vecteurs « gradient » du champ implicatif normalisés et normaux aux lignes équipotentielles, on constate, métaphoriquement, une sorte de « respiration » du phénomène lié aux changements de valeurs des paramètres de la contingence. L'expansion de q, donc de l'intensité d'implication, est bien appréhendée par le sens du vecteur comme le laissaient prévoir les signes des dérivées partielles de q par rapport à n_a (négatif) et à n_b (positif). Plus les paramètres croissent, pour un même nombre de contre-exemples, meilleures sont l'intensité et par conséquent notre confiance en la règle. Notons aussi que l'instabilité de q (courbes qui se « resserrent ») va croissante de gauche à droite comme on le remarquait dans le 6.2.

7 Conclusion

Sur la base des concepts qui ont conduit à la modélisation de la notion de quasi-implication en ASI, nous avons étudié l'indice primitif gaussien qui permet de quantifier la qualité de cette notion et, de ce fait, de qualifier notre agrément ou notre rejet des règles extraites d'un corpus de données binaires (et autres par extension). Notre étude s'est bien précisément centrée sur la sensibilité de l'indice d'implication aux variations des paramètres en jeu dans l'énonciation d'une règle. C'est ainsi que nous avons focalisé l'étude sur le champ scalaire défini par cet indice, puis sur un concept d'analyse, le gradient d'une telle fonction. Nous avons alors examiné les structurations de l'espace vectoriel donné par les paramètres, puis dégagé la notion de champ de gradient (dénommé ici champ implicatif) pour illustrer géométriquement, par des surfaces et des lignes équipotentielles, la sensibilité de l'indice d'implication. Ce pas-de-côté géométrique, par rapport aux approches statistiques et analytiques originelles, restitue à l'A.S.I. une vision dynamique du concept qui en enrichit la représentation mentale. A sa charge d'entretenir nos rêves car selon une formule de G. Bachelard *: « On ne peut étudier que ce que l'on a d'abord rêvé ».*

Références

[1.] Agrawal R., Imielinsky T. et Swami A,(1993), Mining association rules between sets of items in large databases, *Proc. of the ACM SIGMOD'93*, 207-216

[2.] Barbu E. (2003), *Hiérarchie cohésitive (ou implicative*), Mémoire de D.E.A. Extraction des Connaissances à partir des Données, Ecole Doctorale Informatique et Information pour la Société, Equipe COD, LINA, Université de Nantes

[3.] Bernard J.-M. et Poitrenaud S, (1999) L'analyse implicative bayesienne d'un questionnaire binaire: quasi-implications et treillis de Galois simplifié", *Mathématiques, Informatique et Sciences Humaines*, n° 147, 25-46.

[4.] Cadot M., (2009), Graphe de règles d'implication statistique pour le raisonnement courant. Comparaison avec les réseaux bayesiens et les treillis de Galois, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, p.223-250*

[5.] David J, Guillet F., Gras R. and Briand H. (2006): Conceptual hierarchies matching : an approach based on discovery of implication rules between concepts, *In Proc. ECAI 2006, 17th European Conference on Artificial Intelligence, IOS Press, Riva del Garda, Italy.*

[6.] D'Espagnat B. (1981), *A la recherche du réel*, Gauthier-Villars, Paris.

[7.] Fayyad U., Piatetsky-Shapiro G. and Smyth P. From Data Mining to Knowledge Discovery. In Advances In Knowledge Discovery and Data Mining, Fayyad U., Piatetsky-Shapiro G., Smyth P, and Uthurusamy R. eds, AAAI/MIT Press, 1-31.

[8.] Gras R., (1979), Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques, Thèse d'Etat, Université de Rennes 1.

[9.] Gras R., Ag Almouloud S., Bailleul M., Larher A., Polo M., Ratsimba-Rajohn et Totohasina A (1996), *L'implication Statistique*, Collection Associée à Recherches en Didactique des Mathématiques, Grenoble : La Pensée Sauvage.

[10.] Gras R., Kuntz P. et Briand H., (2001b), Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données, *Mathématiques et Sciences Humaines*, n° 154-155, 9-29.

[11.] Gras R., Couturier R.., Blanchard J., Briand H., Kuntz P., Peter P., (2004), Quelques critères pour une mesure de qualité de règles d'association. Un exemple : l'implication statistique, *Mesures de qualité pour la fouille de données, RNTI-E-1, Cépaduès –Editions*, 3-32.

[12.] Gras R., David J., Guillet F., Briand H. (2007). Stabilité en A.S.I. de l'intensité d'implication et comparaisons avec d'autres indices de qualité de règles d'association, *Proceedings atelier « Qualité des données et des connaissances », EGC 07*, Namur

[13.] Gras R., Couturier R., (2010) Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association, *Quaderni di Ricerca in Didattica - GRIM (ISSN on-line 1592-4424)*, Eds : J.C. Régnier, R.Gras, F.Spagnolo, B. Di Paola, Université de Palerme, *p.19-57*

[14.] Lenca P., Vaillant B., Meyer P., et Lallich S. (2007), Association Rule Interestingness Measures : Experimental and Theorical Studies, *F.Guillet and H. J.Hamilton eds, Studies in Computational Intelligence 43, Springer*, p. 51-76.

[15.] Loevinger J. (1947), "A systematical approach to the construction and evaluation of tests of ability", *Psychological Monographs*, n° 61, 1947, p. 1-49

[16.] Ritschard G., Marcellin S., Zighed D.A. (2009), Arbre de décision pour données déséquilibrées : sur la complémentarité de l'intensité d'implication et de l'entropie décentrée, *Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse p.207-21.*

[17.] Sève L. (2005), *Emergence, complexité et dialectique*, Odile Jacob, Paris.

[18.] Thom, R (1980). *Paraboles et catastrophes*, Flammarion, Paris.

Ouvrages de référence

L'implication statistique. Nouvelle méthode exploratoire de donnée, sous la direction de R.Gras, et la collaboration de S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn, A.Totohasina, La Pensée Sauvage, Grenoble (1996)

Mesures de Qualité pour la Fouille de Données, H.Briand, M.Sebag, R.Gras er F.Guillet eds, RNTI-E-1, Cépaduès, 2004

Quality Measures in Data Mining , F.Guillet et H.Hamilton eds, Springer, 2007,

Statistical Implicative Analysis, Theory and Applications, R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, eds, Springer, 2008.

Analyse Statistique implicative.Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, 2009.

Teoria y Aplicaciones del Analisis Estadistico Implicativo, Eds : P.Orus, L.Zemora, P.Gregori, Universitat Jaume-1, Castellon (Espagne), ISBN : 978-84-692-3925-4, 2009..

L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire. Eds : J.C. Régnier, Marc Bailleul, Régis Gras, Université de Caen, ISBN : 978-2-7466-5256-9, 2012

L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, 2013.

UN MARIAGE ARRANGE ENTRE L'IMPLICATION ET LA CONFIANCE ?

Régis GRAS[1], Raphaël COUTURIER[2], Pablo GREGORI[3]

AN ARRANGED MARRIAGE BETWEEN IMPLICATION AND CONFIDENCE

RESUME

La plupart des indices d'association entre variables binaires utilisent la fréquence conditionnelle qu'ils appellent confiance ou expressions algébriques d'instanciations pour décider d'une liaison entre deux variables et en apprécier la qualité. En Analyse Statistique Implicative, une autre mesure, l'intensité d'implication, vise le même objectif en se limitant à l'implication tout en s'appuyant plutôt sur la probabilité d'apparition des contre-exemples à ce type de liaison. Dans cet article nous comparons ces deux mesures en montrant qu'elles sont étrangères mais possèdent des relations analytiques intéressantes. De ce fait, nous concevons et expérimentons une nouvelle mesure de qualité d'implication en deux approches qui associe confiance et intensité. Nous montrons l'intérêt présenté par cette combinaison pour intégrer la contraposée de l'implication, condition nécessaire pour faire jouer à ce nouvel indice une fonction d'analyse causale.

Mots-clés : *confiance, fréquence conditionnelle, intensité d'implication, intensité entropique, implifiance, règle, quasi-règle, contraposée*

ABSTRACT

Most of the indexes used with association rules are based on the conditional probability, also called confidence, to measure the quality of a rule. With Statistical Implicative Analysis, another measure, the implication intensity, focuses on the same objective using only the implication. It is built with the probability of appearance of counter-examples of a rule. In this paper, we compare those two measures, showing they are different but also have interesting analytical properties. Hence, a new measure based on those two approaches, confidence and implication is proposed. The interest of this combination is to integrate the contrapositive of the implication, in order to give this new index a causal property.

Keywords : confidence, conditional probability, implication intensity, entropic intensity, implifiance (implidence), rule, quasi-rule, contrapositive

1 Introduction

1.1 Motivation

Nous nous plaçons dans le cadre du croisement d'un ensemble de sujets et d'un ensemble de variables binaires selon lesquelles les sujets prennent leurs valeurs. De nombreux auteurs (voir Références) ont établi des indices qui permettent de mettre en

[1] École Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, : regisgra@club-internet.fr

[2] FEMTO-ST, Université de Franche-Comté, Belfort, raphael.couturier@univ-fcomte.fr

[3] Dept. Matemàtiques, Universitat Jaume I, Campus del Riu Sec, Castelló E-12071 (SPAIN), gregori@mat.uji.es , http://www3.uji.es/~gregori

évidence des relations implicatives entre des variables binaires telles que, par exemple, a et b. Parmi ces indices et très souvent, ils font référence à la confiance, fréquence conditionnelle de b sachant a et établissent leurs indices d'association sur celle-ci. Dans cet article, nous établirons une relation entre l'intensité d'implication et la confiance, relation qui soulignera à la fois la différence entre ces deux indicateurs et leur éventuelle covariation. *De plus, pour des raisons sémantico-statistiques, nous les combinerons pour construire un nouvel indice composite d'implication statistique qui aura la vertu d'associer la tendance inclusive donnée par la confiance et l'étonnement statistique par l'intensité d'implication.*

Nous présentons également une nouvelle approche de la modélisation en A.S.I. pour des variables binaires prenant en compte autant l'implication directe que sa contraposée. Elle vise à se substituer à la modélisation entropique jusqu'alors utilisée et qui présente un caractère jugé trop ad-hoc par les familiers de l'A.S.I. Elle va donc être construite *contre une connaissance antérieure* comme le dit G. Bachelard dans (Bachelard G. 1967). Cependant, cette précédente modélisation était loin de déplaire aux utilisateurs qui en appréciaient la capacité à accepter plus facilement la grande taille de l'échantillon des sujets considéré. D'où son intérêt pour ce que l'on appelle les « big data ». De plus, prenant en compte la contraposée, elle semble remplir le rôle d'extracteur de relations causales, ce que ne permet pas ou moins finement l'implication dite classique mesurée par l'intensité d'implication. En effet, celle-ci conduit à la même valeur d'intensité autant pour l'implication directe que pour sa contraposée. Elle ne permet pas ainsi d'intégrer l'information nuancée apportée par la contraposée. Rappelons par contre qu'en cas d'implication logique stricte les deux formes implicatives sont équivalentes. Ce qui n'est pas le cas de la quasi-implication, objet central de nos recherches. Ainsi, cette nouvelle modélisation de l'implication entre deux variables binaires, tout en intégrant à la fois l'implication directe et sa contraposée, mais également la confiance, va adopter une méthodologie comparable à celle utilisée dans le cas de l'implication mesurée par l'intensité d'implication définie par R. Gras (Gras, 1979) à la base des premiers fondements de l'A.S.I.

1.2 Remarque épistémologique importante

Il pourrait être objecté à l'égard des auteurs de cet article que ce nouveau concept soit seulement le fruit d'un mariage arrangé entre deux partenaires étrangers l'un à l'autre, regardant dans des directions différentes. Or, l'un et l'autre alimentent l'objectif que nous nous sommes donné sachant que nous traitons essentiellement des problèmes qui relèvent du réel comme en sciences humaines ou relativement à des phénomènes scientifiques non entièrement modélisés (médecine, biologie, etc.) : rendre compte de l'implication en serrant au plus près la notion logique de celle-ci sachant que leurs célibats respectifs ne le permettent pas. Nous prétendons, en plagiant G. Vergnaud (Vergnaud, 2007) s'exprimant au sujet de l'esthétisme, qu'il n'y a pas d'analyse de données sans **confiance**. Mais il n'y pas non plus d'analyse de données sans **surprise**[4],

[4] C'est aussi ce qu'affirme René Thom (« Paraboles et catastrophes », 1980, p.130) : « …le problème n'est pas de décrire la réalité, le problème consiste bien plus à repérer en elle ce qui a de sens pour nous, ce qui est surprenant dans l'ensemble des faits. Si les faits ne nous surprennent pas, ils n'apportent aucun élément nouveau pour la compréhension de l'univers : autant donc les ignorer » et plus loin : « … ce qui

ni sans **correction d'échelle** comme il est fait en analyse en composants principales, par exemple, où une division par $\sqrt{n}$ s'impose dans la détermination de la distance entre sujets. Accoupler confiance et surprise n'a rien de plus artificiel que celui auquel sont soumis les physiciens de l'acoustique, de la thermodynamique, de l'électricité, de la relativité, etc.

Par exemple, sur le problème de chute des corps, l'observateur relève expérimentalement l'influence de certaines variables qui expriment que la résistance de l'air est proportionnelle au carré de la vitesse V, proportionnelle à la surface S du maître-couple du solide, à la masse spécifique r_0 de l'air, à un coefficient de forme C. Tout ceci est résumé par la formule : $R=kCr_0SV^2$. C'est une relation construite sur la base de mesures empiriques qui conduisent à la satisfaire. Il en est de même de la formule expérimentale des gaz parfaits : pV=Cte et toutes les règles énoncées sous le terme de principe[5].

Bien entendu, d'autres phénomènes physiques seront plongés par essence dans un modèle mathématique déduit (par ex. les phénomènes vibratoires). Et d'autres seront extraits de l'expérimentation et de la mise en évidence des variables intervenantes. Nous pourrons dire que notre modèle est bon lorsque les résultats expérimentaux que nous mènerons avec satisferont à la fois l'explicabilité, le bon sens et l'observation mais aussi auront une capacité prédictive. D'autres modèles concurrents du concept implicatif de notre étude présente existent et nous avons mis en évidence leurs adéquation ou leurs limites vis-à-vis des variables requises ou de leur pondération (Gras et Couturier, 2010). De toute façon, quel qu'il soit, le modèle utilisé est bon s'il rend compte au mieux de la réalité jusqu'à sa substitution éventuelle par un autre plus performant. Or c'est cette performance que nous voulons atteindre (cf. aussi note 2).

2 Relation entre confiance et intensité d'implication.

2.1 Rappels sur l'intensité d'implication

Relativement aux cardinaux de E (soit n), espace des sujets, de A (soit n_a), sujets satisfaisant a et de B (soit n_b), sujets satisfaisant b, c'est le poids des contre-exemples (soit $n_{a\wedge\bar{b}}$), i.e. des sujets satisfaisant a et non b, qu'il faut donc prendre en compte pour accepter statistiquement de conserver ou non la **quasi-implication** ou **quasi-règle** $a \Rightarrow b$. Ainsi, c'est à partir de la dialectique entre les exemples et les contre-exemples que la règle apparaît comme le dépassement de la contradiction.

Pour formaliser cette quasi-règle, nous formulons l'hypothèse que a et b sont indépendantes. Puis nous considérons, comme le fait I.C. Lerman dans (Lerman, 1981) pour la similarité, deux parties quelconques X et Y de E, choisies aléatoirement et indépendamment (absence de lien a priori entre ces deux parties) et de mêmes cardinaux

n'est pas possible si l'on ne dispose pas déjà d'une théorie ».

[5] « On nomme **principe physique** une loi physique apparente, qu'aucune expérience n'a invalidée jusque là bien qu'elle n'ait pas été démontrée, et joue un rôle voisin de celui d'un postulat en mathématiques ». (Wikipedia). « Parfois un principe peut être démontré à partir d'un ou plusieurs autres, à charge au physicien de choisir le principe de base pour ses raisonnements : par exemple, le principe de moindre action est équivalent au principe fondamental de la dynamique associé au principe de d'Alembert » (Wikipedia).

respectifs que A et B. Soit $\overline{Y}$ et $\overline{B}$ les ensembles complémentaires respectifs de Y et de B dans E de même cardinal $n_{\overline{b}} = n - n_b$. Soit α un réel quelconque de l'intervalle [0,1].

Définition 1: la quasi-règle $a \Rightarrow b$ est *admissible avec l'intensité* $1-\alpha$ si et seulement si $\Pr\left[\mathrm{Card}(X \cap \overline{Y}) \leq \mathrm{card}(A \cap \overline{B})\right] \leq \alpha$.

Définition 2: On appelle intensité d'implication de la quasi-règle $a \Rightarrow b$, pour la modélisation précédente, le nombre $\varphi(a,b) = 1 - \Pr\left[\mathrm{Card}(X \cap \overline{Y}) \leq \mathrm{Card}(A \cap \overline{B})\right]$ si $n_b \neq n$ et $\varphi(a,b) = 0$ si $n_b = n$.

On démontre (Gras R, 1995, puis autre preuve en 2009) que, pour une certaine modélisation de tirage de X et de Y, $\mathrm{Card}(X \cap \overline{Y})$ suit une loi de Poisson de paramètre $\lambda = \frac{n_a n_{\overline{b}}}{n}$. L'intensité d'implication est une valeur probabiliste, contrairement aux indices implicatifs les plus usités, qui fonde la décision de retenir ou non une relation de quasi-implication entre les variables binaires a et b. Intuitivement, elle représente une sorte d'étonnement statistique (une *surprise* ?) que le nombre de contre-exemples à $a \Rightarrow b$ soit petit alors qu'elles sont supposées indépendantes.

2.2 La confiance en tant que réalisation d'une variable aléatoire

Avec les mêmes notations, dans un corpus de données binaires, la confiance c est la fréquence conditionnelle pour qu'un sujet satisfasse à b sachant qu'il satisfait à a soit

$$c = \mathrm{Fr}[b|a] = \mathrm{Fr}[a \wedge b|a] = [(\mathrm{card}\ A \cap B) / \mathrm{card}\ A] = \frac{n_{a\wedge b}}{n_a} = \frac{n_{a\wedge b}/n}{n_a/n} = 1 - \frac{n_{a\wedge \overline{b}}}{n_a}$$

La variable aléatoire card $(X \cap Y) = N_{a\wedge b}$ est réalisée par card $(A \cap B)$ c'est-à-dire $n_{a\wedge b}$. Aussi, comme nous l'avons fait pour $N_{a\wedge \overline{b}}$ (Gras et Régnier, 2013) et selon une modélisation comparable, nous démontrerions que $N_{a\wedge b}$ suit approximativement, sous l'hypothèse d'indépendance, une loi de Poisson de paramètre estimé $\frac{n_a n_b}{n}$.

Considérons c comme la réalisation d'une variable aléatoire C fonction uniquement de card $(X \cap Y)$ définie par : $C = \frac{N_{a\wedge b}}{n_a}$. Cette confiance aléatoire conditionnelle prend ses valeurs réelles sur [0,1] et a pour loi la loi image, c'est-à-dire transportée de celle de $N_{a\wedge b}$. Ses valeurs sont donc des fractions des valeurs entières dont le numérateur est Poissonnien.

D'où $\Pr[C \geq \alpha] = \Pr[1 - \frac{N_{a\wedge \overline{b}}}{N_a} \geq \alpha] = \Pr[\frac{N_{a\wedge \overline{b}}}{N_a} \leq 1-\alpha]$

Or dans la contingence, la variable aléatoire N_a prend la valeur fixée et égale à n_a.

Par suite, $\Pr[C \geq \alpha] = \Pr[N_{a\wedge \overline{b}} \leq n_a(1-\alpha)]$.

2.3 Comparaison intensité d'implication et confiance

Rappelons que $Q(a,\bar{b}) = \dfrac{\mathrm{Card}(X \cap \bar{Y}) - \dfrac{n_a n_{\bar{b}}}{n}}{\sqrt{\dfrac{n_a n_{\bar{b}}}{n}}}$ est la variable aléatoire gaussienne centrée réduite limite de la variable de Poisson $\mathrm{Card}(X \cap \bar{Y})$ ou $N_{a\wedge\bar{b}}$, nombre aléatoire de contre-exemples à l'implication. D'où l'on tire :

$$\Pr[N_{a\wedge\bar{b}} \leq n_a(1-\alpha)] = \Pr\left[Q(a,\bar{b}) \cdot \sqrt{\frac{n_a n_{\bar{b}}}{n}} + \frac{n_a n_{\bar{b}}}{n} \leq n_a(1-\alpha)\right].$$

Posons $r(a,\bar{b}) = \dfrac{n_a(1-\alpha) - \dfrac{n_a n_{\bar{b}}}{n}}{\dfrac{n_a n_{\bar{b}}}{n}}$. Alors $\Pr[C \geq \alpha] = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{r(a,\bar{b})} e^{-\frac{t^2}{2}} dt$

On vérifie bien que $\Pr[C \geq \alpha]$ est fonction décroissante de α.

A titre de comparaison, rappelons que $\varphi(a,b) = 1 - \Pr\left[Q(a,\bar{b}) \leq q(a,\bar{b})\right] = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$. Posons : $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt = \alpha$.

Alors : si $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{r(a,\bar{b})} e^{-\frac{t^2}{2}} dt \geq \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$ la confiance est plus grande que l'intensité d'implication, sinon, c'est cette dernière qui l'emporte.

Or, avec les mêmes notations, pour α donné, $[\varphi(a,b) \geq \alpha]$ qui est aussi équivalent à $\Pr[N_{a\wedge\bar{b}} \geq n_{a\wedge\bar{b}}] \geq \alpha$ nous rappelle que l'évènement $[\varphi(a,b) \geq \alpha]$ est de moins en moins probable lorsque α croît. Nous y reviendrons plus loin.

Autrement dit, encore, partant de l'inégalité : $n_{a\wedge\bar{b}} \leq N_{a\wedge\bar{b}} \leq n_a\,(1-\alpha)$, pour un seuil comparable, si l'inégalité de gauche est satisfaite l'implication est de bonne qualité. Si celle de droite l'est, ce sera la confiance.

Les deux concepts (confiance et intensité d'implication) répondent donc à des principes relativement distincts mais non contradictoires : la confiance $Fr[b|a]$ est fondée sur la subordination de la variable b à la variable a alors que l'intensité d'implication se fonde sur les contre-exemples à la relation de sujétion de b par a.

Ainsi, si l'on souhaite une bonne qualité d'implication (par ex. $\alpha = 0.95$), il est nécessaire que le nombre de contre-exemples obtenus par hasard soit plus grand que celui de la contingence. En revanche, si l'on veut que la confiance C ait des chances d'être grande (par ex. $\beta = 0.95$), il est nécessaire que le nombre d'exemples que donnerait le seul hasard ne soit pas plus grand que $0.05\ n_a$. On note alors que la probabilité de la variable confiance ne dépend plus de n_b, contrairement à l'intensité d'implication.

Rappelons, à ce sujet, que nous avons prouvé que lorsque le nombre de sujets vérifiant b tend vers n ou est égal à n, la relation implicative devient triviale, banale alors que la confiance sera maximale. Nous avons également montré que la confiance ne varie pas dans toute dilatation de E et des exemples de A et de B. Ce qui nous a conduits à refuser le critère « fréquence conditionnelle » comme seul critère de révélation d'une relation implicative. Elle est cependant, le fondement de la notion de réseau Bayesien qui permet d'organiser un ensemble de variables selon un graphe, la transition d'un nœud du graphe au suivant se faisant sur la base de la fréquence conditionnelle du premier sur le second.

Remarque prospective

Il serait alors possible, dans le cadre d'une recherche, de définir un autre indice probabiliste comme l'est φ(a,b), pour évaluer la qualité de l'implication de a sur b.

3 Lemme

Le comportement asymptotique de φ(a,b), intensité d'implication, est celui d'une variable uniforme sur l'intervalle [0 ,1].

Ce lemme a été démontré dans Gras et al (1996), mais sera rappelé et illustré par un exemple numérique dans l'annexe. Nous utiliserons ce résultat dans la suite de cet article.

4 Indicateur de proximité asymptotique entre confiance et intensité d'implication

Ainsi, $\Pr[\varphi(a,b)\leq\alpha]\sim\alpha$ ou, de façon équivalente vu la continuité de la mesure, $\Pr[\varphi(a,b)\geq\alpha]\sim 1-\alpha$. Or : $\Pr[C\geq\alpha]= 1 - \frac{1}{\sqrt{2\pi}}\int_{r(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}}dt$

Dans ces conditions, le rapport $\Pr[C\geq\alpha]/\Pr[\varphi(a,b)\geq\alpha] \sim \frac{\Pr[C\geq\alpha]}{1-\alpha}$ est un bon indicateur de satisfaction entre confiance et intensité d'implication : plus grand que 1, la confiance est alors meilleure que l'intensité ; inférieure à 1, c'est l'intensité qui est plus forte. Une recherche ultérieure pourrait s'appuyer sur cet indicateur.

5 Une première approche d'une intensité d'implication intégrant la contraposée

Nous présentons cette approche pour des raisons didactiques : partant d'une intuition, un développement théorique se construit mais son débouché conduit à un problème que les auteurs avaient dénoncé et déjà tenté de résoudre. Cela illustre cependant la démarche du chercheur dont les errements peuvent mener à des impasses.

L'implication a $\Rightarrow$ b, où $n_a \leq n_b$, pour des variables binaires, est d'autant plus réalisée que $n_{a\wedge b}$ est peu différent de $n_{a.}$, c'est-à-dire que l'inclusion de A dans B est vérifiée.

La contraposée de cette implication, $\bar{b} \Rightarrow \bar{a}$, est elle-aussi d'autant plus réalisée que l'inclusion de $\bar{B}$ dans $\bar{A}$ est validée, c'est-à-dire que $n_{a\wedge b} + n_{\bar{a}\wedge\bar{b}}$.est peu différent de $n_a + n_{\bar{b}}$, condition nécessaire à la satisfaction de bonnes qualités d'implication tant directe que contraposée.

	b	$\bar{\mathbf{b}}$	**Totaux**
a	$n_{a\wedge b}$	$n_{a\wedge\bar{b}}$	$\mathbf{n_a}$
$\bar{\mathbf{a}}$	$n_{\bar{a}\wedge b}$	$n_{\bar{a}\wedge\bar{b}}$	$\mathbf{n_{\bar{a}}}$
Totaux	$\mathbf{n_b}$	$\mathbf{n_{\bar{b}}}$	**n**

Tab. 1

Soit $N_{a\wedge b}$ et $N_{\bar{a}\wedge\bar{b}}$ les variables aléatoires qui représentent les cardinaux des sous-ensembles aléatoires $X \cap Y$ et $\bar{X} \cap \bar{Y}$ de E respectivement associés à $A \cap B$ et $\bar{A}\cap\bar{B}$ et de mêmes cardinaux. Posons $T = N_{a\wedge b} + N_{\bar{a}\wedge\bar{b}}$ et $t = n_{a\wedge b} + n_{\bar{a}\wedge\bar{b}}$ sa réalisation dans l'expérience de tirage uniforme.

Définition 3

On appelle **intensité d'implication inclusive** associée à a $\Rightarrow$ b, dans l'hypothèse d'indépendance a priori de a et b, la probabilité :

K(a,b) =Pr[T $\leq$ t] ou explicitement : $K(a,b) = \Pr[N_{a\wedge b} + N_{\bar{a}\wedge\bar{b}} \leq n_{a\wedge b} + n_{\bar{a}\wedge\bar{b}}]$

Cette intensité inclusive est la probabilité pour que, si les variables a et b n'avaient aucun lien entre elles, le hasard des inclusions de A dans B et de leurs complémentaires inversés conduise globalement à plus d'exemples que ceux qui ont été observés. Autrement dit, plus les quasi-inclusions de A dans B et de $\bar{B}$ dans $\bar{A}$ seront proches des inclusions strictes, plus le nombre des seuls exemples satisfaisant les inclusions et dues au hasard sera réduit.

Par suite : $\Pr[T \leq t] = \sum_{k=0}^{k=n_{a\wedge b}+n_{\bar{a}\wedge\bar{b}}} \Pr[N_{a\wedge b} + N_{\bar{a}\wedge\bar{b}} = k]$

Choix du modèle hypergéométrique de T.

Adoptant une distribution uniforme sur l'ensemble des sujets la loi de T est hypergéométrique. Par suite :

$$\Pr[T \leq t] = \sum_{k=0}^{k=n_{a\wedge b}+n_{\bar{a}\wedge\bar{b}}} . \sum_{l=0}^{l=n_{a\wedge b}} \Pr[N_{a\wedge b}=l, N_{\bar{a}\wedge\bar{b}} = k-l]$$

$$= \sum_{k=0}^{k=n_{a\wedge b}+n_{\bar{a}\wedge\bar{b}}} \sum_{l=0}^{l=n_{a\wedge b}} \frac{C_{n_{a\wedge b}}^{l} . C_{n_{a\wedge b}-l}^{k}}{C_{n_{a\wedge b}}^{l}}$$

Les valeurs de ce nouvel indice d'implication inclusif utiliseront les moyenne et variance de la loi hypergéométrique et, de ce fait, pourront, dans la plupart des cas, être ramenés à ceux d'une loi binomiale, voire gaussienne comme il est fait dans le cas classique. Une étude ultérieure approfondie serait intéressante en partant de la loi de T.

Remarque

Une modélisation différente pourrait consister à définir séparément les lois de probabilité des variables $N_{a\wedge b}$ et $N_{\bar{a}\wedge\bar{b}}$ en comparant leurs variations aux réalisations respectives à $n_{a\wedge b}$ et $n_{\bar{a}\wedge\bar{b}}$. Cependant, tenant compte des simulations effectuées, nous abandonnerons cette démarche qui, bien qu'intégrant de façon naturelle l'information due à la contraposée, conduit aux mêmes réserves que l'implication classique, à savoir une valeur de l'intensité proche de 1 dès que les occurrences sont grandes. Aussi, nous nous tournons vers une deuxième approche.

6 Une deuxième approche d'une intensité d'implication intégrant la contraposée[6]

6.1 6.1 Motivation sémantique et épistémologique

Puisque implication et confiance admettent des racines communes (cf. § 3), nous associerons, sans agir contre nature, ces deux concepts, à la façon du physicien expérimentaliste comme il a été dit dans l'introduction. Les valeurs de la confiance relative respectivement aux règles $a \Rightarrow b$ et $\bar{b} \Rightarrow \bar{a}$ sont, en utilisant les mêmes notations que dans 2.1 :

$$C_1(a,b) = \text{Fr}\,[Y \mid X] = ([\text{card } X \cap Y]) / (\text{card } X) = \frac{n_{a\wedge b}/n}{n_a/n} = \frac{n_{a\wedge b}}{n_a}$$

$$C_2(\bar{b},\bar{a}) = \text{Fr}\,[\bar{X} \mid \bar{Y}] = ([\text{card } \bar{X} \cap \bar{Y}]) / (\text{card } \bar{Y}) = \frac{n_{\bar{a}\wedge\bar{b}}/n}{n_{\bar{b}}/n} = \frac{n_{\bar{a}\wedge\bar{b}}}{n_{\bar{b}}}$$

Or, compte tenu de leur définition, les deux intensités d'implication $\varphi(a,b)$ et $\varphi(\bar{b},\bar{a})$ sont égales. De ce fait, l'intensité ne nuance pas l'étonnement statistique de l'implication et de sa contraposée. En revanche, les deux confiances sont différentes en général. C_1 et C_2 sont d'excellents indicateurs des inclusions partielles : $A \subset B$ et $\bar{B} \subset \bar{A}$. Mais contrairement à $\varphi(a,b)$ et $\varphi(\bar{b},\bar{a})$, ils sont invariants dans toute homothétie globale des ensembles en jeu. Par suite, ils ne permettront pas de déceler des relations « surprenantes » eu égard à l'échantillon observé de taille n. C'est pour cette raison que nous leur associerons l'intensité d'implication classique,

Nous utiliserons cependant les nuances inclusives apportées par ces indicateurs d'inclusion, pour affecter l'intensité d'implication classique de coefficients permettant à la fois :

- d'intégrer les informations de l'implication directe et de sa contraposée
- d'obtenir une meilleure discrimination des valeurs critiques des contre-exemples pour des grands échantillons où la valeur 1 apparaît comme point d'accumulation de l'intensité d'implication,
- et aussi de crédibiliser le caractère inclusif de A dans B.

Ainsi, en associant les deux indices, nous devrions obtenir une mesure qui intègre à la fois la recherche des implications surprenantes (à la charge de l'intensité

[6] Ne craignons pas le changement que nous introduisons car comme l'affirme G.Bachelard : « *Accéder à la science, c'est, spirituellement rajeunir, c'est accepter une mutation brusque qui doit contredire un passé* ».(Extrait de « La Formation de l'esprit scientifique », 1938).

d'implication classique) et la crédibilité de l'inclusion conditionnelle de A dans B et de $\overline{B}$ dans $\overline{A}$ (à la charge des confiances).

Notre objectif est alors de définir une fonction de C_1 et C_2 qui permette de satisfaire les critères précédents tout en majorant l'effet implicatif dont rend compte l'intensité d'implication afin d'obtenir une relation de la forme : φ(a,b). $f(C_1, C_2)$. Une fonction du type : $f(C_1.C_2) = [C_1(a,b).C_2(\overline{b}, \overline{a})]^r$ avec r < ½ satisfait nos exigences. Nous choisirons arbitrairement r = ¼ et examinerons, à travers nos expériences si cette valeur est satisfaisante pour restituer une information riche, plausible et susceptible de prédictabilité dans un cas causal. Quoi qu'il en soit, cette valeur petite de r affecte le pouvoir de la confiance en lui réduisant son effet d'évidence intrinsèque. De ce fait, il préserve la propriété de « surprise », « d'étonnement statistique » que recèle l'intensité d'implication. Ce qui nous satisfait.

6.2 La corbeille de la mariée.

Définition 4

On appelle **implifiance**[7] la mesure de l'implication statistique qui prend en compte l'implication directe et sa contraposée ainsi que la confiance en chacune de ces deux formes inclusives. Sa valeur est :

$$\Phi(a,b) = \varphi(a,b).\ [C_1(a,b).C_2(\overline{b}, \overline{a})]^{1/4}$$

Par exemple, si l'on extrait une règle dont l'implifiance est égale à 0.95, son intensité d'implication est au minimum égale à 0.95 et chacune des confiances C_1 et C_2 est au moins égale à 0.81. Si l'implifiance est égale à 0.90, les minimas respectifs sont 0.90 et 0.66, ce qui préserve la plausibilité de la règle.

Conclusion partielle

Ainsi, même si les covariations de l'intensité d'implication et la confiance ne sont pas anarchiques, si elles n'obéissent pas ensemble à une référence en termes de probabilité, même si nous nous trouvons devant le choix arbitraire d'une définition composite d'où est partiellement exclue la part de contrôle de l'utilisateur de cette mesure, elles porteront les traces pondérées de deux indices majeurs pour évaluer la qualité implicative.

Quelques propriétés

- P1 : si na = $n_{\overline{b}}$, alors C1 = C2. Nous avons dans ce cas, une même confiance en l'implication directe et sa contraposée ;
- P2 : $\frac{1-C_1}{1-C_2} = \frac{n_{\overline{b}}}{n_a}$. Si ce dernier rapport reste constant, les deux confiances varient dans le même sens ;
- P3 : en exprimant C1 = $\frac{n_{a\wedge b}}{n_a}$ = 1- $\frac{n_{a\wedge\overline{b}}}{n_a}$ et C2 = $\frac{n_{\overline{a}\wedge\overline{b}}}{n_{\overline{b}}}$ = 1 - $\frac{n_{a\wedge\overline{b}}}{n_{\overline{b}}}$ en fonction du nombre de contre-exemples à l'implication directe, on obtient les dérivées partielles de ces confiances par rapport à ce paramètre :

[7] Comme son nom le laisse penser, ce nouveau mot est la contraction de « **Impli**cation » et « Con**fiance** » puisque le concept associé est la combinaison des deux concepts évoqués ici.

* $\frac{\partial C_1}{\partial n_{a\wedge\bar{b}}}$ = - $\frac{1}{n_a}$ donc C_1 décroît quand les contre-exemples augmentent et d'autant plus vite que n_a est petit ,
* $\frac{\partial C_2}{\partial n_{a\wedge\bar{b}}}$ = - $\frac{1}{n_{\bar{b}}}$ donc C_2 décroît quand les contre-exemples augmentent et d'autant plus vite que n_b est grand, les autres paramètres étant constants.

Ces résultats sont compatibles avec la formule établie dans (Gras et Régnier, 2003) : $\frac{\partial q}{\partial n_{a\wedge\bar{b}}} = \frac{1}{\sqrt{\frac{n_a n_{\bar{b}}}{n}}} = \frac{1}{\sqrt{\frac{n_a (n - n_b)}{n}}} > 0$, qui assurent ainsi qu'intensité d'implication et confiance varient plutôt dans le même sens en accord avec le point 2.3.

Remarque

Dans notre première approche, le cadre de la modélisation imposait la nature binaire des variables en jeu. Dans cette seconde approche, les deux notions de confiance, détachée de la sémantique forte en environnement binaire, gardent des propriétés fonctionnelles ce qui autorise la définition d'implifiance dans les cas où les variables sont de nature quelconque (numériques, modales, floues, variables intervalles, etc.).

7 Quelques comparaisons

7.1 Comparaisons entre l'intensité d'implication et l'implifiance

Voici une représentation (Fig. 1) de l'intensité d'implication (II) et de l'implifiance (IC) (en ordonnée) pour des valeurs de n = 100, n_a = 40, de n_b = 50 et où l'on fait varier $n_{a\wedge\bar{b}}$ de 0 à 40 en abscisses. Sur cette même représentation, nous faisons figurer, dans un objectif de comparaison, comment varient les 3 modes de valuation de la qualité de l'implication II, IC et IE (intensité entropique utilisée précédemment). Cette dernière sera également présente dans les figures 2,3, 4, 5, 6, 7, 8 et 9 où nous faisons varier les paramètres fondamentaux : n, n_a et n_b.

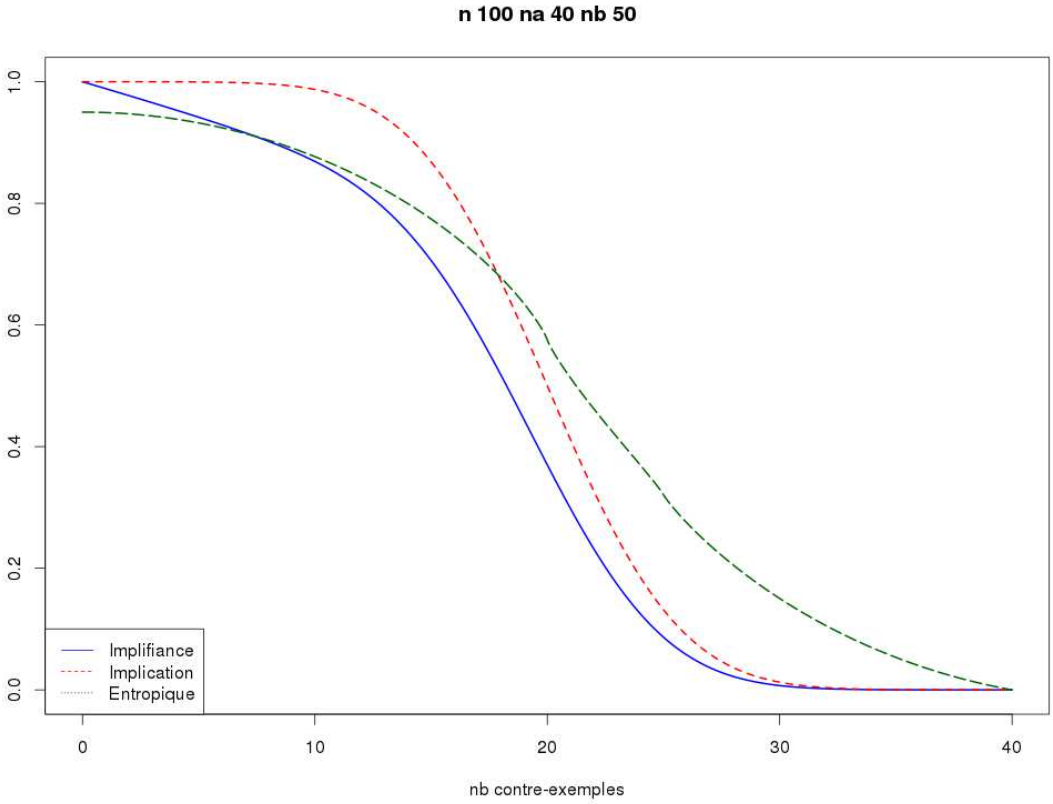


Figure 1

On remarque immédiatement, ce qui était attendu compte tenu de sa définition, que l'implifiance est toujours inférieure à l'intensité d'implication classique. Une valeur de 0.95 pour l'intensité d'implication pourra être accompagnée d'une valeur de l'implifiance compris entre 0.85 et 0.90. Que l'usager ne s'en inquiète pas. Par exemple, des confiances C_1 et C_2 supérieures à 0.50 conduisent à une implifiance supérieure à 0.67 alors que l'intensité est 0.95. De plus, l'IC n'ayant pas de dérivée nulle à l'origine décroît plus vite que l'II mais cette décroissance est lente avant de s'accélérer à n/4, voisinage du point d'inflexion de la courbe. En revanche, comme nous le savons, l'II décroît très lentement au début de la croissance des contre-exemples mais cette lenteur est préjudiciable pour des valeurs de n grandes car elle devient alors peu discriminante comme nous le voyons par une homothétie x100, dans la Fig. 2 où cette fois n = 10000, n_a = 4000, de n_b = 5000 et où l'on fait varier $n_{a \wedge \bar{b}}$ de 0 à 4000. L'II (intensité d'implication) ne varie que très peu de la valeur maximum 1 jusqu'à $n_{a \wedge \bar{b}}$ = 2000, puis plonge brusquement plus vite, ce qui la valorise, que IE, l'intensité entropique.

Remarquons que l'IC est plus fidèle à l'II que ne l'était la version précédente de la mesure de qualité entropique.

D'autres situations avec n= 10 et n_a = 1, n = 100 et n_a = 10 (Fig. 3), n = 1000 et n_a = 1000 montrent que le rejet de l'implication, en raison de la faiblesse de la valeur de IC, n'intervient pas trop tôt et seulement au voisinage du cardinal de $X \cap \overline{Y} = n_a/2$.

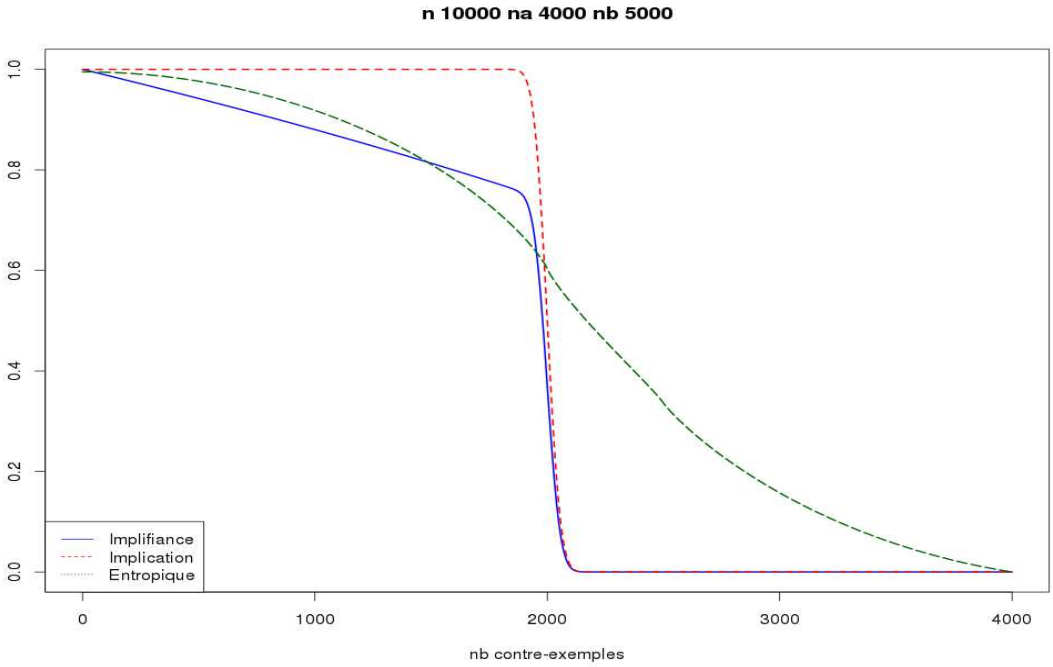


Figure 2

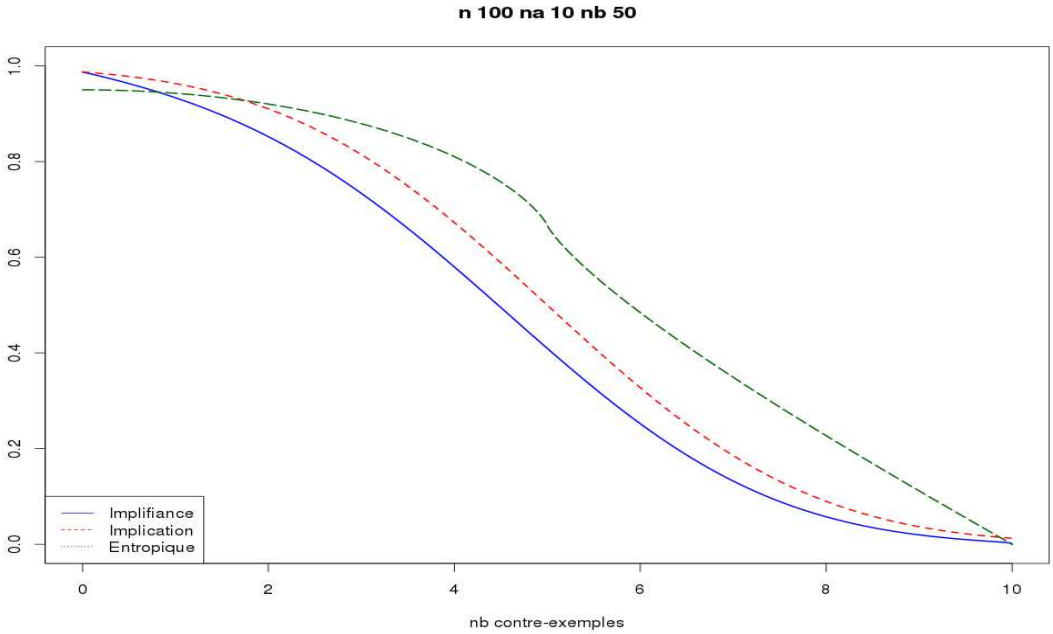


Figure 3

Lors de l'examen des spécificités de l'ASI disposant de l'intensité φ(a,b) comme outil de mesure implicative, nous avions souligné la propriété suivante : au voisinage de $N_{a\wedge\bar{b}} = 0$, l'intensité d'implication varie peu et reste très voisine de 1, comme le montrent les courbes (cf Fig.1 et Fig. 2), dont la dérivée à l'origine est presque nulle. Cette propriété a la vertu de permettre une certaine résistance de la fonction implicative, donc également prédictive, avant de céder à de trop forts écarts autour de 0. Certes, ceci

évite le rejet brutal de l'implication pour quelques contre-exemples qui pourraient n'être qu'accidentels ou dus à des erreurs de mesure. Mais le prix à payer pour garantir cette sagesse se facture en des valeurs de l'intensité difficilement discriminantes comme nous venons de le voir. Aussi, l'implifiance semble associer harmonieusement cette réserve à l'égard des écarts admissibles à l'implication et cependant la mise en garde du chercheur envers l'accumulation de contre-exemples devient insoutenable.

7.2 Comparaisons du comportement de l'implifiance par rapport à celui des confiances C1 et C2

Une autre propriété de l'intensité d'implication II a été soulignée. Ses variations en fonction de la variable $N_{a\wedge\bar{b}}$ ne sont pas linéaires contrairement à la seule confiance des multiples et autres indices d'association. Or, nous avons dénoncé l'inadéquation de la linéarité en tant que modélisante dans le fonctionnement de la pensée. La philosophie structuraliste nous rappelle que le « tout est plus riche que la somme de ses parties ». « *Autrement dit, dans le passage* ***non additif***, ***non linéaire*** *des parties au tout, il y a apparition de propriétés qui ne sont d'aucune manière précontenues dans les parties et ne peuvent donc s'expliquer par elles* » (Sève L., 2005). Ce phénomène qui conduit bien souvent à l'émergence d'une propriété du tout nécessite analytiquement un caractère non dérivable ou à points d'inflexion de la fonction qui mesure le phénomène. D'où cette mise à l'écart de la linéarité. En effet, en un point d'inflexion de l'implifiance, celui-ci marque le passage de la phase d'accélération de la décroissance de l'implifiance à sa phase de décélération. Les courbes qui décrivent les variations de l'implifiance montrent bien ce caractère non linéaire qui tient à la présence multiplicative de l'intensité d'implication dans sa définition.

En revanche, comme le montrent les courbes représentant C_1 et C_2 des figures 4a, 5a et 6a, la linéarité de celles-ci ne respecte pas la philosophie dont s'arroge la nouvelle mesure implicative. Rappelons que C_1 mesure la fréquence de b sachant a et que C_2 mesure celle de non a sachant non b. Ces deux fréquences ne sont qu'exceptionnellement égales et se présentent avec une valuation indifférente : tantôt C_1 est au-dessus de C_2 et tantôt le contraire.

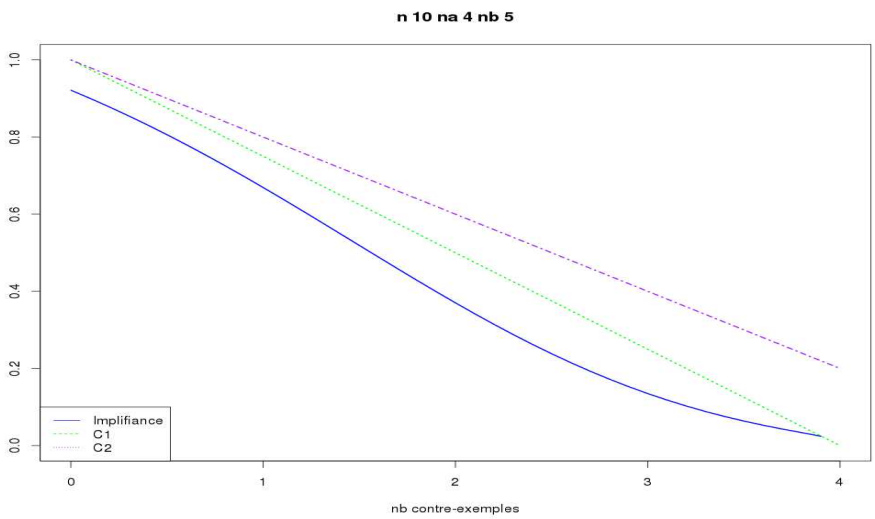


Figure 4a

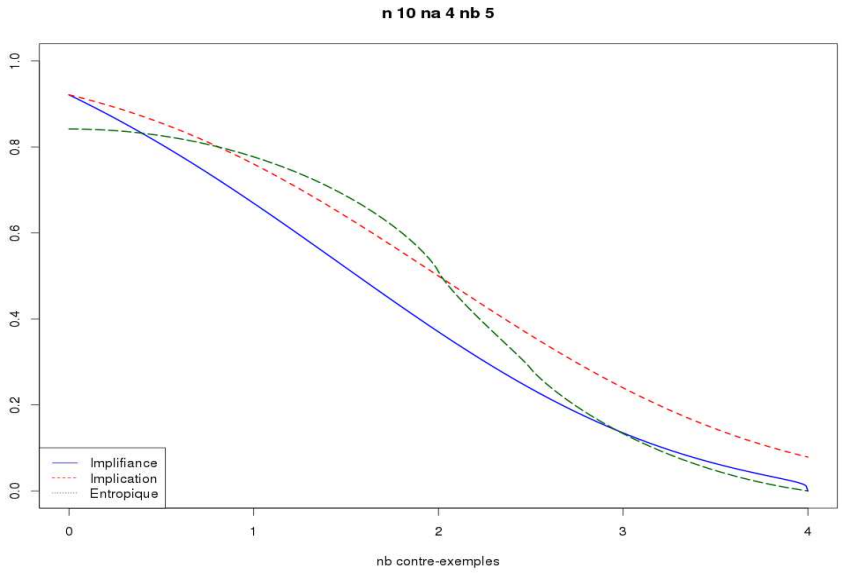


Figure 4

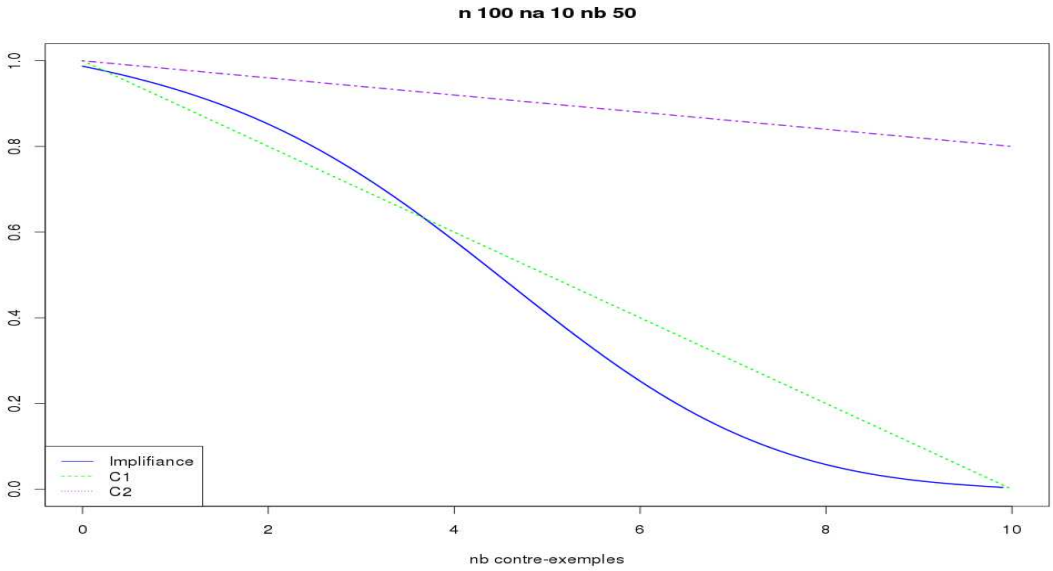


Figure 5a

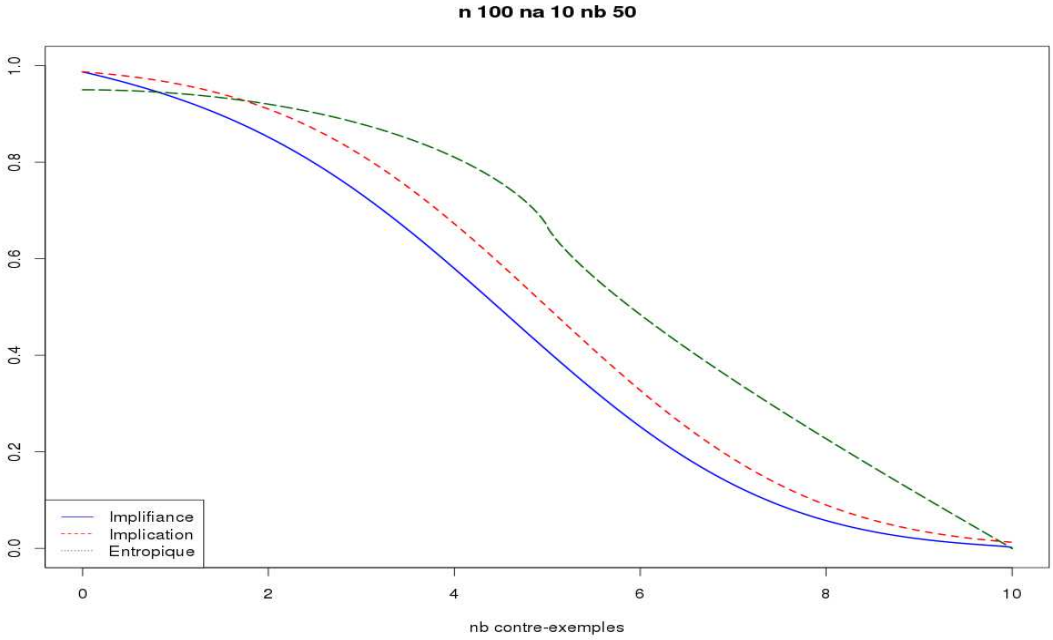


Figure 5b

En revanche, la figure 6a montre clairement l'effet « lissage » produit par les coefficients des deux facteurs de confiance mais sans en supporter le défaut majeur de linéarité sur l'ensemble des nombres de contre-exemples : décroissance de l'implifiance de 0 à $n_a/2$, puis plongeon de celle-ci lorsque la règle implicative associée devient insoutenable. Cet effet « lissage » est répercuté sur l'implifiance comme on le voit sur la figure 6.

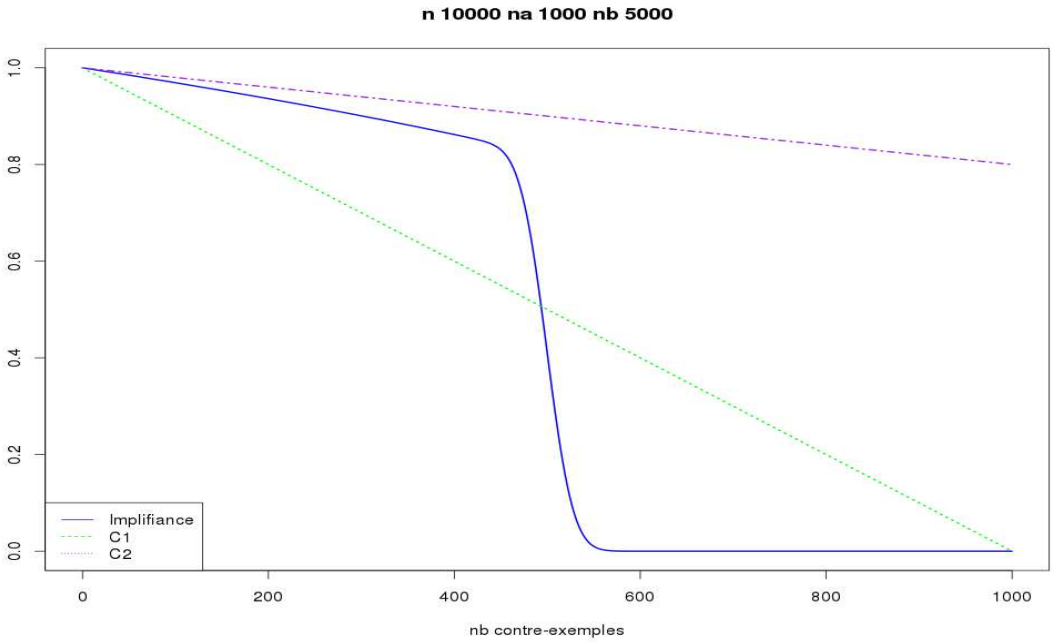


Figure 6a

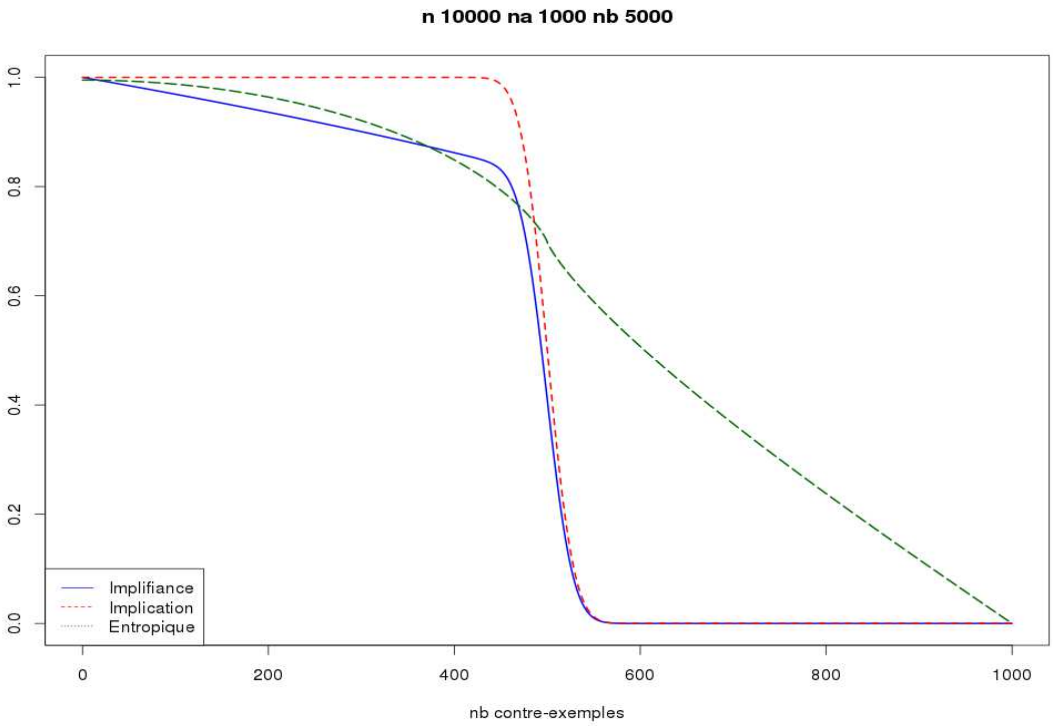


Figure 6b

7.3 Comparaisons entre l'intensité d'implication, l'implifiance et l'intensité entropique

En réaction au comportement de l'intensité d'implication II pour les corpus volumineux, insuffisamment discriminante sur une grande plage de contre-exemples, nous avions défini un nouvel indice pour mesurer la qualité des règles implicatives. La définition s'appuyait sur la notion d'entropie conditionnelle des évènements associés à l'implication directe et sa contraposée. Comme nous l'avons dit dans l'introduction, il lui fut reproché son côté quelque peu ad-hoc ainsi que la nécessité de modifier l'entropie à partir de $n_a/2$ pour lui ôter sa propriété de symétrie incompatible avec la philosophie implicative. Autre grief : comme l'II la décroissance de l'intensité entropique semblait insuffisante pour rendre compte de l'admissibilité du caractère implicatif des règles évaluées. Ceci nous a conduits à la nouvelle mesure envisagée dans cet article : l'implifiance.

Les courbes 7, 8 et 9 qui suivent, obtenues également dans la situation où le nombre de sujets est important : n =10000, illustrent la comparaison entre les 3 mesures : l'intensité d'implication classique, l'intensité entropique et l'implifiance.

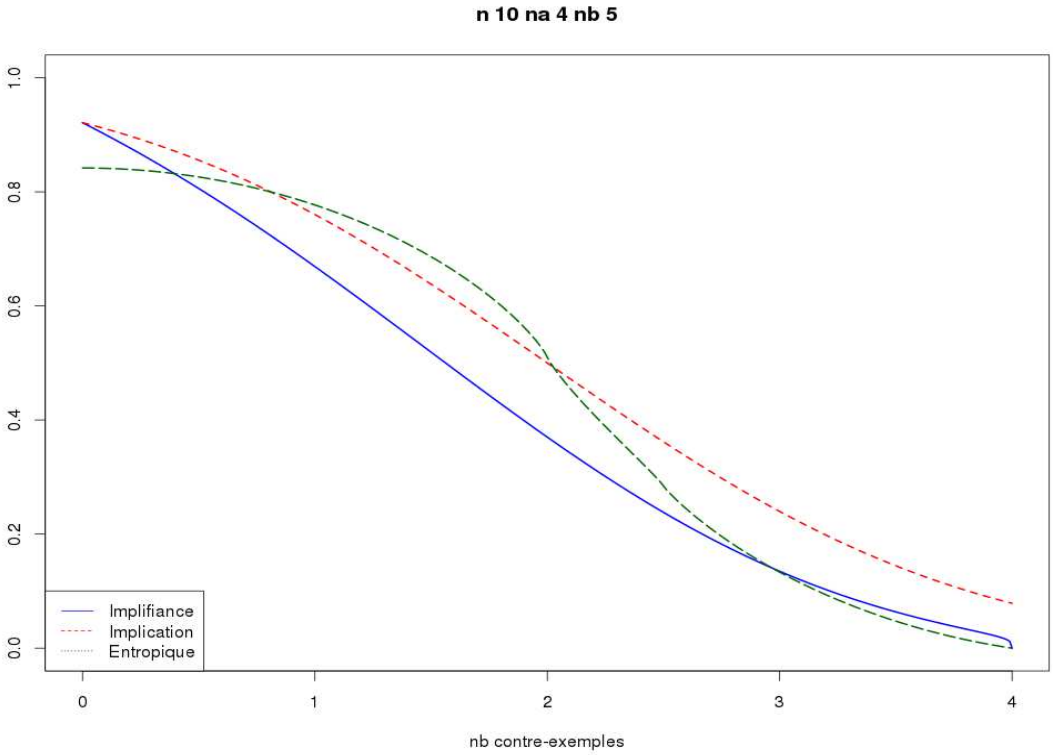


Figure 7

On remarquera, en effet, la plus grande résistance au rejet de la règle implicative dès que le nombre de contre-exemples devient trop importants eu égard aux valeurs de n_a et n_b, et particulièrement selon la courbe 9 des grandes valeurs des données.

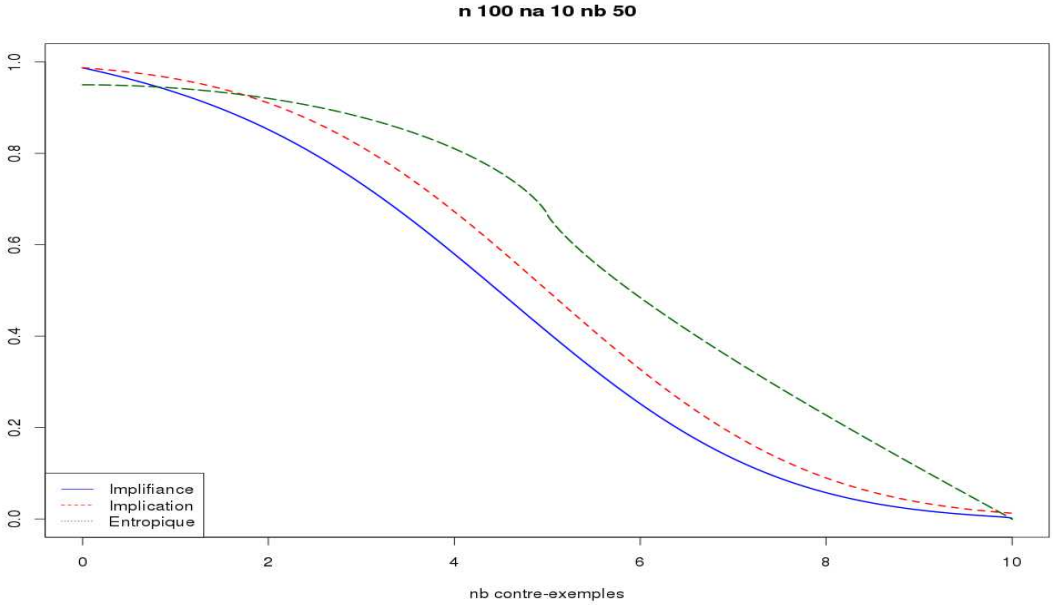


Figure 8

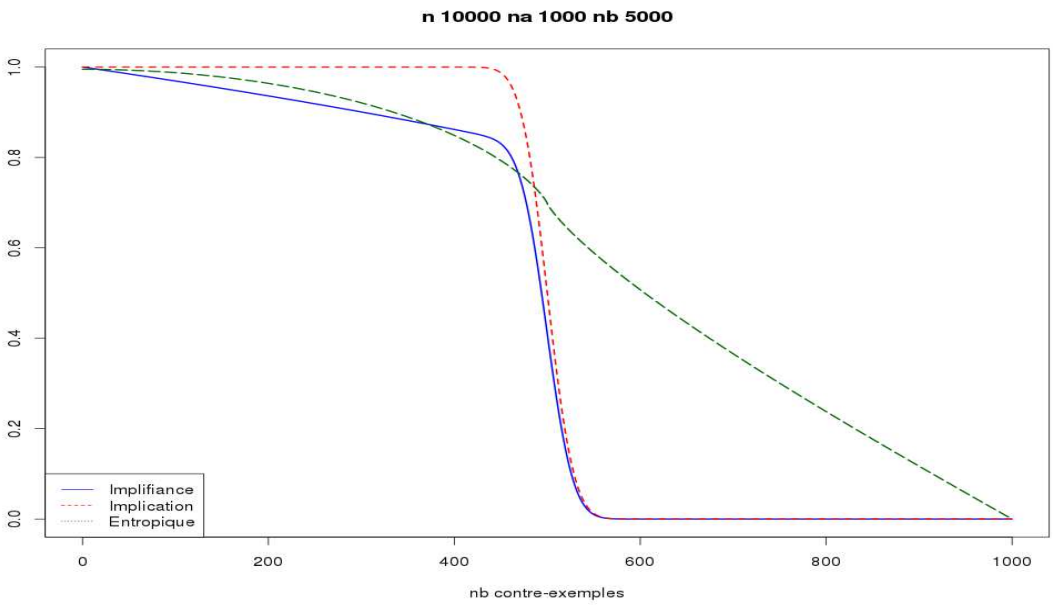


Figure 9

Comme nous l'avons déjà souligné, nous notons, avec ces 3 exemples 7, 8 et 9, une plus grande affinité entre l'intensité II et l'implifiance IC qu'entre II et l'intensité entropique IE. La fig. 9 est significative à cet égard où IE tarde à décroître avec le nombre croissant de contre-exemples. Ceci n'empêche pas l'implifiance de restituer à l'II les nuances de « confiance » que lui confèrent les deux composantes de la fréquence conditionnelle C_1 et C_2. Ceci confirme le contenu du § 3 où nous avions montré la relation entre intensité d'implication et confiance.

8 Conclusion

Nous proposons dans cet article la construction d'une nouvelle mesure qui permette d'évaluer la qualité de règles implicatives sur la base de l'importance relative des contre-exemples à l'implication. *Nous poursuivons donc la consolidation de notre édifice* ***non-réductionniste*** *de la mise en évidence et de représentation de relations présumées causales, cachées au sein d'un large ensemble de variables, édifice intégrant aussi la dualité sujets-variables, édifice qui fait de l'ASI un modèle original et fécond*. D'ailleurs, nous avons pu laisser ouvertes quelques pistes de recherche ultérieure.

Cette mesure, *l'implifiance*, comme l'intensité entropique jusqu'alors utilisée en complément de l'intensité d'implication, a la vertu de prendre en compte la contraposée de l'implication faute de quoi la fonction de recherche causale nous semblerait imparfaite. *Par rapport à l'intensité entropique, elle satisfait le principe philosophique du rasoir d'Occam*[8] *car sa définition est moins complexe*. Elle intègre cette fois et en outre, la notion de confiance ou fréquence conditionnelle afin de limiter à l'examen des règles celles qui présentent un caractère de liaison conditionnelle entre la prémisse et la

[8] Ce principe est généralement retenu par les scientifiques qui affirment qu'en présence de plusieurs théories en charge de la même réalité, il est préférable de choisir la plus simple.

conclusion. Tout comme ses mesures aînées, cette implifiance subira les assauts des données réelles *(« Expérience, source unique de vérité »* écrit Henri Poincaré dans « Science et hypothèse ») pour pouvoir prétendre être un bon prédicteur de relation causale et un bon outil d'extraction de pépites de connaissance. Et de toute façon, la qualité de la prévision restera entachée des incertitudes inhérentes au choix du modèle probabiliste. Car, comme le dit avec humour le physicien Niels Bohr : « *La prévision est un art difficile, surtout quand elle concerne l'avenir* ».

Références

[1] Agrawal R., T. Imielinsky and A. Swami (1993), Mining association rules between sets of items in large databases, *Proc. of the ACM SIGMOD'93*, 207-216.

[2] Amarger S., D. Dubois and H. Prade (1991), Imprecise quantifiers and conditional probabilities, In *Symbolic and quantitative approaches to* uncertainty (R. KRUSE, P. SIEGEL), Springer-Verlag, 33-37.

[3] Bachelard G. (1967), *La Formation de l'esprit scientifique*, Paris, 5e édition, Librairie philosophique J. Vrin.

[4] Briand H., M. Sebag, R. Gras et F.Guillet (2004), *Mesures de Qualité pour la Fouille de Données*, H.Briand, M.Sebag, R.Gras et F.Guillet eds, RNTI-E-1, Cépaduès, 2004.

[5] Brin S., R. Motwani and C. Silverstein (1997), Beyond market baskets: generalizing association rules to correlations, *Proc. Of ACM SIGMOD Conf. On Management of Data SIGMOD'97*, 265-276.

[6] Couturier, R. (2008). CHIC: cohesive Hierarchical Implicative Classification, In *Statistical implicative analysis*. Volume 127 of Studies in Computational Intelligence, Springer Verlag, p. 41–54.

[7] Fayyad U., G. Piatetsky-Shapiro and P. Smyth (1996), From Data Mining to Knowledge Discovery. In *Advances In Knowledge Discovery and Data Mining*, Fayyad U., Piatetsky-Shapiro G., Smyth P, and Uthurusamy R. eds, AAAI/MIT Press, 1-31.

[8] Frawley W., G. Piatetski-Shapiro and C. Matheus (1992), Knowledge discovery in databases: an overview. *AI Magazine*. 14(3), 57-70.

[9] Gras R. (1979), *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'Etat, Université de Rennes 1.

[10] Gras R., S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn et A. Totohasina (1996), *L'implication Statistique*, Collection Associée à Recherches en Didactique des Mathématiques, Grenoble : La Pensée Sauvage.

[11] Gras R., R. Couturier, J. Blanchard, H. Briand, P. Kuntz et P. Peter (2004), Quelques critères pour une mesure de qualité de règles d'association. Un exemple : l'implication statistique, *Mesures de qualité pour la fouille de données, RNTI-E-1, Cépaduès –Editions*, 3-32.

[12] Gras R. et R. Couturier (2010), Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association, *Quaderni di Ricerca in Didattica - GRIM (ISSN on-line 1592-4424)*, Eds : J.C. Régnier, R.Gras, F.Spagnolo, B. Di Paola, Université de Palerme, *p.19-57.*

[13] Gras R. et J.-C. Régnier (2013), Fondements théoriques de l'Analyse Statistique Implicative, *Méthode exploratoire et confirmatoire à la recherche de causalités*, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, p 25-186.

[14] Gras R., J.-C. Régnier et F. Guillet (2009), *Analyse Statistique implicative.Une méthode d'analyse de données pour la recherche de causalités,* sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse.

[15] Gras R., J.-C. Régnier, C. Marinica et F. Guillet (2013), *L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités*, Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8.

[16] Gras R., E. Suzuki, F. Guillet and F. Spagnolo (2008), *Statistical Implicative Analysis, Theory and Applications,* R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, eds, Springer.

[17] Guillet F. et H. Hamilton (2007), *Quality Measures in Data Mining* , F.Guillet et H.Hamilton eds, Springer.

[18] Hipp J., U. Guntzer and J. Nakhaeizadeh (2000), Mining association rules: Deriving a superior algorithm by analyzing today's approach , *Proc. of 4th Eur. Conf. on Principles of Data Mining and Knowledge Discovery, Lect. N. in Art. Int. 1910*, 160-168.

[19] Lallich S., O. Teytaud et E. Prudhomme (2007), Association Rule Interestingness: Measure and Statistical Validation, *F.Guillet and H. J.Hamilton eds, Studies in Computational Intelligence 43, Springer*, p. 251-275.

[20] Lenca P. et S. Lallich (2011), *Le choix d'une bonne mesure de qualité, condition du succès d'un processus de fouille de données*, Atelier Data Mining, Applications, Cas d'Etudes et Success Stories, Extraction et Gestion des Connaissances, *5-8.*

[21] Lerman I.-C. (1981), *Classification et analyse ordinale des données*, Paris : Dunod.

[22] Orús P., L. Zamora et P. Gregori (2009), *Teoria y Aplicaciones del Analisis Estadistico Implicativo,* Eds: P. Orús, L. Zamora y P. Gregori, Universitat Jaume I Castellon (Espagne), ISBN : 978-84-692-3925-4.

[23] Pearl J. (1988), *Probabilistic Reasoning in intelligent systems*, San Mateo, CA, Morgan Kaufmann.

[24] Régnier J.-C., M. Bailleul et R. Gras (2012), *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire*. Eds : J.C. Régnier, Marc Bailleul, Régis Gras, Université de Caen, ISBN : 978-2-7466-5256-9, 2012.

[25] Saporta G. (2006), *Probabilités, Analyse de Données et statistique*, Paris : Ed. Technip.

[26] Schektman Y., J. Trejos et M. Troupe (1992), Un générateur de règles floues à partir de bases de données volumineuse, *Actes des 3èmes Journées "Symboliques-Numériques", mai 1992*, Paris.

[27] Sève L. (2005), *Emergence, complexité et dialectique*, Odile Jacob, Paris.

[28] Vaillant B., S. Lallich and P. Lenca (2008), On the behaviour of the generalisations of the intensity of implication: a data-driven comparative study, *Statistical Implicative Analysis, Theory and Applications*, R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, Studies in Computational Intelligence, 127, Springer-Verlag Berlin Heidelberg.

[29] Vergnaud G. (2007), *Activités humaines et conceptualisation*, Presses Universitaires du Mirail, p.29.

[30] Xuan-Hiep Huynh, F. Guillet, J. Blanchard, P. Kuntz, H. Briand et R. Gras (2007), A Graph-based Clustering Approach to Evaluate Interestinguess Measures: A Tool and and a Comparative Study, *F.Guillet and H. J.Hamilton eds, Studies in Computational Intelligence 43, Springer*, p. 25-50.

ANNEXE

Lemme

Le comportement asymptotique de $\varphi(a,b)$, intensité d'implication, est celui d'une variable uniforme sur l'intervalle [0 ,1].

Notons $p=\varphi(a,b)$, p étant une probabilité, ne peut être considérée formellement comme une variable aléatoire qu'en changeant le modèle probabiliste jusqu'alors adopté. En effet, l'univers des possibles Ω introduit pour l'étude de l'étonnement à l'observation de la valeur $q(a,\bar{b})$, suppose que les valeurs également observées n_a et n_b soient des réalisations des cardinaux des variables aléatoires indépendantes X et Y. Si l'on veut alors mesurer un "étonnement" au sujet de la valeur de la cohésion, il faut considérer un autre univers des possibles Ω' sur lequel p suivrait une certaine loi de probabilité. Le problème revient alors à choisir la "bonne" loi de probabilité que suivrait p.

Nous allons la calculer pour n suffisamment grand, de telle façon que l'on puisse se considérer dans le cas asymptotique. Pour cela, on suppose que la valeur observée $n_{a \wedge \bar{b}}$ est la réalisation d'une variable aléatoire qui n'est plus Card $(X \cap \overline{Y})$, mais Card $(A \cap \overline{B})$, où A et B sont des sous-ensembles aléatoires indépendants dans E et fonctions d'éventualités $\omega \in \Omega' \neq \Omega$.

On obtient dans ces nouvelles conditions:

Prob [$p \leq \alpha$] = Prob Ω' [Prob Ω (Card $(X \cap \overline{Y}) \leq n_{a \wedge \bar{b}}$)$\leq \alpha$]

= Prob Ω' { ω' | Prob Ω { ω | Card $(X \cap \overline{Y})(\omega) \leq$ Card $(A \cap \overline{B})(\omega')$ } $\leq \alpha$}

= Prob Ω' { ω' | Prob Ω { ω | $N(\omega) \leq N'(\omega')$}$\leq \alpha$}}

où N et N' admettent des lois identiques, par exemple des lois normales centrées réduites et dans ce cas :

Prob [$p \leq \alpha$] = Prob Ω' { ω' | $\int_{-\infty}^{N'(\omega')} e^{\frac{-t^2}{2}} dt \leq \alpha$ }

= Prob Ω' { ω' | $N'(\omega') \leq F^{-1}(\alpha)$}, où F est la fonction de répartition de la loi normale,

$$= \int_0^{F^{-1}(\alpha)} e^{\frac{-t^2}{2}} dt = F(F^{-1}(\alpha))$$

$$= \alpha \qquad (1)$$

Ainsi, la loi de p est uniforme sur l'intervalle [0,1]. Cela signifie que, si n est très grand, un tirage au hasard de deux variables a et b étant donné, l'évènement [[$\varphi(a,b) \leq \alpha$]=α] ou [$\varphi(a,b) \geq \alpha$]=1-α] est presque sûr. De plus, la loi de l'intensité d'implication, en tant que variable aléatoire, tend vers une loi uniforme sur [0,1].

Exemple numérique fictif : 11 variables, 1081 objets (des véhicules)

On dispose de 11x10/2 soit 55 couples envisageables pour définir une règle de type a $\Rightarrow$ b respectant $n_a \leq n_b$. Le tableau des intensités d'implication donne des résultats n'invalidant pas la relation (1) :

- 5 intensités supérieures ou égales à 0.95 (1-a = 0.05) pour 5% x 55 = 2.75 intensités attendues d'après (1) ;
- 6 intensités supérieures ou égales à 0.90 (1-a = 0.10) pour 10% x 55 = 5.5 intensités attendues d'après (1) ;
- 8 intensités supérieures ou égales à 0.80 (1-a = 0.20) pour 20% x 55 = 11.5 intensités attendues d'après (1) ;

Ce qui statistiquement et pour ces trois mesures représente 19 résultats observés pour 19.75 attendus confortant notre lemme.

ANALYSE D'UN QUESTIONNAIRE « ENSEIGNANTS DE MATHEMATIQUES » PAR DIFFERENTES METHODES

Régis GRAS[1] et Antoine BODIN[2]

RÉSUMÉ

Les différents acteurs éducatifs (parents, enseignants, élèves) s'interrogent**,** particulièrement en mathématiques, sur les objectifs généraux de leur enseignement. Un groupe de travail de l'Association des Professeurs de Mathématiques a élaboré un questionnaire pour interroger les enseignants eux-mêmes sur ces objectifs tels qu'ils les perçoivent à travers ce qu'en attendent l'institution et la société. Les réponses fournies par un corpus de professeurs de filières différentes sont analysées par différentes méthodes de traitement de données dont l'A.S.I., l'A.C.P. et autres. Nous en comparons dans cet article les résultats obtenus en insistant sur l'émergence de deux variables latentes qui discriminent la population interrogée.

Mots-clés – *objectifs, analyse statistique implicative, analyse en composantes principales, analyse hiérarchique, nuées dynamiques, typicalité.*

ABSTRACT

The various agents of the Educational System try to better understand, what are, really, the general objectives of teaching mathematics. To do that, a research team from the Association of Mathematics Teachers set a questionnaire to ask the teachers themselves about what, according to them, is being expected by the educational institution and by the society at large. The answers provided by a corpus of teachers of different streams have been analyzed by different data processing methods including I.S.A., P.C.A. and some others. In this paper, we compare the results obtained with emphasizing the emergence of two latent variables that discriminate the population surveyed.

Keywords : *Objectives, Implicative Statistical Analyse, Principal Component Analysis, Hierarchical Clustering, k-means Clustering, Typicality.*

1 Introduction. Pourquoi un tel questionnaire ?

Vers la fin des années 90, un groupe de travail de l'Association des Professeurs de Mathématiques de l'Enseignement public (A.P.M.E.P.) a été chargé d'organiser une étude à grande échelle sur les acquis des élèves des classes terminales (17-18 ans) des lycées, toutes séries confondues. L'opportunité d'élaborer de telles épreuves ne s'était pas immédiatement inscrite dans le prolongement des épreuves des classes antérieures et, en particulier, de la classe de première (16-17 ans. Mais différents facteurs se sont conjugués pour que le comité de l'APMEP accepte de confier cette nouvelle (et lourde) tâche à l'Observatoire EVAPM (Évaluation des Programmes de Mathématiques. Cette étude apparaissait comme une suite logique de ce qui avait été présenté dans toutes les autres classes des premier et second cycles (de 11 à 18 ans). Pourquoi donc s'arrêter avant la terminale ? Les épreuves apparaîtront-elles comme un obstacle psychologique

[1] École Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241 : regisgra@club-internet.fr

[2] Institut de Recherche sur l'Enseignement des Mathématiques, Université d'Aix-Marseille, 163 Avenue de Luminy - Case 901, 13288 Marseille Cedex 09, antoinebodin@mac.com

au climat pré-baccalauréat ? Au contraire agiront-elles comme un stimulant et un récapitulatif de ce qui paraît essentiel de savoir et de savoir-faire ? Ne permettraient-elles pas, comme les évaluations précédentes, de mettre en évidence les obstacles communs rencontrés par les élèves, d'éclairer ainsi sur les difficultés liées aux programmes, aux limites de leur assimilation et aux conditions institutionnelles de leur réalisation ?

Parallèlement, mais de façon indépendante, un mouvement de fond avait fait émerger des critiques sévères et convergentes au sujet du baccalauréat lui-même, de son inadéquation au moins partielle à l'esprit dans lequel veut être pratiqué l'enseignement, tout au moins jusqu'en classe de première. Dès 1997, des réunions inter-associatives (APMEP, UPS, SMAI, SMF[3]), en présence de l'Inspection Générale de Mathématiques, avaient permis de dresser un bilan très réservé au sujet de certains acquis des élèves à la sortie des classes terminales et avaient responsabilisé, en partie, le baccalauréat dans les appauvrissements constatés par effet induit. En effet, personne ne contestait plus la relation de cause à effet de l'évaluation de sortie du lycée sur l'enseignement-même qui s'y pratiquait : à un examen peu adapté à la mesure de compétences recherchées répondrait un enseignement souvent rétréci à la préparation aux conditions dans lesquelles se présente l'examen. Sans oblitérer les objectifs spécifiques d'EVAPM (cf. plus haut) et leur opérationnalisation, les épreuves de ce type s'offraient en instrument d'évaluation des curricula des classes terminales et ne pouvaient que contribuer à objectiver les jugements portés sur l'enseignement conduit dans ces classes.

En préalable à l'étude proprement dite, donc à la rédaction des épreuves, ce groupe, dans le cadre de l'observatoire EVAPM, a décidé de sonder, par un questionnaire portant sur des objectifs généraux, les enseignants ayant accepté de faire passer les épreuves dans leurs classes. Le questionnaire proposé portait sur les représentations des enseignants dans les classes terminales de lycée quant à leur rôle dans la formation en mathématique à leur charge. Il devait permettre d'obtenir une typologie et une hiérarchie plus claires des attentes des professeurs. Que privilégient-ils ? Que jugent-ils essentiel ? Qu'est-ce qui distingue les opinions des enseignants des classes littéraire ou technique ou scientifique ? Y a-t-il des invariants, des relations stables entre les opinions, des éléments prédicteurs de gestes pédagogiques ? La présente étude approfondit une partie de l'analyse publiée en 1999 (Bodin et Gras, 1999) : d'une part elle se centre sur l'analyse de la première partie du questionnaire, d'autre part elle complète l'utilisation de l'analyse implicative et s'ouvre à d'autres méthodes d'analyse des données. Les méthodes d'analyse qui vont tenter de répondre aux questions posées ci-dessus, font en effet référence aux statistiques descriptives classiques, mais également à des méthodes multidimensionnelles qui globalisent tous les comportements de réponse afin d'en dégager les dimensions majeures et les structures sous-jacentes. On distingue de manière classique les variables manifestes et les variables latentes. Les premières, qui relèvent du champ empirique, sont directement observées et / ou mesurées. Les secondes qui relèvent d'un champ inférentiel, renvoient à des caractéristiques qui ne sont ni directement observables, ni directement mesurables, ce qui est le cas des

[3] UPS : Union des Professeurs de Spécialités.
SMAI : Société de Mathématiques Appliquées et Industrielles,
SMF : Société Mathématique de France

processus mentaux. En psychologie, on rendra compte des relations entre les données observées par le jeu de variables inférées de ces données. L'inférence peut se limiter à un versant qualitatif. Par exemple, on expliquera la covariation de l'âge mental et du poids de l'enfant par un processus de maturation et de croissance. Elle peut aussi s'exprimer sur le versant qualitatif. Par exemple, l'analyse factorielle des données observées permet de dégager des dimensions latentes dont les scores factoriels pourront être calculés pour chacun des individus. Dans le cadre de l'ASI, l'analyste met en évidence des chemins implicatifs entre les variables observées. On peut également calculer les scores (contributions et / ou typicalités) de chacun des individus sur ces dimensions implicatives. Pour l'analyse factorielle et pour l'ASI, la question centrale revient à l'interprétation et à la validation des facteurs ou des chemins produits de l'analyse. A une nuance près : pour l'analyse factorielle on cherche à interpréter les regroupements corrélationnels des variables observées ; à ce premier niveau symétrique, l'ASI en ajoute un second non symétrique, l'interprétation de l'ordre implicatif des variables. En d'autres termes, sont-ce les mêmes variables latentes qui expliquent à la fois regroupement et ordre des variables observées, ou bien faut-il considérer deux strates de variables latentes explicatives ?

2 Le questionnaire : objectifs de la formation mathématique

La consigne précédant le questionnaire était la suivante : *A votre avis, quels sont les objectifs essentiels_de la mission d'un professeur de mathématiques dans la série pour laquelle vous répondez. Pour répondre à cette question, classez par ordre préférentiel décroissant de 1 à 6 (1 : le plus important,...) six des objectifs majeurs de cette formation en les choisissant parmi les objectifs proposés ci-dessous :*

A- acquisition de connaissances
B- préparation à la vie professionnelle
C- préparation à la vie civique et sociale
D- préparation aux examens, concours, au passage dans l'enseignement supérieur
E- développement de l'imagination et la créativité
F- développement de la capacité à prouver et valider sa preuve
G- développement de la capacité d'accepter des points de vue différents
H- développement de la volonté et la persévérance
I- développement de l'esprit critique
J- développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers
K- développement de compétences utiles dans les autres disciplines
L- développement de la pratique de calculs formels, donc sans nécessité de signification
M- développement de la capacité à mathématiser et à formaliser
N- acquisition de savoir-faire
O- participation au développement d'une culture générale

Codage utilisé :
Exemple de réponse d'un enseignant x : en 1 : I ; en 2 : G ; en 3 : M ; en 4 : D ; en 5 : A ; en 6 : J)

Pour un traitement par le logiciel CHIC (Couturier et Ag Almouloud, 2009), ces variables pourraient être considérées comme des variables modales. C'est la stratégie que nous avions utilisée dans notre ancienne présentation de l'analyse complète par l'ASI de ce questionnaire (cf. Bodin et Gras, 1999). Cela nous permettait de conserver

l'information provenant des rangs choisis par chaque enseignant. Ici, nous nous contentons de coder 1 à chaque fois qu'un objectif est choisi par l'enseignant, quel que soit le rang qu'il lui a attribué ; et nous codons 0 lorsque la variable n'est pas classée.

On s'intéresse alors à trois sujets : le classement des items selon la fréquence de choix des items ; la structure implicative des items ; la structure des items à l'aide d'autres approches et comparaison avec la structure implicative.

Nombre de sujets ayant répondu au questionnaire :

311 professeurs ont répondu à ce questionnaire :

- 50 % enseignent en série Scientifique (S),
- 21 à 22 % enseignent en Economique et Sociale (ES) ou en diverses séries technico-professionnelles (TE),
- 7% enseignent en classe littéraire (LI).

2.1 Fréquence des choix :

Nous récapitulons ci-dessous (Tab. 1) les fréquences de choix quels que soient les rangs retenus.

item	n	fréquence
D préparation aux examens, concours, au passage dans l'enseignement supérieur	230	73,95%
F développement de la capacité à prouver et valider sa preuve	230	73,95%
J développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers	205	65,92%
M développement de la capacité à mathématiser et à formaliser	178	57,23%
I développement de l'esprit critique	164	52,73%
A acquisition de connaissances	161	51,77%
K développement de compétences utiles dans les autres disciplines	161	51,77%
N acquisition de savoir-faire	137	44,05%
H développement de la volonté et la persévérance	120	38,59%
O participation au développement d'une culture générale	101	32,48%
E développement de l'imagination et la créativité	81	26,05%
G développement de la capacité d'accepter des points de vue différents	50	16,08%
B préparation à la vie professionnelle	24	7,72%
C préparation à la vie civique et sociale	19	6,11%
L développement de la pratique de calculs formels, donc sans nécessité de signification	12	3,86%

Tableau 1. Fréquences des choix.

Les fréquences des choix sont franchement inégales, de 3,85% à 73,72%. Ces fréquences observées valident les constats précédents : « ...à un examen peu adapté à la mesure de compétences recherchées répondrait un enseignement souvent rétréci à la préparation aux conditions dans lesquelles se présente l'examen. »

Des compétences plus larges, transversales ou génériques utiles aux adaptations futures apparaissent comme des objectifs secondaires. Le contexte scolaire réduit les objectifs aux seules compétences scolaires et ferme l'ouverture à la citoyenneté, à la future vie sociale et professionnelle. On peut s'attendre, compte tenu de notre expérience d'enseignants- chercheurs, à une structure double des choix :

- Facteur 1 : l'enseignement des mathématiques se présente comme une discipline institutionnelle conservatrice
- Facteur 2 : l'enseignement des mathématiques se présente comme une occasion de libre créativité, d'inventivité et de préparation à la vie citoyenne.

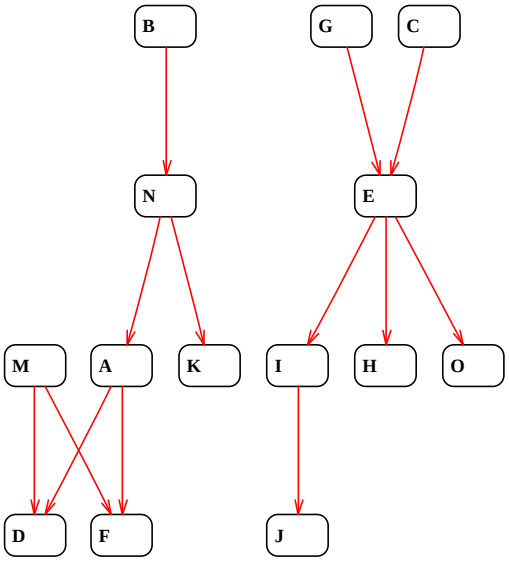
2.2 Structure implicative des choix :

On a demandé à CHIC de générer un graphe implicatif des items. Les variables supplémentaires correspondent aux différentes séries : S s pour série scientifique ; ES s pour série économique et sociale ; LI s pour classe littéraire ; TE s pour séries technico-professionnelles. On obtient la table d'implication ci-dessous (Tab. 2). Les intensités sont diversifiées et s'étendent de 0 à 100 pour le couple D - F.

item	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
A	0	48	22	**77**	28	**77**	25	11	15	0	14	45	29	**65**	35
B	37	0	**58**	29	**51**	29	37	7	17	4	**95**	37	6	**79**	36
C	1	**59**	0	0	**99**	0	33	**62**	**54**	**78**	21	46	7	4	**63**
D	**72**	46	16	0	0	**100**	24	29	9	13	**50**	**51**	**66**	**74**	7
E	18	**54**	**86**	0	0	0	**79**	**74**	**97**	**95**	46	49	24	16	**97**
F	**72**	46	16	**100**	0	0	24	29	9	13	**50**	**51**	**66**	**74**	7
G	9	41	40	4	**83**	4	0	58	50	**87**	6	47	2	3	24
H	6	27	**58**	16	**72**	16	**58**	0	20	39	2	47	2	1	40
I	15	38	**54**	3	**91**	3	**53**	26	0	**69**	0	35	5	2	49
J	0	29	**61**	9	**84**	9	**72**	45	**68**	0	13	41	6	6	47
K	14	**74**	42	45	**50**	45	22	4	0	8	0	45	13	**92**	35
L	23	32	45	38	40	38	43	32	6	12	23	0	**58**	36	19
M	30	30	33	**67**	35	**67**	14	5	6	4	15	**55**	0	**58**	5
N	**67**	**65**	27	**82**	25	**82**	14	2	1	2	**94**	49	**56**	0	12
O	27	46	**59**	0	**95**	0	34	36	45	41	27	41	1	7	0

Tableau 2. Table d'implication.

En utilisant le mode classique pour calculer avec CHIC l'intensité d'implication, le graphe implicatif au seuil 0,66 (par la méthode entropique dont on connaît la sévérité) met en évidence une dichotomie entre deux groupes d'items, tout en excluant l'item L (Fig. 1).



Le graphe situé à droite renvoie explicitement au facteur 2 : l'enseignement des mathématiques se présente comme une occasion de libre créativité, d'inventivité et de préparation à la vie **citoyenne**.

G- développement de la capacité d'accepter des points de vue différents ou C- préparation à la vie civique et sociale impliquent E- développement de l'imagination et la créativité et de là soit O- participation au développement d'une culture générale, soit H- développement de la volonté et la persévérance soit I- développement de l'esprit critique qui implique J- développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers.

Le graphe situé à gauche renvoie au facteur 1 : l'enseignement des mathématiques se présente comme une discipline institutionnelle conservatrice :

B- préparation à la vie professionnelle implique N- acquisition de savoir-faire et de là soit K- développement de compétences utiles dans les autres disciplines, soit A- acquisition de connaissances qui implique F- développement de la capacité à prouver et valider sa preuve ou D- préparation aux examens, concours, au passage dans l'enseignement supérieur. M- développement de la capacité à mathématiser et à formaliser implique également soit F- développement de la capacité à prouver et valider sa preuve, soit D- préparation aux examens, concours, au passage dans l'enseignement supérieur.

On examine maintenant les typicalités des différents chemins (Tab. 3).

facteur	chemin	S	ES	LI	TE
2	C-E-I-J			X	
	C-E-O			X	X
	C-E-H			X	
	G-E-H		X		
	G-E-O		X	X	
	G-E-I-J				
1	B-N-A-D				X
	B-N-A-F				X
	B-N-K		X		X
	M-D		X	X	
	M-F		X	X	

Tableau 3. Typicalités des chemins**.**

S n’est typique d’aucun des chemins.

ES est typique de deux des départs G-E-x (F2) : développement de la capacité d'accepter des points de vue différents => développement de l'imagination et la créativité.

ES est typique de l’un des départs B-N-x (F1), soit B-N-K : préparation à la vie professionnelle => acquisition de savoir-faire => développement de compétences utiles dans les autres disciplines.

ES est enfin typique des départs M-x (F1) : développement de la capacité à mathématiser et à formaliser.

LI est typique des départs C-E-x (F2) : préparation à la vie civique et sociale => développement de l'imagination et la créativité.

LI est typique de l’un des départs G-E-x (F2), soit G-E-O développement de la capacité d'accepter des points de vue différents => développement de l'imagination et la créativité => participation au développement d'une culture générale.

Enfin, comme ES, LI est enfin typiques des départs M-x (F1) : développement de la capacité à mathématiser et à formaliser.

TE est typique des départs C-E-O (F2) : préparation à la vie civique et sociale => développement de l'imagination et la créativité => participation au développement d'une culture générale.

TE est typique des départs B-N-x : préparation à la vie professionnelle => acquisition de savoir-faire.

Il serait intéressant, et certainement enrichissant, d’étudier les relations contributives entre les modalités de choix des enseignants. Il suffirait pour cela de considérer, techniquement possible avec CHIC, les variables principales, comme des variables supplémentaires. Dans cet article, la place manque pour pousser l’étude plus loin.

2.3 Structures non-implicatives des choix

La question est ici de savoir si des analyses classiques font ressortir cette structure en deux grands facteurs. On utilisera successivement la classification hiérarchique, les nuées dynamiques et l'analyse en composantes principales.

2.3.1 La classification hiérarchique :

Par le moyen d'une classification hiérarchique utilisant la distance moyenne entre classes on obtient deux classes relativement proches du résultat de l'ASI (Fig. 2).

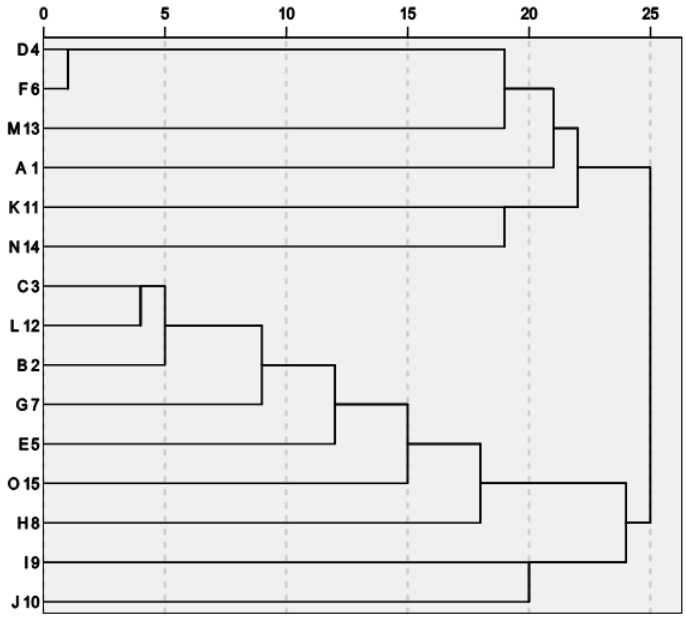


Figure 2. Arbre hiérarchique utilisant la distance moyenne entre classes.

Soit pour le facteur 2 : C- préparation à la vie civique et sociale, G- développement de la capacité d'accepter des points de vue différents, E- développement de l'imagination et la créativité, O- participation au développement d'une culture générale, H- développement de la volonté et la persévérance, I- développement de l'esprit critique et J- développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers. L- développement de la pratique de calculs formels, donc sans nécessité de signification non classé par l'ASI rejoint ce groupe et B- préparation à la vie professionnelle joue le transfuge.

Et soit pour le facteur 1 : D- préparation aux examens, concours, au passage dans l'enseignement supérieur, F- développement de la capacité à prouver et valider sa preuve, M- développement de la capacité à mathématiser et à formaliser, A- acquisition de connaissances, K- développement de compétences utiles dans les autres disciplines et N- acquisition de savoir-faire.

2.3.2 La classification en nuées dynamiques :

Là encore, pour la classification en nuées dynamiques, les classes se rapprochent de celles obtenues par l'ASI, à une variable près (Tab. 4).

items	classe 1	classe2
A	1	0
B	1	0
D	1	0
F	1	0
K	1	0
M	1	0
N	1	0
C	0	1
E	0	1
H	0	1
I	0	1
J	0	1
O	0	1
G	0	0
L	0	0

Tableau 4. Classification en nuées dynamiques.

On retrouve intégralement la classe 1. L n'est pas classée. Mais G non plus.

2.3.3 L'analyse factorielle en composantes principales :

H, L et G restent isolées. On extrait une structure en cinq facteurs, dont quatre bipolaires. Ces derniers opposent des items appartenant aux deux classes du graphe implicatif issu de l'ASI (Tab. 5).

items	Classes ASI	F1	F2	F3	F4	F5
D	1	0,99				
F	1	0,99				
E	2	-0,99				
K	1		0,73			
N	1		0,55			
I	2		-0,72			
A	1			0,80		
J	2			-0,70		
O	2				0,69	
M	1				-0,79	
B	1					0,80
C	2					0,48

Tableau 5. Analyse factorielle en composantes principales.

F1 *opposition logique, imagination* **:** F développement de la capacité à prouver et valider sa preuve et D préparation aux examens, concours, au passage dans l'enseignement supérieur *vs* E développement de l'imagination et la créativité.

F2 *opposition utilités, esprit critique* : K développement de compétences utiles dans les autres disciplines et N acquisition de savoir-faire *vs* I développement de l'esprit critique

F3 *opposition savoirs, communication* **:** A acquisition de connaissances *vs* J développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers

F4 *opposition culture générale, discipline* **:** O participation au développement d'une culture générale *vs* M développement de la capacité à mathématiser et à formaliser

F5 *préparation à la vie citoyenne* **:** B préparation à la vie professionnelle et C préparation à la vie civique et sociale

2.3.4 Introduction des variables supplémentaires (sections d'enseignement)

Pour chacune des sections d'enseignement, on a calculé la signification de la différence pour chacun des facteurs (Tab. 6).

	S	ES	LI	TE
F1				
F2	X -	X +		X +
F3				
F4	X -	X +	X +	
F5	X -			X +

Tableau 6. Résumé de l'ANOVA par section.

On ne relève aucune différenciation pour les facteurs F1 opposition logique, imagination et F3 opposition savoirs, communication.

Les enseignants des sections S, scientifiques, ont des scores différenciés pour les trois autres facteurs, ces sections présentant un score négatif sur ces facteurs, les autres sections un score positif. Pour F2 *opposition utilités, esprit critique* (-0,35 *vs* 0,35) ces enseignants valorisent le développement de l'esprit critique. Pour F4 *opposition culture générale, discipline* (-0,24 *vs* 0,24) ils valorisent le développement de la capacité à mathématiser et à formaliser. Pour F5 ils n'adhèrent pas à la *préparation à la vie citoyenne* (-0,15 *vs* 0,15).

Les enseignants des sections ES, économique et sociale, ont des scores différenciés sur les facteurs F2 et F4, y présentant des scores moyens positifs, à l'inverse des S. Pour F2 *opposition utilités, esprit critique* (0,37 *vs* -0,10) ces enseignants valorisent le développement de compétences utiles dans les autres disciplines et l'acquisition de savoir-faire. Pour F4 *opposition culture générale, discipline* (0,30 *vs* -0,08) ils valorisent la participation au développement d'une culture générale.

Les enseignants des sections LI, littéraire, ne se différencient que sur F *opposition culture générale, discipline* (0,39 vs -0,03) en privilégiant la participation au développement d'une culture générale.

Enfin, les enseignants des sections TE, technologiques, se différencient sur F2 et F5. Ils rejoignent les ES pour le facteur 2 *opposition utilités, esprit critique* (0,42 *vs* -0,11) valorisant le développement de compétences utiles dans les autres disciplines et l'acquisition de savoir-faire. Ils rejoignent également leurs collègues LI pour F5 opposition culture générale, discipline (0,46 *vs* -0,12) privilégiant la participation au développement d'une culture générale.

On note la spécificité des enseignants de la section scientifique S étroitement centrés sur les aspects formels *a contrario* de leurs collègues des autres sections plus ouvert sur la culture générale et plus largement sur la préparation à la vie professionnelle et citoyenne.

3 Conclusion

Ce papier tente de cerner et d'identifier les liens entre ASI et autres méthodes d'analyses en classes ou en facteurs. Par rapport aux données utilisées, on observe que chaque méthode de calcul pratiquée fait ressortir une opposition entre deux groupes d'items, soit sous forme de deux classes, soit sous forme de facteurs bipolaires. Cette quadruple approche méthodologique présente donc un vif intérêt alors que les outils d'analyse sont très différents. On peut inférer de cette partition en deux classes l'existence d'un processus mental qui opère ces classifications selon deux conceptions distinctes : préserver l'enseignement des mathématiques en tant que discipline institutionnelle conservatrice, conception caractérisant les professeurs de section S, gardiens du temple de l'orthodoxie scientifique ; utiliser le « prétexte » de l'enseignement des mathématiques comme occasion de libre créativité, d'inventivité et de préparation à la vie citoyenne. Mais n'y a-t-il pas variation du rapport institutionnel des enseignants en fonction de la situation (référence à la théorie anthropologique du didactique).

Toutefois, on ne sait pas si ce clivage fermeture / ouverture au principe de la relation d'appartenance à l'une des deux classes commande également la relation d'ordre implicatif structurant ces deux pôles. Il devient de plus en plus nécessaire de se donner les moyens d'avancer sur cette problématique des déterminants de l'appartenance et de l'ordre propres aux facteurs implicatifs

Références

[1] Bodin A., Gras R. (1999) : Analyse du préquestionnaire enseignants avant EVAPM-Terminales, *Bulletin n°425 de l'Association des Professeurs de Mathématiques de l'Enseignement Public, 772-786, Paris*

[2] Couturier R. et Ag Almouloud S. (2009), Historique et fonctionnalités de CHIC, *Analyse Statistique Implicative, Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd, invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse p.279-293*

[3] Pasquier D. et Gras R., [2012], De l'intérêt de l'Analyse Statistique Implicative (A.S.I.) pour la recherche exploratoire en psychologie**,** *Psychologie Française, Elsevier-Masson***,** p 161-173

[4] Gras R., Régnier J.-C. [2013], Fondements théoriques de l'Analyse Statistique Implicative, *Méthode exploratoire et confirmatoire à la recherche de causalités*, sous la direction de Gras R., eds Gras R., Régnier

J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, p 25-186.

[5] Pasquier D. et Gras R., [2012], Analyse exploratoire a priori et analyse confirmatoire a posteriori : paradigme pour la comparaison de deux structures en analyse statistique implicative, *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire*, Eds J.C.Régnier, M.Bailleul R.Gras, Université de Caen, ISBN 978-2-7466-5256-9, p.245-261.

[6] Pasquier D., Gras R. [2013], Analyse Statistique Implicative et psychométrie, *Méthode exploratoire et confirmatoire à la recherche de causalités*, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, p. 365-378.

Ouvrages de référence

[1] *L'implication statistique. Nouvelle méthode exploratoire de donnée*, sous la direction de R.Gras, et la collaboration de S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn, A.Totohasina, La Pensée Sauvage, Grenoble (1996)

[2] *Mesures de Qualité pour la Fouille de Données*, H.Briand, M.Sebag, R.Gras er F.Guillet eds, RNTI-E-1, Cépaduès, 2004

[3] *Quality Measures in Data Mining* , F.Guillet et H.Hamilton eds, Springer, 2007,

[4] *Statistical Implicative Analysis, Theory and Applications*, R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, eds, Springer, 2008.

[5] *Analyse Statistique implicative.Une méthode d'analyse de données pour la recherche de causalités*, sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, 2009.

[6] *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, Eds : P.Orus, L.Zemora, P.Gregori, Universitat Jaume-1, Castellon (Espagne), ISBN : 978-84-692-3925-4, 2009..

[7] *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire*. Eds : J.C. Régnier, Marc Bailleul, Régis Gras, Université de Caen, ISBN : 978-2-7466-5256-9, 2012

[8] *L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités*, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, 2013.

VARIABLE NODALE ET CONE IMPLICATIF

Dominique LAHANIER-REUTER[1], Régis GRAS[2], Marc BAILLEUL[3]

NODUL VARIABLE AND IMPLICATIVE CONE

RÉSUMÉ.

Dans le cadre de l'Analyse Statistique Implicative, il est procédé à l'extraction de règles implicatives pondérées par une mesure appelée intensité d'implication. Un graphe dit implicatif permet de représenter l'ensemble des relations non symétriques qui associent les variables. Au sein du graphe, on observe quelquefois, à partir d'une variable centrale, des relations amont et aval telles qu'il soit possible d'isoler ces deux types de part et d'autre de ce nœud, ce confluent, sous la métaphore de cône implicatif à deux grappes. Donner une signification à ce cône, sous la condition d'une certaine homogénéité et de la connexité de l'ensemble, permet de l'isoler conceptuellement du tout en des termes de causalité-conséquences. Dans cette communication, nous définissons selon deux approches un critère d'homogénéité d'un cône et des conditions d'existence de relations causales. Nous illustrons le propos théorique par l'examen de plusieurs exemples significatifs.

Mots-clés. *Analyse statistique implicative, règle, graphe, nœud, cône, sommet du cône, arête, cause et conséquence,*

ABSTRACT

Inside the Statistical Implicative Analysis frame, implicative rules may be extracted while pondered by a measure called Intensity of Implication. The set of the non-symmetrical relationships that associate the different variables may be represented by a graph so called implicative. Inside the graph, relationships upstream and downstream from a central variable parted on either sides of this node can be sometimes noticed. This situation of a confluent is metaphorically said a two clusters implicative cone. On condition that the set is rather homogeneous and linked, giving conceptual meaning to this cone allows some kind of causalities/consequences isolation of it from the whole set. In this communication, we define criteria for the homogeneity of a cone and conditions for existence of causal relationships. Some significant examples illustrate our theoretical discourse.

Keywords *Statistical Implicative Analysis, Rule, Graphic, Node, Cone, Edge, Top of the cone, Causes and consequences*

1 Introduction

Disposant d'un corpus de données croisant sujets et variables, l'Analyse Statistique Implicative offre des réponses à certains objectifs du chercheur :

– Extraire de cet ensemble de données où les sujets sont éléments d'un ensemble discret E et les variables sont de natures diverses, des règles implicatives par

[1] Équipe Théodile CIREL, Université Lille 3, Villeneuve d'Ascq, France, E-mail: dominique.lahanier@univ-lille3.fr

[2] École Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, : regisgra@club-internet.fr

[3] CERSE (Centre d'Etudes et de Recherches en Sciences de l'Education), Université de Caen Basse-Normandie, courriel : marc.baillleul@unicaen.fr

exemple du type : « si la variable a est instanciée alors généralement la variable b l'est aussi », dans le cas de variables booléennes,

- Affecter à chaque règle une valeur numérique de [0,1] croissante avec la qualité prédictive de la règle,
- Représenter par un graphe dit implicatif, non symétrique, pondéré et sans cycle, l'ensemble des règles de qualité au moins égale à un certain seuil d'acceptabilité,
- Interpréter en donnant du sens aux arêtes du graphe, mais aussi aux structures connexes qui le constituent.

C'est ainsi qu'il existe, pour un seuil d'intensité d'implication donné, certaines sous-structures connexes où des nœuds du graphe admettent isolément des antécédents (des « pères ») et des successeurs (des « fils »). Ces sous-structures sont extraites « à la main » par l'utilisateur lui-même au vu du graphe implicatif donné par CHIC. Antécédents et successeurs peuvent éventuellement être interprétés respectivement comme causes et conséquences du sommet du graphe dans une perspective causale.

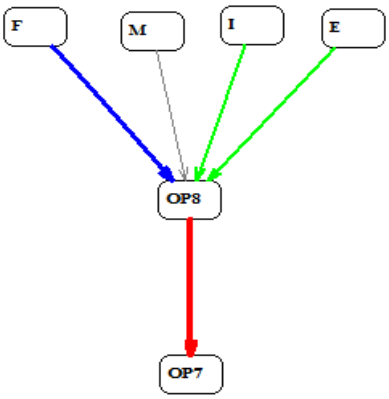


Figure 1

Par exemple, sur la figure 1, on a extrait, d'un graphe implicatif d'une quarantaine de variables et au seuil 0.69, la sous-figure ci-dessus que nous appelons *« cône implicatif »*, avatar géométrique car il admet un sommet et deux nappes dont l'amont et l'aval représentent respectivement les « pères » et les « fils ». Le sommet de ce cône, OP8, admet pour « pères » (pluriel, comme c'est bizarre !) ou « causes » les variables F, M, I et E. Il n'a qu'un « fils » ou « conséquence : OP 7. Ce sommet joue un rôle que l'on peut désigner comme « confluent » ou « variable nodale» de variables. Le mot « cône » a été retenu car il symbolise aussi, comme un entonnoir, un objet de type déversoir » où confluent divers affluents et d'où repartent de nouveaux réseaux absorbant les flux de l'amont. Jean-Marc Lévy-Leblond, dans (J.-M. Lévy-Leblond, 1996, p. 320) parlerait même d'un **champ causal** ou un écheveau enchevêtré qui « ne laisse plus individualiser les fils continus au long desquels pourrait se concevoir la propagation d'influences causales ».Au sujet des mots constitutifs de la phrase, François Gaudin dans (Sève, L., 2005) dirait que l'on voit « s'y dessiner une topologie du sens ».

Ce cône pouvant se prolonger le long de ses deux nappes, on constate que la notion de « cause » (ou de « pères ») et de « conséquence » (ou « d'effet » ou de « fils ») est toute relative. Un père n'est père qu'en tant qu'il a un fils. Nous rencontrons ici aussi le raisonnement logique de type dialectique où l'analyse a pour fonction la résolution de la contradiction. La cause détermine l'effet tout autant que l'effet détermine la cause (cf L.Sève, 2005, p.71). Son sens ne se construit que dans son rapport avec l'effet.

Notre intention est d'utiliser ce cône pour mettre en évidence un ensemble de causes non liées entre elles afin de pouvoir identifier leurs rôles, leurs responsabilités dans l'apparition de l'attracteur. Autrement dit, nous nous intéresserons de façon privilégiée aux cônes dont les « pères » du sommet sont **indépendants ou très faiblement dépendants**. Ce sommet pourrait définir une sorte de « bassin d'attraction » au sens de la théorie des catastrophes. Sous cette restriction *d'indépendance*, nous nous limiterons donc ici, un graphe étant donné comme image d'un ensemble de règles, à ce type de structure conique et à son interprétation. Mais celle-ci exige que ce cône ait des chances d'offrir une certaine consistance sémantique et, pour cela, que sa constitution présente des caractères analytiques révélateurs d'une certaine homogénéité implicative. *N'y observerait-on pas un phénomène comparable à celui que L. Sève désigne par « propriété émergente » ?* Cette problématique constituera notre premier paragraphe.

2 Indice d'homogénéité d'un cône implicatif

2.1 Rappels

Au cours de la formalisation de l'implication statistique, nous avons introduit un indicateur de qualité de la mesure d'une règle de type a=>b. Dans le cas des variables binaires, nous avons comparé les occurrences des contre-exemples à la règle à celles qui serait observables au hasard si les variables étaient indépendantes. Pour quantifier cette proposition, aux sous-ensembles A et B de E des observations de sujets effectives et respectives de a et b, nous avons associé des parties X et Y de mêmes cardinaux que A et B. En modélisant un tirage au hasard dans E des parties X et Y, nous avons comparé le cardinal aléatoire $N_{a\wedge\bar{b}}$ des contre-exemples à l'implication (soit $\mathrm{card}(\overline{X\cap Y})$), au cardinal $n_{a\wedge\bar{b}}$ (soit $\mathrm{card}(\overline{A\cap B})$) des contre-exemples observés. L'intensité d'implication qui mesure la qualité de l'implication est alors la probabilité que le hasard conduise à plus de contre-exemples que la contingence soit :

$$\varphi(a,b) = = \Pr\left[\mathrm{Card}(X\cap\overline{Y}) > n_{a\wedge\bar{b}}\right] = 1-\Pr\left[\mathrm{Card}(X\cap\overline{Y}) \le n_{a\wedge\bar{b}}\right] \quad (1)$$

Si l'on centre et l'on réduit la variable aléatoire $\mathrm{Card}(X\cap\overline{Y})$ sous la forme :

$$Q(a,\bar{b}) = \frac{\mathrm{Card}(X\cap\overline{Y}) - \frac{n_a\, n_{\bar{b}}}{n}}{\sqrt{\frac{n_a\, n_{\bar{b}}}{n}}} \quad (2)$$

Celle-ci dans des conditions à grands effectifs est une variable gaussienne N(0,1) en tant que valeur approchée d'une variable aléatoire de Poisson centrée et réduite (cf.

Partie 1, Chapitre 1 de Gras et Régnier, 2013). Sa réalisation contingente est $q(a,\bar{b})$. Par suite, l'intensité d'implication devient dans ce cas gaussien :

$$\varphi(a,b) = 1 - \Pr\left[Q(a,\bar{b}) \le q(a,\bar{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt \quad (3).$$

2.2 Rôle fonctionnel des intensités d'implication des arêtes

Une première approche pourrait consister à déterminer la *variance implicative* intra-classe du cône implicatif (cf. Partie 2, Chapitre 2 de Gras et Régnier, 2013). Cette variance du cône, basée sur une interprétation géométrique de l'implication statistique, offre une mesure qui permet d'apprécier la consistance d'une grappe connexe du graphe comme l'est un cône implicatif. L'absence de programmation des calculs associés dans CHIC nous incite, puisque nous visons des calculs conduits à la main, à retenir une autre méthode plus intuitive et aisée sur le plan calculatoire.

Une deuxième approche, plus simple et plus intuitive, consiste à examiner la consistance d'un tout à partir de celles de l'ensemble de ses arêtes, même si un tout est plus informatif que l'ensemble de ses parties. Ainsi, un cône implicatif étant donné, une mesure naturelle pour en qualifier la qualité d'homogénéité se nourrit de l'ensemble des qualités implicatives des arêtes. Il est alors possible de mesurer cette qualité à travers, par exemple, la moyenne arithmétique ou la moyenne géométrique (afin de valoriser une intensité nulle pour viser un rejet) de toutes les intensités d'implication des arêtes. Cependant, cette méthode simple a le défaut de ne pas accorder à la mesure un caractère relatif comme nous l'obtenons pour l'intensité $\varphi(a,b)$ qui est une probabilité.

Aussi, nous faisons appel à une troisième approche qui a la vertu épistémologique de s'harmoniser avec la philosophie de la modélisation employée pour évaluer la qualité d'une quasi-implication. Comme l'intensité d'implication, elle présente aussi l'avantage de fournir une échelle de mesure (une probabilité, nombre de [0 ,1]), c'est-à-dire une valeur ordinale où le « zéro » sera signe d'une hétérogénéité maximale du cône et le « un » sera signe d'une excellente homogénéité dissymétrique ou d'un flux implicatif régulier et riche dans ses liaisons.

Associons à chacune des p arêtes une variable aléatoire relative à l'implication a => b ; on la note $Q(a,\bar{b})$ (formule (2) du §2.1). Elle conduit, nous l'avons vu, à une mesure de la qualité de l'implication entre les deux extrémités de l'arête a ---> b. Notons de façon générique, Q_i , i = 1 à p, l'indice d'implication aléatoire dont la réalisation est q_i. Nous savons que les p variables Q_i , supposées indépendantes, suivent approximativement la loi normale N(0,1). Or, ces p variables indépendantes, suivant la même loi satisfont, en conséquence, les hypothèses du théorème central limite.

Posons $S_p = \sum_1^p Q_i$ et sa forme centrée-réduite $T_p = \frac{S_p - E(S_p)}{\sigma(S_p)}$ où $E(S_p)$ est l'espérance de S_p, somme des espérances des variables Q_i qui sont toutes nulles et $\sigma(S_p)$ est l'écart-type de S_p, soit $\sqrt{p}$.

Par suite, $T_p = \frac{S_p}{\sqrt{p}}$ converge en loi vers la variable aléatoire de loi N(0,1). Pour une valeur de p convenable, par exemple $p \geq 4$ [4], T_p est donc approximable par la variable gaussienne centrée réduite N(0,1).

Comparant cette valeur attendue à la valeur observée $t_p = \frac{\sum_1^p q_i}{\sqrt{p}}$, nous obtenons alors un indice probabiliste pour mesurer, sur son versant d'entrée, la qualité du cône à savoir : $\Pr[T_p \geq t_p] = \frac{1}{\sqrt{2\pi}} \int_{t_p}^{\infty} e^{-\frac{u^2}{2}} du$

3 Recherche des conditions de causalités sans liaison entre elles

Si l'on veut identifier les « pères » du sommet du cône comme étant des causes ne dérivant pas de phénomènes causaux communs ou proches, comme le seraient par exemple « il pleut » et « le sol est mouillé », il nous faut exiger que ces variables amont ne soient pas liées ou ne le soient que faiblement. Si tel est le cas, on pourra affirmer que telle variable (par ex. le sommet du cône en l'occurrence) dépend essentiellement de telles causes séparées, tout au moins dans les conditions expérimentales sources des données traitées. Ces données qu'elles soient quantitatives ou qualitatives, pour un traitement avec CHIC, sont numériques ou binaires ou décimales. Envisager une étude la corrélation linéaire de Bravais-Pearson est donc justifié puisque le rho de Spearman n'aurait été indiqué que si nous avions conservions les rangs des valeurs des variables. Or, dans l'exemple initial, la relation d'ordre des variables ordinales –des préférences- a été commuée en relation d'ordre sur R. Aussi, nous écartons le rho de Spearman.

Un test de significativité du coefficient de corrélation linéaire peut alors être joint à l'analyse interprétative pour réfuter la liaison entre deux variables amont. Par exemple, utilisant ce test, si le nombre de sujets est 150, on dira que la liaison est significative au risque de 5% si on observe une corrélation r supérieure à 0.16. Mais si r = 0.10, on admettra l'absence de liaison entre les deux variables considérées. Il est possible aussi de pratiquer un test de corrélation partielle[5] entre d variables « pères ». Pour d= 5, le risque de 5% est atteint pour une corrélation partielle de 0.164.

Il en est de même pour les « fils » en aval du sommet qui seront désignées comme conséquences non liées entre elles de ce sommet.

[4] Pour un nombre d'arêtes plus petit, l'indice envisagé risque de n'avoir qu'un rôle indicateur de tendance.

[5] Rappelons que la corrélation partielle se calcule de façon récurrente. Par exemple, si d=3 : $r_{1,2,3} = \frac{r_{1,2} - r_{1,3}\, r_{2,3}}{\sqrt{(1-r_{1,3}^2)(1-r_{2,3}^2)}}$. Pour obtenir $r_{1,2,3,4}$, on remplace dans la formule précédente les corrélations simples par les corrélations partielles.

4 Retour sur l'exemple initial

4.1 Les données

Reprenons l'exemple du § 1 Fig. 1. Ce cône est obtenu par extraction au sein d'un graphe implicatif de données issues d'un questionnaire présenté à des professeurs de mathématiques de classes terminales de lycée (17-18 ans). Nous avons recueilli et analysé (Bodin et Gras., 1999) les réponses de 311 professeurs, à des classements attendus (de 1 à 6) portant sur quinze objectifs que ces professeurs assignent à leur enseignement (A, B, C, …O)[6] et sur leurs opinions relatives à dix phrases susceptibles d'être communément énoncées (OP1, OP2,..OPX)[7]. Les 26 variables correspondantes ne sont pas binaires, mais ordinales (valeurs (1, 0.8, 0.6, 0.4, 0.2, 0.1, 0) pour les objectifs et (1, 0.5, 0) pour les opinions). Ainsi l'analyse intègre l'intensité des attitudes, d'un choix prioritaire d'un objectif à un choix plus secondaire, voire non retenu.

La consigne était la suivante :

« A votre avis, quels sont les objectifs essentiels de la mission d'un professeur de mathématiques dans la série pour laquelle vous répondez. Pour répondre à cette question, classez par ordre préférentiel décroissant de 1 à 6 (1 : le plus important,...) six des objectifs majeurs de cette formation en les choisissant parmi les objectifs proposés ci-dessous :

A- acquisition de connaissances

……

E- développement de l'imagination et la créativité

F- développement de la capacité à prouver et valider sa preuve…….

I- développement de l'esprit critique…….

M- développement de la capacité à mathématiser et à formaliser……..

Puis, Voici quelques opinions entendues dans la salle des profs. Vous pouvez être d'accord, ou un peu d'accord ou pas d'accord avec l'une ou l'autre. Entourez votre choix :

1-(OP1) C'est vrai que les math constituent un instrument de sélection excessif.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

……..

A la sortie de la terminale de la série sur laquelle vous répondez, un élève devrait....

7-(OP7) ....pouvoir reconnaître si un nombre entier écrit dans la base 10 est divisible par 4.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

8-(OP8) ....pouvoir donner un exemple ou un contre-exemple personnels à l'affirmation :

"si deux applications *f* et g sont strictement croissantes sur un intervalle, l'application produit *fxg* y est également croissante".

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

[6] Par exemple, E symbolise l'objectif : « développement de l'imagination et de la créativité »

[7] Par exemple, OP4 symbolise : « Pour corriger, j'aime bien un barème très détaillé sur les résultats à obtenir »

4.2 L'homogénéité du cône

Le logiciel CHIC nous fournit les valeurs des intensités d'implication des variables en jeu dans le graphe de la Fig. 1. De ces valeurs de l'intensité d'implication qui sont des probabilités donc des nombres de [0,1] affichées par CHIC, nous pouvons remonter à la valeur q_i prise par Q_i où i est l'un des couples qui nous intéressent du cône implicatif. Il suffit de lire la table de la loi normale N(0,1) de façon inverse.

Règle i	Intensité d'implication φ_i	Indice d'implication observé q_i
F -> OP8	0.96	- 1.75
M -> OP8	0.86	- 1.08
I -> OP8	0.93	- 1.48
E -> OP8	0.91	- 1.34
F -> OP7	0.81	- 0.88
M -> OP7	0.61	- 0.28
I -> OP7	0.75	- 0.68
E -> OP7	0.81	- 0.88
OP8 -> OP7	1	- 4.00

Tableau 1

Appliquons les relations établies dans le 2.2 pour p = 9, $t_9 = \dfrac{\sum_1^9 q_i}{\sqrt{9}} = -4.13$

D'où $\Pr[T_9 \geq t_9] = \dfrac{1}{\sqrt{2\pi}} \int_{t_p}^{\infty} e^{-\frac{u^2}{2}} du \cong 1$

Ceci signifie que la qualité d'homogénéité sémantique et dissymétrique du cône, eu égard à notre modèle aléatoire, est excellente puisqu'une distribution au hasard des indices d'implication conduirait à des sommes de Q_i bien supérieures à celle qui est observée à savoir – 4.13. Autrement dit, le « flux implicatif » qui traverse le cône respecte bien la dissymétrie attendue de l'implication et cela dans le même sens puisque les valeurs des q_i sont toutes négatives.

Remarquons, en passant, que la liaison OP8 => OP7 donne du sens à OP7 comme engageant la réflexion personnelle et critique de l'élève et non pas l'application d'un critère appris en cours puisque c'est précisément ce qu'exprime l'objectif OP8. La compétence de l'élève qui lui est associée apparaît comme « émergente » de la conjugaison des objectifs qui seraient réalisés en amont. Cette interprétation est renforcée par la nature de I, une des « causes » amont où l'esprit critique est explicitement visé au lycée. Cette remarque illustre le propos tenu dans l'avant-dernier paragraphe du § 1.

4.3 Les liaisons entre les « causes » et entre les « conséquences »

Dans cet exemple, seules les éventuelles « causes » sont plurielles : F, M, I et E. Mais nous n'avons qu'une conséquence : OP7. Les coefficients de corrélation linéaire - à la limite d'applicabilité en raison des faibles effectifs en jeu- entre les variables « pères » sont les suivantes (toujours données par CHIC) :

r (F ; M) = 0.13 ; r (F ; I) = - 0.04 ; r (F ; E) = 0.1 ;

r (M ; I) = - 0.1 ; r (M ; E) = - 0.07 ; r (I ; E) = 0.15

Faisant référence à la table des valeurs critiques du coefficient de corrélation d'un échantillon issu d'une population normale, nous constatons qu'un coefficient au hasard R, pour tous les couples en jeu, dans le cas d'une liaison entre ses termes, est supérieur à 0.13 avec un risque de 5%. Ceci est le cas pour (F et M) et (I et E), mais non pour les autres couples dont l'affirmation d'une liaison conduirait à un risque supérieur à 10%. On peut donc faire l'hypothèse d'ordre cognitif que deux causes fondamentales se conjuguent pour impliquer OP8 puis OP7 : l'un de type non normatif dans le cadre scolaire (imagination E, esprit critique I), l'autre de type cartésien (preuve F, formalisation M), à l'intérieur desquels des modalités seraient liées. L'examen des tâches nécessaires à la résolution de OP8 (exemple, contre-exemple, concepts mathématiques) montre que celle-ci nécessite la mise en commun des deux types causaux identifiés en tant que « pères ».

Cependant, l'examen des corrélations partielles montre que admettre l'absence de liaison globale reste légitime et donc que les « causes » présentent des caractères spécifiques non inclusives. En effet, dans les limites de seuil de 5% (0.164) et même à 10% (0.137), on obtient des corrélations partielles en permettant la validation :

r(F, M, I) = 0.127 ; r(M, I, E)) _0.092 ; r(F, I, E) = - 0.056 ; r(F , M, E) = 0.143 ;

et finalement r(F, M, I, E) = 0.131.

5 Un deuxième exemple

Relations et Co variations

D'un point de vue statistique la problématique principale que nous abordons est celle de la définition des liens statistiques entre variables lorsqu'ils sont déduits des mises en relation de leurs variations. Il s'agit de donner un sens statistique aux questions suivantes : si les valeurs d'une des variables différencient deux groupes de sujets (si ces groupes de sujets sont affectés par un changement de valeurs de cette variable ou par une modification importante de ces valeurs) en est-il de même pour l'autre variable? Et ces changements sont-ils comparables en importance ? Sont-ils significatifs ou négligeables ? etc.

Nous proposons de différencier les réponses à ces questions selon un critère particulier, celui de la symétrie exigée ou non des mises en relation des variations des variables. Ainsi nous distinguons les liens statistiques dont la définition inclut une exigence de réciprocité entre les variations des variables intéressées, et ceux dont la définition n'accorde pas le même statut aux variables considérées : les variations de la seconde (disons « impliquée » ou « seconde » ou « fils ») ne sont étudiées qu'au regard des variations de la première (disons le « père » ou l' « initiale »).

Ainsi nous comparons des traitements « classiques » qui permettent d'établir des relations symétriques entre variables tels ceux que le coefficient de corrélation linéaire r

permet (ou le chi²) aux traitements issus de l'ASI, qui permettent au contraire d'établir des dissymétries.

Ces problématiques et les décisions que nous venons d'évoquer sont inscrites dans le champ de la statistique. Notre projet est de montrer ici l'intérêt qu'elles peuvent présenter dans le champ des didactiques disciplinaires ou plus largement, dans celui des sciences de l'éducation. Pour ce faire, il est nécessaire de considérer l'interprétation ou les points de la modélisation qui posent problème d'un point de vue didactique.

Il a été dit plus haut que la problématique des liens entre variables reposait sur l'identification de modifications, de changements de valeurs d'une des variables sur un *groupe* de sujets. Par conséquent, les changements de valeurs à étudier sont plutôt à considérer du point de vue d'un groupe de sujets et non pas d'un individu isolé.

Ceci nous amène à interroger les groupes pertinents pour les études didactiques. Nous envisageons ici quelques-uns des groupes pertinents du point de vue didactique. Pour que cette pertinence soit reconnue, nous ne relevons que des groupes relativement stables afin que leurs frontières soient identifiables ainsi que leurs variations. Il nous semble que parmi ces groupes font consensus les groupes classes, les groupes de niveaux scolaires, les groupes de classes disciplinaires (nous pensons par exemple aux groupes d'élèves pour les enseignements de langues). Il nous semble également que sont moins consensuels, au regard toujours de leur pertinence pour les didactiques, les groupes identifiés par des choix pédagogiques revendiqués des enseignants ou encore les regroupements d'élèves selon leur « catégorie sociale »[8]. C'est pourquoi nous avons choisi dans ce cadre, de présenter une étude dans laquelle ces derniers regroupements seront l'objet d'une attention particulière.

5.1 La recherche « conscience disciplinaire »

Nous disposons d'un vaste corpus, construit à partir de la passation de trois questionnaires. Ces questionnaires ont été conçus et analysés dans le cadre d'une recherche en didactiques, dirigée par Cora Cohen-Azria, qui a donné depuis lieu à un ouvrage collectif. Cette recherche a pour but de questionner ce que nous avons appelé la « conscience disciplinaire » d'élèves de CM1 et CM2. Pour le dire très rapidement, nous appelons « conscience disciplinaire » les façons dont chaque élève identifie (ou ignore) les disciplines scolaires, les façons dont il se munit pour identifier ces différents espaces disciplinaires, les façons dont il se reconnait sujet de ces espaces et y agit. Ainsi par exemple, les façons de reconnaître dans quelle discipline on travaille – ce qui n'est pas toujours aussi évident que l'on peut le croire – ce qui y est valorisé et ce qui y est dénigré, ce qui y est possible, ce qui ne l'est pas, sont autant de dimensions de cette conscience disciplinaire : déclarer qu'en Français « ce que le prof attend c'est surtout du bla bla **bla** » tandis qu'en Mathématiques « c'est surtout un raisonnement correct », est une trace de la conscience disciplinaire du Français et des Mathématiques (en tant que disciplines scolaires) du lycéen interrogé. Durant cette recherche, nous avons tenté d'identifier certaines dimensions de ces consciences disciplinaires, dans quatre disciplines : le Français et les Mathématiques (disciplines scolaires reconnues comme « très importantes » et « très présentes »), les Sciences et l'Histoire-Géographie (qui le

[8] Guillemets, car ce sont plutôt celles de leur famille, avec tous les risques que cela comporte quant à l'interprétation du terme famille.

sont sans doute un peu moins). Nous avons choisi de les étudier auprès d'élèves de l'école primaire, pour plusieurs raisons : celle tout d'abord de l'organisation institutionnelle, en France, des disciplines scolaires, qui varie fortement entre l'école maternelle, l'école primaire, le collège et les divers types de lycées (nous mettons en note quelques différences). Les autres raisons tiennent à l'ancrage du travail didactique de notre équipe de recherche (Théodile CIREL) dans l'exploration des effets des modes de pédagogies alternatives (références) ainsi que dans l'intérêt que nous avons vu à explorer ces degrés de conscience disciplinaire à un moment du curriculum « charnière ».

C'est pourquoi ces questionnaires ont été soumis à quelques trois cents élèves de CM1 et CM2. Ces trois questionnaires les interrogent sur les façons dont ils dénomment, identifient, reconstruisent… les disciplines scolaires choisies: le Français, les Mathématiques, les Sciences et l'Histoire-Géographie (en note, nous expliquons que quelques questions plus globales sont importantes aussi). Ils sont pratiquement uniquement constitués de questions ouvertes.

Nous nous focalisons ici sur les variables issues de l'étude de quatre questions :

(1) Qu'est-ce que tu as appris d'important en Mathématiques ? (Questionnaire 1)
(2) Qu'est-ce que tu as appris d'important en Français ? (Q1)
(3) Qu'est-ce qui est important pour réussir en Mathématiques ? (Q3)
(4) Qu'est-ce qui est important pour réussir en Français ? (Q3)

Nous avons relevé les expressions/ locutions (terme à fixer) des réponses des élèves (présents à Q2 et à Q3). Et nous avons défini ainsi un ensemble de 51 variables binaires : elles valent 0 ou 1 selon que cette expression/locution a été employée par l'élève dans la 1e, 2e, 3e ou 4e question. Par exemple la variable FIOrtho correspond à la question (2) (F pour Français, I pour Important Appris) et à l'usage du terme « orthographe » dans la réponse de l'élève.

5.2 L'importance du flux implicatif et de la dissymétrie

Nous avons fait passer ces questionnaires dans des établissements implantés dans des quartiers « socialement défavorisés », et dans d'autres qui le sont au contraire dans des zones « socialement favorisées ». Et de fait, savoir s'il est possible, à partir des origines ou des milieux sociaux des élèves d'induire des conséquences quant à leurs représentations de l'école et des disciplines scolaires est une question très étudiée et particulièrement polémique. Les discours d'opinion abondent sur ce problème récurrent qu'ils tranchent dans un sens ou dans un autre. Du point de vue scientifique, en sociologie de l'éducation, dans les différentes didactiques (et peut-être plus particulièrement en didactique du français), les références sont nombreuses également (voir les travaux du groupe ESCOL, ceux de Lahire…). Formulons autrement notre question : est-ce que les expériences sociales possibles et effectives des *enfants* sont déterminantes pour leurs compréhensions, lectures de l'organisation de l'école où ils sont *élèves* ? Cohen-Azria, Lahanier-Reuter et Reuter, 2013).

Reprenons les termes que nous venons d'utiliser : « induire des origines sociales des conséquences sur les représentations ». Ceci recouvre quelques questions suivantes :

- est-ce que le type de quartier est considéré dans cette question comme un facteur/déclencheur de certaines dimensions des représentations? (corrélée ou « cause », « père » dans le schéma implicatif à des locutions/termes qui seraient distincts : par exemple, « être dans un milieu défavorisé » (est corrélé à) implique que/ a pour conséquence « dire que pour réussir en maths, il faut savoir compter » et « pour réussir en Français il faut bien écouter », tandis que « être dans un milieu favorisé » (est corrélé à) implique que/ a pour conséquence « dire que pour réussir en maths, il faut faire des exercices, s'entrainer » et « pour réussir en Français il faut bien écouter »). Le type d'établissement est alors « antécédent », « père » de certaines dimensions.
- est-ce qu'à un type d'établissement correspond un certain type de représentation ? (presque une fonction, donc des relations corrélées/implicatives comme ci-dessus, mais l'exigence de ne relever aucun terme/locution partagée par des élèves de milieux différents./pas de consensus possible, pas d'attitude commune entre ceux qui appartiennent à des types d'établissements différents).
- est-ce qu'un type d'établissement est particularisé/singularisé par plusieurs dimensions des représentations ? Ce qui est assez différent de ce qui précède. Dans ce cas, ce sont certaines dimensions qui seront propres à des élèves de ce type d'établissements : elles ne seront pas communes à tous les élèves de ces derniers, mais ne se relèveront que dans ceux-ci (tout cela doit être tempéré par « statistiquement »). Par exemple, le fait de reconnaître la lecture comme moyen de réussite en français est un discours que seuls tiennent des élèves des établissements d'un type particulier. Pour l'analyse que nous menons, ces dimensions apparaîtront alors comme des « pères », des antécédents des types d'établissement. La situation est l'inverse de la première.

Nous avons donc interrogé les discours des élèves qui ont été recueillis sous cet angle : les caractéristiques de ces discours, autour des contenus importants enseignés en Mathématiques et en Français et autour des contenus importants pour réussir dans ces deux disciplines, sont confrontées aux inscriptions des établissements scolaires qu'ils fréquentent, dans des quartiers « socialement favorisés » ou au contraire « socialement défavorisés ». A nos variables précédentes s'en ajoutent donc deux : « Favorisé » et «Très Défavorisé », qui sont encore des variables binaires. Les cônes implicatifs sont centrés tour à tour sur ces deux variables.

5.2.1 Les établissements situés dans des quartiers « très défavorisés »

Les facteurs « pères »

Commençons par le cône implicatif qui apparaît sous CHIC au seuil .60 centré sur la variable «Très Défavorisé ».

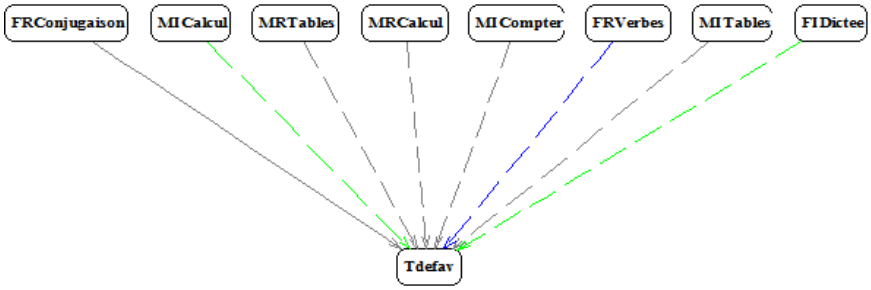


Figure 2 : Cône implicatif au seuil .60 centré sur la variable « Très défavorisé »

Nous constatons que le fait d'être inscrit dans un établissement (public) situé dans un quartier «Très défavorisé » n'a **aucune *conséquence* implicative**, mais apparaît comme une « conséquence » implicative.

Il semble donc que dans les établissements étudiés, situés dans des quartiers très défavorisés, *on puisse identifier des discours d'élèves qui sont propres à ces établissements.*

- Les contenus importants en Mathématiques sont compter/calculer/les tables de multiplication (MICompter, MICalcul, MITables);
- Il est important pour réussir en Mathématiques de [savoir] les tables de multiplication (MRTables) ;
- Les contenus importants en Français sont [faire] des dictées (FIDictées);
- Il est important pour réussir en Français de [savoir] les verbes et/ou les conjugaisons (FRVerbes, FRConjugaison).

Ces discours, spécifiques (au sens statistique du terme) des établissements étudiés situés dans des quartiers très défavorisés, sont cohérents. Ils mettent en évidence des constructions des deux disciplines essentiellement centrées sur des contenus :

- emblématiques de l'école primaire (compter, calculer, les tables de multiplication ou les dictées, les verbes et les conjugaisons) ;
- ancrés dans les représentations communes ;
- utiles ;
- gérables/contrôlables par les parents ou l'entourage ;

Pour le dire très brièvement, ces traits sont ceux d'une conscience disciplinaire qui serait celle du « certificat d'études », c'est-à-dire d'un enseignement non basé sur des sous-disciplines, destinés à des enfants dans lesquels on lit l'adulte, les futurs travailleurs et non pas de futurs collégiens par exemple.

Les corrélations partielles

Si nous considérons à présent les corrélations partielles entre ces différents facteurs (voir Annexe), seules les variables (MRCalcul, MRTables) et (FRConjugaison, FRVerbes) peuvent être considérées comme couples de variables dépendantes au seuil

de risque de 5%. Remarquons tout d'abord que ces couples de variables concernent des réponses distinctes aux deux questions :

- qu'est-ce qui est nécessaire selon toi pour réussir en Mathématiques ?
- qu'est-ce qui est nécessaire selon toi pour réussir en Français ?

Il n'y a pas trace de corrélation entre des réponses concernant des disciplines scolaires différentes, ce qui conforte notre hypothèse de conscience disciplinaire, c'est-à-dire du poids des disciplines scolaires dans les « attitudes » des élèves. Plus précisément, les deux conditions à la réussite en Mathématiques telles qu'elles sont principalement énoncées par des élèves des établissements très défavorisés (la maîtrise du calcul *vs* la connaissance des tables de [multiplication]) seraient des conditions exclusives[9] (r = -0,23). Cette tendance à l'exclusion peut être interprétée comme l'indice de deux modes d'appréhension de la réussite, de ce qui est attendu par l'enseignant, de ce qui est évalué en mathématiques par ces élèves. Y aurait-il une différence entre concevoir les Mathématiques comme l'art de calculer (au sens de « jongler avec les chiffres) et les résumer par un « slogan » (il faut savoir [ses] tables) ? Ou n'est-ce là qu'une caractéristique du discours des élèves, qui en quelque sorte « choisissent » l'un des termes (calculer ou les tables) de façon tendanciellement exclusive tant ils leur paraissent synonymes ou redondants ?

Il est plus simple nous semble-t-il d'interpréter la corrélation positive entre les deux facteurs de réussite en Français, [maîtriser] les conjugaisons et [connaître/savoir] les verbes, qui sont en effet parfois rapprochés dans le discours des élèves. Certains d'entre eux choisissent donc de considérer ces deux termes comme non redondants, non synonymes, mais plutôt sans doute comme complémentaires.

5.2.2 Les établissements situés dans des quartiers favorisés

Considérons à présent les établissements situés dans des quartiers considérés comme plutôt favorisés.

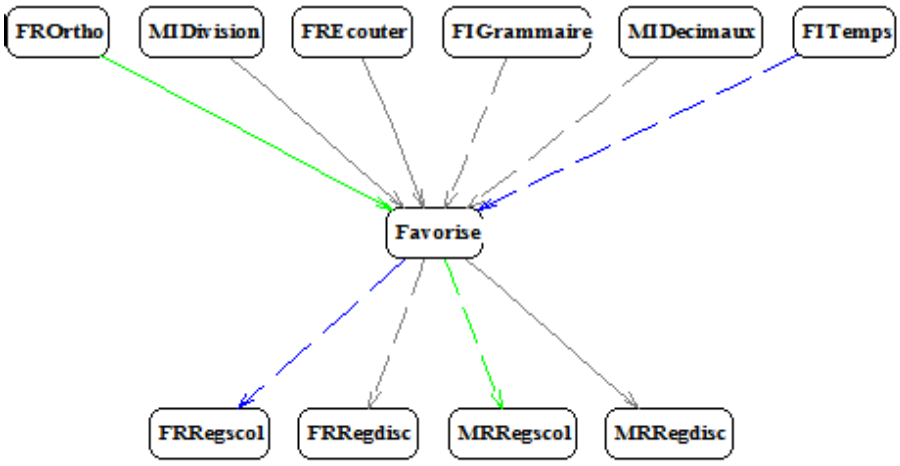


Figure 3 : Cône implicatif à .60 centré sur la variable « Favorisé »

[9] Ou des façons de dire exclusives plutôt.

Cette fois, les établissements « favorisés », centre d'un bassin d'attraction, sont à la fois des « pères » pour certaines dimensions et des « fils » pour d'autres (on retrouve la relation dialectique où la cause détermine l'effet et réciproquement, voir plus haut). Commençons par les discours qui sont ceux d'élèves de ces établissements, et d'eux (statistiquement) uniquement :

- Les contenus importants en Mathématiques sont la division et les nombres décimaux (MIDivision, MIDécimaux) ;
- Pour réussir en mathématiques… (il n'y en a pas de cités) ;
- Les contenus importants en Français, c'est la grammaire et les temps (FIGrammaire, FITemps) ;
- Pour réussir en Français, [il faut bien] écouter (FREcouter).

Par conséquent, les contenus qui sont désignés comme importants en Mathématiques sont des contenus qui sont *nouveaux* (au CM1 et au CM2), la division et les nombres décimaux. La perception de cette nouveauté présuppose son ancrage dans un apprentissage continu, ce qui particularise sans doute la conscience disciplinaire de ces élèves en ce qui concerne les Mathématiques. Ceux qui le sont en Français sont pour l'un, une sous-discipline (la grammaire) pour l'autre une désignation (scolaire?), les temps.

Les discours de ces élèves d'établissements (publics) « favorisés » privilégient parfois des contenus disciplinaires qui portent la marque de l'école: apprendre des temps (peu employés parfois hors de l'école, comme le passé antérieur), la grammaire, la division (qui est vraiment un savoir nouveau et complexe), les nombres décimaux (qui ne sont pas toujours les plus usités à l'extérieur de cette école, pensons par exemple à l'écriture 0,9764). Nous avançons, avec beaucoup de prudence, comme précédemment, l'idée que ces discours sont les traces de consciences disciplinaires qui seraient de l'ordre de l'apprentissage de contenus académiques, dont la maîtrise inscrit son possesseur dans une trajectoire/curriculum.

Cette fois, deux variables seulement sont corrélées positivement, au seuil de risque de 5% : Le fait de citer les décimaux comme contenus importants vus en Mathématiques et celui de citer les Temps comme contenus importants vus en Français : une vision globale des savoirs disciplinaires conforme à celle des programmes ?

5.2.3 Comparaison

Si nous comparons les « antécédents » mis au jour des deux types d'établissements (favorisés et très défavorisés), il nous semble qu'ils montrent des différences sensibles. Identifier des sous-disciplines et non pas des objets, identifier les savoirs nouveaux comme importants, ceux qui sont bien évalués (la maîtrise de la division) et non pas ceux qui sont intéressants à maîtriser (bien compter, savoir ses tables de multiplication) et que l'entourage ou l'opinion commune désigne comme tels seraient des positions (?) qui permettraient de parier à presque coup sûr sur l'appartenance à un établissement inscrit dans un quartier « favorisé ».

En ce qui concerne les « conséquences » de l'inscription dans un établissement « favorisé », elles présentent des aspects similaires, qui concernent les conditions de réussites :

- Que ce soit en Français ou en Mathématiques, les élèves de ces établissements s'entendent (statistiquement toujours) sur le fait que l'on réussit en obéissant à des règles scolaires : [bien] écouter/ ne pas faire de bêtises/faire ce que le maître/la maîtresse dit de faire etc.
- Que ce soit en Français ou en Mathématiques, les élèves de ces établissements s'entendent (statistiquement toujours) sur le fait que l'on réussit en obéissant à des règles cette fois disciplinaires : en Français en consultant son Bled, en Mathématiques en tenant bien droit sa règle pour tracer, etc.

Sur le corpus dont nous disposons, cette analyse confirme des résultats déjà établis : il y a des modes d'appréhension variables de l'école, ici de la structuration des enseignements scolaires en enseignements disciplinaires. Et ces variations, de ce qui est pour nous des traces des consciences disciplinaires, sont réglées (soumises ?)/ expliquées par les variations des milieux sociaux dominants des quartiers dans lesquels les établissements scolaires sont implantés. C'est le respect formel des règles qui semble dériver de l'appartenance au milieu favorisé : une institutionnalisation acceptée ?

Mais il nous est aussi possible, par le recours à ce type d'analyse, de détruire (ou de reconsidérer ?) la symétrie des relations (statistiques) entre dimensions de la conscience disciplinaire et type d'environnement social dans lequel les élèves évoluent. Ces relations (à une variation conséquente du milieu correspond une variation conséquente des dimensions reconstruites) comme nous venons de le voir, ne sont pas, loin de là, des relations simples de cause à effet. En montrant que certaines de ces dimensions sont des caractéristiques « partagées » tandis que d'autres semblent plutôt être le fait de « groupes d'élèves », l'ASI démontre aussi son intérêt.

Cependant, notre analyse n'est pas terminée. En effet, notre projet global est l'évaluation (dans un sens très large) les effets des modes de travail pédagogiques dans des quartiers populaires (voir Reuter etc.).

5.2.4 Les pédagogies dont se réclament les enseignants

Par conséquent, les variations les plus intéressantes pour nous sont celles des modes de travail pédagogiques : dans le corpus dont nous disposons, deux établissements (dans des quartiers « très défavorisés ») revendiquent plus ou moins l'inscription dans une démarche pédagogique particulière. Dans le troisième et dernier des établissements inscrits dans ce type de quartier, les enseignants en revanche, ne revendiquent pas de démarche pédagogique spécifique.

Voici un tableau descriptif des inscriptions pédagogiques déclarées par les enseignants et les codages :

Mode de travail pédagogique	**Établissements**[10]	**Codages des classes correspondantes**
Mouvement Freinet	Hélène Boucher (HB)	HB1
		HB2

[10] Les noms ont été modifiés, sauf celui d'Hélène Boucher.

Fonctionnement collectif : échanges de pratiques, de groupes de classe etc.	Pascal (Pa)	Pa1
		Pa2
Pas d'inscription particulière	Veermer (Ve)	Ve

Tableau 2

Nous nous posons donc la question de savoir si aux différentes pratiques pédagogiques des maîtres correspondent des différences entre les discours des élèves et des [reconstructions de leur] conscience disciplinaire.

L'analyse a été menée sur le même corpus, en centrant cette fois les cônes implicatifs sur les 5 classes de ces trois établissements.

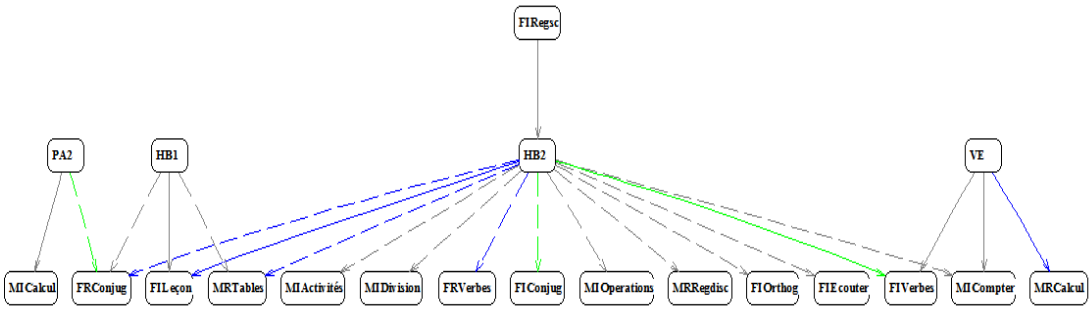


Figure 4

La particularité de la classe HB2 « saute aux yeux ». C'est tout d'abord celle où les discours sont les plus variés. C'est aussi celle où les discours des élèves sont originaux : ainsi la dimension « les contenus importants en Mathématiques sont les activités spécifiques telles les recherches mathématiques (MIActivités). Nous notons enfin que quelques-unes des « conséquences » c'est-à-dire quelques-unes des dimensions manifestées par l'appartenance à cette classe sont celles qui particularisaient des discours des élèves des établissements situés dans un autre type de quartier : MRReglesdisciplinaires, MIDivision, FIOrthographe, FIécouter.

Nous ne creuserons pas davantage ici cette singularité, pour laisser la place à une question qui illustre la relation dialectique cause-effet. Cette classe HB2 nous est familière : elle est celle d'un enseignant qui est un des piliers de ce mouvement de l'ICEM, qui a été un des premiers enseignants de cette équipe[11]. Faut-il en conclure que ce dernier peut être qualifié d'expert ? (et, parallèlement, qualifier le second enseignant de l'établissement HB, l'enseignant de la classe HB2 de « novice » ?). Nous hésitons à attribuer ces étiquettes à des sujets : ne faut-il pas plutôt parler dans la classe HB2 de pratiques pédagogiques réellement alternatives ?

6 Conclusion

[11] Nous avons suivi cette expérience (une école entière fonctionnant en pédagogie Freinet depuis 2001. L'enseignant responsable de la classe HB1 est un jeune enseignant, qui a rejoint cette école depuis un an, mais qui s'inscrit lui aussi dans ce mouvement pédagogique.

A la faveur de l'analyse non symétrique d'un corpus de données et à travers la représentation en graphe implicatif de l'ensemble des règles d'association obtenues, nous avons isolé un sous-graphe constitué d'une sorte de cône à deux nappes. Son intérêt réside dans la mise en évidence du rôle de bassin d'attraction que remplit le sommet du cône. Dans le cas où la nappe supérieure supposée causale est constituée de sommets indépendants, on dispose d'une confluence causale des pères du sommet comme effet résumé de plusieurs causes et, en un même temps dialectique, explicatif de celles-ci. De la même façon, l'examen des « fils » de ce sommet permet de donner un sens de conséquences éclatées de cet attracteur que joue le sommet. Plusieurs situations concrètes, dont l'un plus fouillée portant sur la notion de « conscience disciplinaire », ont permis d'illustrer ces démarches. Il serait intéressant de se priver de l'indépendance des « causes » pour, au contraire, examiner en quoi le sommet pourrait relever de l'agglutination de ces causes qui ne seraient alors que des modalités faisant en quelque sorte « cause commune ». Cette autre problématique rejoindrait celle de cet article dans notre ambitieuse entreprise de l'élucidation des relations causales entre phénomènes car comme l'écrit Francis Bacon : « *Vraiment connaître, c'est connaître par les causes* » et par ailleurs : « *Savoir pour prévoir, prévoir pour pouvoir* ».

Références

[1] Bailleul, M. & Godard, S. (2009), Derrière les réseaux de variables, il y a des individus… à écouter!, In Gras R. (dir.), Régnier J-C., Marinica C., Guillet F., *L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités,* Cépaduès Ed., Toulouse, 437-453.

[2] Bodin, A. & Gras, R. (1999), Analyse du préquestionnaire enseignants avant EVAPM-Terminales, *Bulletin de l'Association des Professeurs de Mathématiques,* n° 425, 772-786.

[3] Briand, H., Sebag, M., Gras, R.et Guillet F. (eds), (2004), *Mesures de Qualité pour la Fouille de Données,* RNTI-E-1, Cépaduès, Toulouse.

[4] Cohen-Azria, C., Lahanier-Reuter, D., Reuter, Y. (2013), *La conscience disciplinaire,* P.U.R., Rennes.

[5] Gras R. (dir.) Régnier J.-C., Marinica C., Guillet F (eds.), (2013), L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités, Cépaduès Editions, Toulouse.

[6] Gras, R. (dir.), Almouloud, S. Ag, Bailleul, M., Larher, A., Polo, M., Ratsimba-Rajohn, H., Totohasina, A. (coll.), (1996), *L'implication statistique. Nouvelle méthode exploratoire de donnée,* La Pensée Sauvage, Grenoble.

[7] Gras, R., Kuntz, P. et Briand, H. (2001), Les fondements de l'analyse statistique implicative et quelques prolongements pour la fouille de données, *Mathématiques et Sciences Humaines,* n° 154-155, 9-29.

[8] Gras, R., Suzuki, E., Guillet, F., Spagnolo, F. (eds), (2008), *Statistical Implicative Analysis, Theory and Applications,* Springer.

[9] Guillet, F. & Hamilton, H. (eds), (2007), *Quality Measures in Data Mining,* Springer.

[10] Lahire, B. (1993), Culture écrite et inégalités scolaires, P.U.L., Lyon.

[11] Lévy-Leblond, J.-M. (1996), *Aux contraires*, Gallimard, Paris.

[12] Orus, P., Zemora, L., Gregori, P., (2009), *Universitat Teoria y Aplicaciones del Analisis Estadistico Implicativo*, Eds : Jaume-1, Castellon (Espagne).

[13] Pasquier, D. & Gras, R. (2012), De l'intérêt de l'Analyse Statistique Implicative (A.S.I.) pour la recherche exploratoire en psychologie, *Psychologie Française*, Elsevier-Masson, 161-173.

[14] Régnier, J-C., Bailleul, M., Gras, R. (eds.), (2012), *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire*, Université de Caen, Caen.

[15] Sève, L. (2005), *Émergence, complexité et dialectique*, Odile Jacob, Paris.

Annexes

Coefficients de corrélation (Figure 2)

	MR Tables	FR Conjugaison	FR Verbes	MI Tables	MI Calcul	MI Compter	FI Dictées
MR Calcul	-0,23	0,07	0,08	-0,02	0,20	0,04	0,05
MRTables		0,14	0,19	0,06	0,01	0,02	0,11
FRConjugaison			0,28	0,05	0,09	0,02	0,08
FRVerbes				0,21	0,04	0,03	0,14
MITables					-0,06	-0,03	0,07
MICalcul						0,06	-0,05
MICompter							-0,04

Corrélations partielles

r(MRC, MRT,FRC) = **-0,24** r(MRC,MRT, FRV) = **-0,25**

r(MRC, FRC, FRV)= 0,05 r(MRT,FRC, FRV) =0,09

Pour la Figure 3

	FREcouter	FROrtho	MIDivision	MIDécimaux	FITemps	FIGrammaire
FREcouter		-0,06	-0,02	0,02	0,12	0,11
FROrtho			0,05	0,2	0,09	0,18
MIDivision				0,09	0,22	0
MIDécimaux					0,32	0
FITemps						-0,08
FIGrammaire						

Cette fois, deux variables seulement sont corrélées positivement, au seuil de risque de 5% : Le fait de citer les décimaux comme contenus importants vus en Mathématiques et celui de citer les Temps comme contenus importants vus en Français : une vision globale des savoirs disciplinaires conforme à celle des programmes ?

HIERARCHIE DE REGLES EN A.S.I. ET CONCEPTUALISATION

Régis GRAS[1], Nadja ACIOLY-REGNIER[2]

HIERARCHY OF RULES IN S.I.A. AND CONCEPTUALIZATION

RÉSUMÉ

Nous présentons dans cet article une méthode alternative et complémentaire au graphe implicatif pour représenter les relations implicatives nouées au sein de variables qualifiant ou quantifiant des sujets ou des objets. Mais ici les règles extraites des données sont de degré supérieur et apparaissent comme métarègles ou règles de règles. Nous axiomatisons ces règles généralisées et les représentons selon une hiérarchie ascendante orientée. En psychologie différentielle et en didactique de sciences (champs conceptuels), la complexité cognitive nous semble apparaître comme métaphore ou avatar privilégié de ce type de hiérarchie. Nous présentons deux exemples illustratifs en montrant l'intérêt prédictif de cette représentation.

***Mots-clés** : Hiérarchie, classes orientées, axiome, distance ultramétrique**,** psychologie différentielle, champ conceptuel.*

ABSTRACT

In this article we present an alternative and complementary method to the implicative graph to represent the implicative relationship to present the knotted within the variables qualifying or quantifying either subjects or objects. But here, the data extracted rules are of a superior level and appear as metarules or as rules of rules. We axiomatise those generalised rules and we represent them according to an oriented ascendant hierarchy. In differential psychology and in science didactic (conceptual fields), cognitive complexity seems to appear as a privileged metaphor or avatar for this type of hierarchy. We present two illustrative examples by showing the predictive interest of this representation.

***Keywords :** hierarchy, oriented classes, axiom, ultrametric distance, differential psychology, conceptual field.*

1 Introduction

A notre connaissance, d'une part, les développements existant en matière de mesures de qualité de règles d'association s'arrêtent généralement à la proposition d'un indice d'implication partielle pour des données binaires et, au mieux, à la représentation arborescente des règles (par exemple : Agrawal, R. et al, 1993, Guillet F. et Hamilton H. 2007, Hiep J. et al 2000, Lenca P. et al, 2007 et tous les travaux portant sur les réseaux bayesiens et les treillis de Galois, voir Cadot M. 2009). D'autre part, cette notion n'est pas étendue à l'extraction de règles de règles où les prémisses et les conclusions peuvent

[1] Laboratoire d'Informatique de Nantes-Atlantique (LINA) Site Polytech Nantes - La Chantrerie, rue C. Pauc, BP 44306, Nantes cedex 3, e-mail : regisgra@club-internet.fr,

[2] ESPE-École Supérieure du Professorat /Université Claude Bernard – Lyon 1. E-mail: nadja.acioly-regnier@univ-lyon1.fr

être elles-mêmes des règles. A l'instar de certaines méthodes d'analyse de données, l'intérêt de passer, non linéairement, du niveau relationnel à celui de hiérarchie nous est apparu évident, comme une réponse à la question brûlante, dans la construction des concepts, de non-linéarité[3]. Nous pouvions nous appuyer sur notre expérience en didactique des mathématiques et en psychologie cognitive. Nous avons alors prolongé notre recherche de règles dans un corpus de données à son extension théorique, celle de règles de règles ou méta-règles (ou règles généralisées). *Ainsi, notre recherche de relations causales sera* ***non-réductionniste*** *puisqu'elle dépassera le simple dénombrement et conduira au nouveau choix d'un indice, ceci en élargissant le champ opérationnel de l'ASI à des règles de niveau supérieur, à leur représentation, à l'identification et à la mesure du rapport dual sujets-variables.*

Nous proposons *donc* ici ces prolongements dans le cadre de l'A.S.I., en formalisant la notion de hiérarchie orientée. Celle-ci vise la prise en charge de la représentation graphique de la structure de l'ensemble des associations des variables selon **des règles de règles**, susceptibles pour certaines de se situer en toute hypothèse à un niveau conceptuel supérieur. Une première version formelle a été présentée dans (Gras et, 1996). L'approche en était intuitive et pragmatique. Depuis, (Gras et Kuntz, 2005) a permis de réviser la notion de hiérarchie orientée par une présentation algébrique plus formelle mais en accord avec les propriétés rigoureuses de la notion de hiérarchie orientée.

La forme graphique obtenue peut nous faire penser, toutes réserves faites et par analogie de structure, dans son architecture logique, au modèle théorique épistémologique de J. Piaget. Ce modèle, aujourd'hui relativisé pour sa rigidité, sinon contesté, conçoit le développement cognitif de l'enfant selon une organisation élaborée passant par une succession de stades emboîtés successifs. Le passage d'un stade à un autre selon une conception établie sur un mode implicatif strict n'est pas sans poser problème (Pellois C., 2002, 2010 et 2013) car il se fait, notamment, mais pas seulement, par le bais de ce que Piaget appelle l'abstraction réfléchissante « Elle est réfléchissante aux deux sens suivants : elle transpose sur un plan supérieur de conceptualisation ce qu´elle emprunte à un palier précédent » écrit Sylvie Lucas dans Le Tome 52 du Bulletin de Psychologie, juillet-août 1999. Ecartant provisoirement toute référence à cette théorie, l'objectif central de ce texte réside dans l'étude de l'adéquation métaphorique de la représentation de règles de règles avec une démarche conceptualisante.

Rappelons qu'une représentation structurée des relations quasi-implicatives dans l'ensemble des variables instanciées a été obtenue, dans le cadre de l'A.S.I., par un graphe implicatif sans cycle, pondéré par les intensités d'implication, fermé transitivement à un seuil donné. Cette première représentation serait-elle comparable aux mécanismes de conceptualisation qui semblent se mettre en œuvre formellement pour aboutir, in fine à l'organisation cognitive du sujet épistémique, telle que la présente

[3] Relisons G. Bachelard : « Pour un esprit scientifique, toute connaissance est une réponse à une question. S'il n'y a pas eu de question, il ne peut y avoir connaissance scientifique. Rien de va de soi. Rien n'est donné. Tout est construit ». (Extrait de *La Formation de l'esprit scientifique* 1938).

Piaget (1970) ? Répondre à cette question nécessiterait très certainement bien des discussions et quelques approfondissements, mais dans des situations d'apprentissage ou d'évaluation, elle nous a permis d'interpréter des chemins de ce graphe, constitués de suites d'arcs qui lient certaines variables, en termes de genèses relevant d'une forme particulière de conceptions différentielles. Cette théorie piagétienne du développement cognitif du sujet, dans sa forme originelle exprimée en termes de stades ordonnés indépendants des contenus est cependant, comme nous l'avons dit, à relativiser dans sa rigueur formelle. Elle fait actuellement la place, non seulement à la psychologie différentielle sous différentes conceptions (Cf. par exemple, Pellois, C., 2008, en particulier p. 91.), mais aussi à la théorie des champs conceptuels[4] où le développement cognitif serait fortement lié au contenu conceptuel des domaines dans lesquels il s'opère plus qu'à des structures logico-mathématiques de la pensée. On en fera l'observation dans l'exemple 2 du § 4.2.

La méthode d'organisation hiérarchique des données, non symétrique que nous présentons maintenant, va doubler les informations fournies par les règles d'association implicative en organisant leur ensemble selon une structure ordonnée en méta-règles, en méta-méta-règles, etc.. Nous verrons, par la suite, quelle proximité de structure (quel homomorphisme ?) existe entre ce type de représentation et l'organisation de règles nécessaires à la conceptualisation. Les deux premiers paragraphes 2 et 3 présenteront une modélisation d'une hiérarchie orientée de règles de règles. Le paragraphe 4 étudiera en quoi la hiérarchie ainsi modélisée pourrait être une métaphore d'un processus de conceptualisation.

2 Hiérarchie de classes de variables[5]

Afin de soutenir l'intuition, nous baserons le modèle de la **hiérarchie orientée cohésitive** sur la correspondance linguistique suivante :

1. *les variables* (ou attributs) de l'ensemble V (card V=p) constitueront l'ensemble les *lettres de l'alphabet V*,
2. *les classes de k variables*, éléments de V^k, par exemple (a_1, a_2,…, a_k), constitueront les *syllabes du vocabulaire*,
3. *les classes maximales*, i.e. telles qu'aucune variable ne la complète, constitueront *les mots* du vocabulaire,
4. *l'organisation hiérarchique* de l'ensemble des classes constituera une *phrase*, structurée par des propositions incises.

D'autres métaphores peuvent illustrer le modèle que nous allons construire comme, par exemple, l'ensemble des séquences constituant le génome ou encore une théorie

[4] Vergnaud (Vergnaud 1981 et 1990) définit la notion de champ conceptuel « comme un espace de problèmes ou de situations-problèmes dont le traitement implique des concepts et des procédures de plusieurs types en étroite connexion. »

[5] Cette section §2 est une composante de l'article Gras et Kuntz (2005), reprise d'ailleurs dans Gras et al. (2009 et 2013). On la retrouve présentée dans Régnier et Acioly-Régnier (2007). Dans un souci d'homogénéisation, nous tenions à le rappeler ici.

mathématique organisée en théorèmes et corollaires. Mais nous verrons que ces métaphores ne satisfont pas totalement la structure du modèle.

On va également constater que ce modèle hiérarchique, où l'ordre intervient, ne s'apparente pas au modèle classique d'une hiérarchie ascendante, par exemple celle basée sur un indice de similarité entre attributs, car les classes d'une telle hiérarchie sont des sous-ensembles de variables et non pas des k-uplets.

2.1 Hiérarchie orientée. Définitions. Propriétés

Ce paragraphe théorique rend compte de la formalisation de la construction d'une hiérarchie dans laquelle l'ordre ascendant de classification se doit de respecter le fondement asymétrique de l'implication. C'est donc sur ce projet, cette base épistémologique que nous construirons ce nouveau modèle hiérarchique.

Définition 1: On appelle hiérarchie orientée **H** sur l'ensemble des variables V, une suite d'arrangements (au sens de la combinatoire, donc suite de k-uplets) des éléments de V, vérifiant les axiomes 1. 2 et 3 énoncés ci-dessous. Ces arrangements sont appelés classes de H.

Par exemple, {(j), (f,g), (b,c), (e,f,g), (b,c,d), (h,i), (a,b,c,d), (e,f,g,h,i)} est une hiérarchie orientée sur V={a,b,c,d,e,f,g,h,i} et (a,b,c,d) est une classe de cette hiérarchie.

Définition 2: On appelle classe **C** de degré k de la hiérarchie **H** un arrangement (ou k-uplet) de k éléments de V appartenant à **H**. On notera $\prec$ la relation d'ordre induite sur **C** par le tirage d'un arrangement.

Par exemple, $(a_1, a_2,\ldots, a_k)$, pour $k \leq p$, est une classe de degré k et $a_1 \prec a_2 \prec \ldots \prec a_k$. Mais également, par convention, (a) est une classe de degré 1. Elle est dite élémentaire

Définition 3: On appelle troncation de **C**, tout sous-arrangement des éléments de **C** respectant la structure d'ordre $\prec$ et la consécutivité.

Par exemple, si **C** = $(a_1, a_2,\ldots, a_k)$, la classe **C'** = $(a_i, a_{i+1},\ldots, a_j)$ où $1 \leq i$ et $j \leq k$, est une troncation de **C**.

Définition 4: On note **C'** $\hat{\subset}$ **C** si et seulement si **C'** est une troncation de **C**. Une classe est dite maximale s'il n'existe pas de classe qui la contienne dans **H**. Elle est dite minimale si elle ne contient aucune classe de la hiérarchie **H**. En particulier, une classe élémentaire est donc minimale (mais elle peut être aussi maximale).

On peut comparer, sans la confondre cependant, cette relation à l'inclusion ensembliste. Dans l'exemple initial, les classes (a,b,c,d), (e,f,g,h,i) et (j) sont maximales. Cette dernière est aussi minimale.

Définition 5: La trace de **C** sur **C'** est constituée d'éléments communs à **C** et **C'** et elle respecte la structure d'ordre $\prec$ et la consécutivité. La trace est une opération commutative notée $\hat{\cap}$.

Ainsi on peut comparer, sans la confondre, cette opération à l'intersection ensembliste.

Par exemple, (d,f,g,a,e) $\hat{\cap}$ (f,g,a,e,b,h)= (f,g,a,e)

Définition 6: Si les deux classes quelconques **C'** et **C''** ont une trace vide (**C'** $\hat{\cap}$ **C''**=$\varnothing$), la concaténation de **C'** et **C''** notée **C'** $\hat{\cup}$ **C''** est la classe **C** dont les éléments appartiennent à **C'** et **C''** et à elles exclusivement. Elle respecte les ordres au sein de **C'** et **C''**et le plus grand élément de **C'** précède le plus petit de **C''**. On dira que **C'** et **C''** sont des classes génératrices de **C'** $\hat{\cup}$ **C''.**

Cette opération, comparable à la concaténation ordinaire, ainsi qu'à la réunion ensembliste sans se confondre avec elle, est non commutative et respecte donc un ordre.

Par exemple, si **C'** = (d,f,g,a) et **C''** = (b,u,r,p,y), **C'** $\hat{\cup}$ **C''** = (d,f,g,a,b,u,r,p,y), alors que **C''** $\hat{\cup}$ **C'** = (b,u,r,p,y,d,f,g,a)

2.2 Axiomes d'une hiérarchie orientée

Axiome 1 : $\forall$C et $\forall$C' classes de H , C $\hat{\cap}$ C''$\in$ {$\varnothing$, C, C'}

Axiome 2 : $\forall$C$\in$H, si C n'est pas élémentaire ou minimale, elle est la concaténation de classes de H

Axiome 3 : Il existe une permutation des éléments de V qui coïncide avec la concaténation de toutes les classes maximales de H

2.3 Algorithme de construction de l'ensemble des classes

Nous définissons ci-dessous un critère algébrique en vue de nous permettre de construire de façon ascendante, la hiérarchie organisatrice de l'ensemble V des variables et qui respecte les trois axiomes d'une hiérarchie orientée.

2.3.1 Critères algébriques

Définition 7: La cohésion d'une classe de degré 2, correspondant au couple (a,b) est définie, à partir de l'entropie au sens de Shannon, par la formule, $coh(a,b) = \left(1 - [-p(\log_2(p)) - (1-p)(\log_2(1-p))]^2\right)^{\frac{1}{2}}$ où p=φ(a,b)≥0,50 et $coh(a,b) = 0$ si p=φ(a,b)<0,50.

Cette notion de cohésion nous conduit, dans le cadre de l'ASI, de parler aussi bien de **hiérarchie orientée** que de **hiérarchie cohésitive.**

Définition 8: La cohésion d'une classe **C** = (a_1, a_2,…, a_r) de degré r, est définie par la formule : $coh(C) = \left[\prod_{i=1,\dots,r-1}^{j=2,\dots,r;\, j>i} coh(a_i, a_j)\right]^{\frac{2}{r(r-1)}}$

Définition 9 : La cohésion d'une classe **C** = (a) de degré 1, est définie par $coh(a) = 1$

2.3.2 Algorithme de construction de la hiérarchie

Cet algorithme est implémenté dans le logiciel CHIC (logiciel de traitement non symétrique de données, Couturier R. et Ag Almouloud S., 2013).

Niveau 0

Les classes sont élémentaires et toutes les cohésions sont égales à 1

Niveau 1

On compare toutes les cohésions des arrangements 2 à 2 de V.

On conserve celle, notée C_1, qui correspond au maximum, soit par ex. C_1=(a,b).

Définition 1: On appelle nœud 1, la règle $a \Rightarrow b$

Niveau 2

On compare toutes les cohésions des classes à 2 éléments, sauf C1, à celles des classes à 3 éléments du type (x, a, b) et (a, b, x).

On conserve celle, notée C_2, correspondant au maximum obtenu.

Le nœud 2 sera :

1. soit une classe à 2 éléments, et dans ce cas le nœud sera du type : c=>d
2. soit une classe à 3 éléments, et dans ce cas le nœud sera noté (a=>b)=>c ou c=>(a=>b).

Ces dernières règles sont dites composées ou généralisées. Pour restituer l'ordre dans lequel est constituée la classe, on notera, par exemple ici, ((a,b),c) ou, dans l'autre cas, (c,(a,b)).

Niveau k

On compare toutes les cohésions des concaténations de 2 des classes déjà formées aux niveaux inférieurs, du type C_i et C_j avec i<k et j<k. On conserve celle $C_k=C_i \cup C_j$ qui satisfait le maximum et est la concaténation de C_i et C_j.

Le nœud k correspondant sera noté par extension C_i=>C_j. Mais on peut expliciter des nœuds correspondant à la formation des troncations respectives et génératrices de C_i et C_j aux niveaux inférieurs.

Par exemple, la règle composée $((f \Rightarrow (e \Rightarrow u)) \Rightarrow ((a \Rightarrow b) \Rightarrow (c \Rightarrow d))$ est l'explicitation d'un nœud particulier. Afin de faire apparaître les classes formées à des niveaux successifs, on notera aussi la règle sous la forme : ((f(eu))((ab)(cd)))

L'algorithme s'arrête, au plus tard, au niveau p-1, lorsque toute concaténation conduirait à une classe de cohésion nulle ou à une permutation de l'ensemble V des variables Les classes formées au niveau ultime et qui n'admettent pas de classes qui les contiennent sont donc maximales. Certaines classes maximales peuvent aussi être élémentaires. La hiérarchie est composée de l'ensemble des classes maximales et de toutes leurs parties.

2.4 Conformité de la construction aux axiomes d'une hiérarchie orientée

La hiérarchie ainsi construite vérifie les trois axiomes d'une hiérarchie orientée. En effet :

Axiome 1 :

Deux classes de **H**, **C'** et **C''** étant données,

1. ou bien elles sont associées dans une même concaténation et la constituent entièrement, alors **C'** $\hat{\cap}$ **C''**=$\varnothing$
2. ou bien l'une est la concaténation de l'autre et d'une troisième et alors **C'** $\hat{\subset}$ **C''** ou **C''** $\hat{\subset}$ **C'**
3. ou bien elles ne sont pas associées dans une concaténation et alors elles sont des arrangements sans élément commun, donc **C'** $\hat{\cap}$ **C''**=$\varnothing$

Axiome 2 :

Pour toute classe C de la hiérarchie :

1. ou bien elle est constituée d'un élément et c'est une classe élémentaire
2. ou bien elle est constituée de plus d'un élément et elle est alors la concaténation de deux ou plusieurs classes par construction.

Axiome 3 :

On range toutes les classes maximales par ordre croissant de la cohésion ; les classes élémentaires seront les éléments maximaux de cet ordre. Toutes les classes sont 2 à 2 disjointes et tous les éléments de V appartiennent à l'une et l'une seulement des classes. La concaténation de leur ensemble constitue alors une permutation particulière de tous les variables de V.

Notons qu'à une permutation de V peuvent correspondre plusieurs hiérarchies.

2.5 Exemples

Exemple 1: Si l'on range les classes maximales de la hiérarchie donnée au début du texte par ordre croissant de la cohésion, on obtient par exemple :

coh(e,f,g,h,i)$\leq$coh(a,b,c,d)$\leq$coh(j) et (e,f,g,h,i,a,b,c,d,j) est une permutation de A.

Mais à cette permutation, peut aussi correspondre la hiérarchie :

{(g,h), (b,c), (e,f), (a,b,c), (g,h,i), (d,j), (a,b,c,d,j)} dont les classes maximales sont (e,f), (g,h,i) et (a,b,c,d,j).

Exemple 2 : Reprenant encore l'exemple initial, l'autre hiérarchie {(j), (f,g), (b,c), (e,f,g), (b,c,d), (h,i), (a,b,c,d), (h,i,e,f,g)}ne coïncide pas avec la première. La permutation correspondante de A est (h,i,e,f,g,a,b,c,d,j).

Exemple 3

La figure ci-dessous montre la hiérarchie obtenue, artificiellement, à partir de 7 variables. Des interprétations de telles règles généralisées sont quelquefois complexes, comme par exemple, la règle $(x \Rightarrow (y \Rightarrow z)) \Rightarrow (t \Rightarrow v)$. Mais quelques règles sont réductibles à des assemblages plus aisément interprétables. Par exemple, la règle $x \Rightarrow (y \Rightarrow z)$ se ramènerait, dans le cas formel, à $x \wedge y \Rightarrow z$. Une cohésion de qualité (proche de 1) autoriserait cette extension sémantique. La règle $(d \Rightarrow b) \Rightarrow (a \Rightarrow f)$ ou ((db)(af)), illustrée par la figure (Fig. 1), peut s'interpréter par la phrase: le « théorème » $d \Rightarrow b$ a généralement pour conséquence le « théorème » $a \Rightarrow f$. Cette figure montre aussi que la variable *e* n'a ni prémisse ni conclusion.

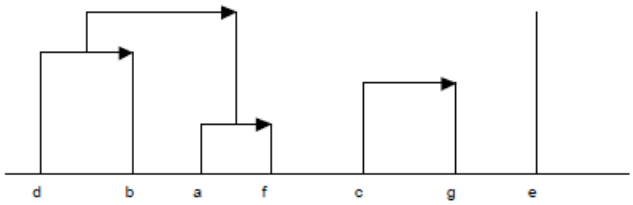


Figure 1 Un exemple de hiérarchie orientée

Exemple 4

Par exemple, de théorèmes comportementaux, on peut dégager **un trait** ou une **conception** (ex : conceptions du hasard par (Lahanier-Reuter, 1999)). Un indice statistique permet en outre de repérer les niveaux hiérarchiques correspondant à une **significativité des classes** formées à ces niveaux (en rouge sur la Fig.2). La hiérarchie de la Fig. 2 est produite, à l'aide de CHIC, à partir d'un fichier constitué par les réponses à un questionnaire figurant en annexe et présenté à des enseignants de mathématiques dans des classes terminales de lycée (18-19 ans). Ces enseignants devaient choisir et pondérer les objectifs qu'ils assignaient aux cours de mathématiques de leur programme. Nous reviendrons plus loin sur cet exemple.

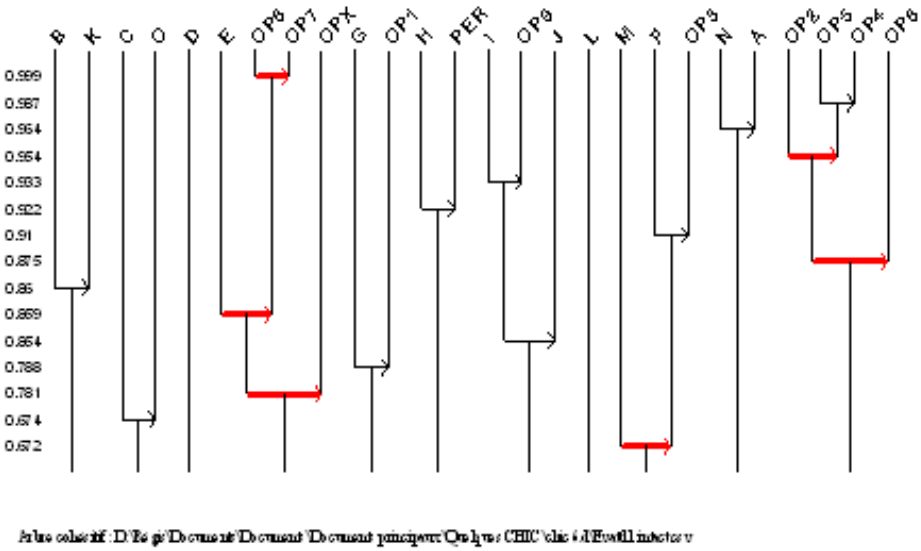


Figure 2 Hiérarchie des objectifs et attitudes propres à l'enseignement des mathématiques données par le logiciel CHIC

Il n'y a pas de raison fondamentale d'utiliser le type de représentation (« vers le haut ») de la fig.1 ou celui de la fig. 2 (« vers le bas »). La première est traditionnellement utilisée dans les hiérarchies de la théorie des graphes. La seconde, obtenue par pliage axial, organise la hiérarchie selon la cohésion croissante. Elle permet un accès privilégié aux noms des variables.

3 La hiérarchie cohésitive basée sur une distance ultramétrique

On part de l'algorithme de construction de H déjà défini et par lequel la cohésion de la classe en voie de formation, à un niveau donné, est inférieure à la valeur de la

cohésion au niveau immédiatement inférieur et est supérieure à celle du niveau immédiatement supérieur. C'est donc une fonction décroissante des niveaux et, a fortiori, avec l'inclusion des parties de H. C'est la propriété de la cohésion qui va permettre de définir la distance ultramétrique d, pour laquelle tous les triangles sont isocèles. Cette propriété d'ultramétricité de la cohésion va justifier, a posteriori, le bien-fondé de l'expression « hiérarchie » employée pour la construction.

Il suffit de choisir pour un couple quelconque de variables (x, y) :

$$d(x, y) = 1 - coh(C_{(x,y)})$$

où $coh(C_{(x,y)})$ est la cohésion de la plus petite classe $C_{(x,y)}$ contenant x et y. Rappelons les propriétés suivantes de la hiérarchie :

1. quelles que soient les classes C et C' de H, ou bien $C \hat{\subset} C'$ ou bien $C \hat{\supset} C'$ ou bien $C \hat{\cap} C' = \varnothing$
2. si $C \hat{\subset} C'$, alors $coh(C) \geq coh(C')$ par construction.

Vérifions alors que les trois axiomes d'ultramétricité sont bien valides sur l'ensemble V des variables

Axiome 1 :

Pour tout $x \in V$, d(x,x) = 0 par construction de d car $coh(x,x) = 1$

Axiome 2 :

Pour tout couple (x,y)∈ VxV, d(x,y)=d(y,x) par construction

Axiome 3 :

(x,y,z)∈ VxVxV, d(x,y)≤sup[d(x,z), d(y,z)] (1)

La définition adoptée respecte aussi l'axiome 3. En effet, soit $C_{x,y}$, $C_{x,z}$, $C_{y,z}$ les plus petites classes contenant respectivement x et y, x et z, y et z.

Alors $z \in C_{(x,z)} \hat{\cap} C_{(y,z)} \neq \varnothing$ d'où $C_{(x,z)} \hat{\subset} C_{(y,z)}$ ou bien $C_{(x,z)} \hat{\supset} C_{(y,z)}$ d'après les propriétés des classes de H. Supposons alors: $C_{(x,z)} \hat{\supset} C_{(y,z)}$. On en déduit d(x,z)≥d(y,z) et par suite d(x,z)=sup[d(x,z) ; d(y,z)] (2). De plus, on a à la fois : $x \in C_{(x,z)}$ et , par conséquent $C_{(x,y)} \hat{\subset} C_{(x,z)}$

Comme l'indice d croît avec l'inclusion en raison de la décroissance de la cohésion, alors : d(x,z)≥ d(x,y) (3)

Par (2) et (3) on obtient donc (1) :d(x,y)≤sup[d(x,z), d(y,z)]

H est donc bien une **hiérarchie indicée** la distance d au sens strict de hiérarchie mathématique.

4 Hiérarchie cohésitive, modèle de la conceptualisation ?

En psychologie du développement, et, ceci afin de poursuivre l'analogie avec les conceptions piagétiennes, la notion d'abstraction réfléchissante associée à celle d'abstraction empirique et d'abstraction « réfléchie », constituant différentes étapes de « prise de conscience », tente de rendre compte du mécanisme d'équilibration majorante et de conceptualisation. Celle-ci fait passer d'un niveau de l'organisation conceptuelle (par exemple, un schème, une règle d'action, un stade donné du développement cognitif) à un niveau supérieur de cette organisation conceptuelle (le stade suivant du développement cognitif) ; chacun des niveaux étant constitué de règles, par exemple des théorèmes en acte dirait G. Vergnaud, portant sur des contenus à chaque fois différents, d'opérations portant sur ces contenus, enfin sur des opérations sur ces opérations. On retrouve, d'ailleurs, dans une théorie mathématique ces mêmes élargissements lorsque l'on passe d'un théorème qui établit l'implication d'une propriété sur une autre, à un corollaire qui fait découler d'un théorème un autre théorème ou une simple propriété. C'est le cas, par exemple, dans l'étude des fonctions, puis à l'étude de fonction de fonctions en analyse fonctionnelle. D'où l'idée que nous avons eue, de construire un second plan de relations implicatives, celui de **règles de règles** selon une **hiérarchie dite cohésitive** en raison de l'indice de **cohésion** qui permet d'engendrer des classes orientées de règles. Cet indice est à une règle généralisée ce que l'intensité d'implication est à une règle. C'est-à-dire que contrairement aux mathématiques où la validité est absolue, celle d'une règle de règle en ASI est relativisée à la valeur de la cohésion, nombre compris entre 0 et 1, croissant avec la qualité de l'implication d'une règle sur une autre.

4.1 Un détour par la philosophie des sciences.

A travers ces deux types de représentations de règles simples (graphe implicatif) ou généralisées (hiérarchie cohésitive), qui mettent au jour deux types de structures dans l'ensemble des variables, nous répondons à la philosophie structuraliste, donc non réductionniste, tout comme la théorie générale des systèmes de L. Von Bertalanffy (1973, p.66) qui estime que le **« tout » est plus riche que la somme de ses « parties ».** Citons, à ce sujet, deux extraits du livre de L. Sève (ib. p. 58) « *…le **tout** ne se compose de rien d'autre que de ses **parties**, et pourtant il présente, en tant que tout, des propriétés n'appartenant à aucune de ses parties. Autrement dit, dans le passage **non additif**, **non linéaire** des parties au tout, il y a apparition de propriétés qui ne sont d'aucune manière précontenues dans les parties et ne peuvent donc s'expliquer par elles* ». Et plus loin *« Tout se passe donc comme si se produisait une génération spontanée de propriétés du tout…C'est le paradoxe de **l'émergence**».*

4.2 Illustrations

Exemple 1

Appuyons-nous sur l'exemple 4 du § 2.5. La classe M => (F => OP3) est constituée de 3 objectifs et attitudes attendus par les professeurs de mathématiques au lycée :

- OP3 : « Dans ma notation, j'attache plus d'importance à la démarche qu'au résultat »,
- F : « développement de la capacité à prouver et valider sa preuve »,
- M : « développement de la capacité à mathématiser et à formaliser.

OP3 apparaît comme une opérationnalisation de l'objectif spécifique F. M exprime une position supérieure sur le plan cognitif, quasi-finalité de la discipline « Mathématiques ». Il y a donc bien élévation du niveau conceptuel du « théorème » F => OP3 à M. M apparaît comme large enveloppe émergeant d'une spécification.

On retrouve, de la même façon, le passage des comportements spécifiques exprimés par E => (OP8 => OP7) vers un objectif général E qui n'est pas la seule addition de OP8 et OP7 :

- OP7 : ....pouvoir reconnaître si un nombre entier écrit dans la base 10 est divisible par 4 impliqué par …
- OP8 : pouvoir donner un exemple ou un contre-exemple personnels à l'affirmation….
- E : « développement de l'imagination et la créativité ».

Nous constatons, ici aussi, le passage, par un saut qualitatif, d'un niveau d'activité élémentaire OP7 à un niveau supérieur E via un niveau intermédiaire OP8 dont OP7 est un exemple. La classe (E(OP8, OP7)) traduirait, en fait, l'émergence d'une qualité intellectuelle E génératrice d'une règle opératoire inférée par E.

Cette propriété disparaît ou s'estompe lorsque, sur le même fichier, en utilisant le même type d'algorithme de base[6], on construit avec CHIC la hiérarchie de similarité sur l'ensemble des variables. On observe (voir la Figure 3) des classes regroupant des variables vis-à-vis desquelles les enseignants ont adopté des attitudes identiques ou voisines. Par exemple, la classe {E, I, OP9, J} rassemble 3 objectifs généraux (E,I,J) de mise à distance non nécessairement disciplinaire et un savoir de type critique (réfuter vs accepter). Contrairement aux classes cohésitives ordonnées, celle-ci est homogène : les comportements qui l'illustrent relèvent du même modèle attendu de l'élève. Ce n'est donc pas comme avec la hiérarchie cohésitive l'image d'un modèle en construction mais en action.

[6] La similarité entre deux variables est évaluée à partir du nombre statistiquement étonnant d'exemples satisfaisant simultanément les deux variables eu égard aux nombres de sujets satisfaisant séparément chacune des variables. Des paires sont alors réunies en une classe. La liaison entre deux classes est établie de façon comparable.

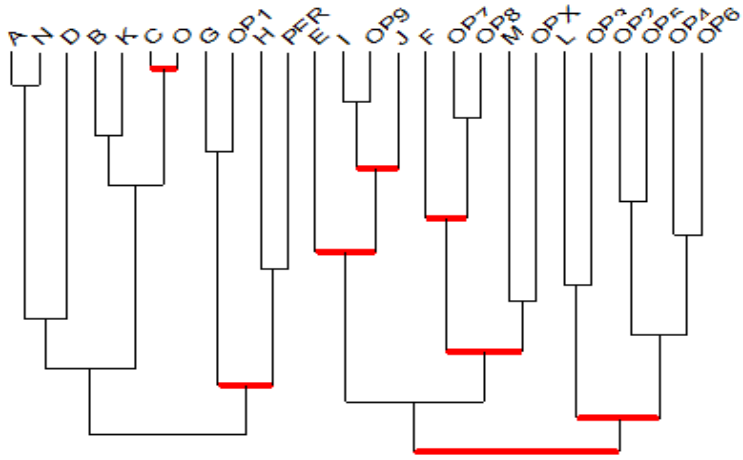


Figure 3 Hiérarchie des similarités

Exemple 2

Dans sa thèse où il a fondé l'A.S.I., Gras R. (1979) a également construit, dans le cadre de l'apprentissage des mathématiques, une taxonomie d'objectifs cognitifs selon 5 niveaux allant de la « Connaissance des outils de préhension de l'objet et du fait mathématique » jusqu'à la « Critique et évaluation ». Cette taxonomie ne décrit pas des phases d'apprentissage conceptuel cumulatives qui s'étireraient linéairement dans le temps au cours d'une scolarité, mais une hiérarchie de niveaux de complexité cognitive. Dans le but de valider ponctuellement ou simplement interroger la taxonomie[7], un questionnaire portant sur la notion de symétrie centrale (géométrique et algébrique) a été proposé à 401 élèves de 13-14 ans. Il était constitué d'une cinquantaine d'items supposés opérationnalisant a priori la taxonomie, c'est-à-dire organisant les réussites attendues selon un préordre respectant peu ou prou l'ordre taxonomique.

Les résultats obtenus par une analyse implicative montrent bien le phénomène attendu en tant qu'illustration des paragraphes plus haut. L'échelle de complexité prévue est relativement bien validée. Les items de niveaux supérieurs géométriques (resp. algébriques) (critique, évaluation,..) se placent généralement en racines des classes cohésitives en impliquant ceux de niveaux inférieurs (connaissance de faits, de définitions, de techniques,…) avec des décalages comme il s'en trouve dans des tests de psychologie différentielle. L'observation de certains décalages montre en particulier des différences de réussite très nettes entre les items algébriques et les items géométriques supposés de même niveau « piagétien ». Mais ces items, opérant dans des domaines non isomorphes, conformément à la théorie des champs, ne mobilisent pas les mêmes structures logico-mathématiques de Piaget, les mêmes schèmes, les mêmes invariants opératoires, les mêmes théorèmes et concepts en acte. On trouvera ainsi des classes

[7] C'est d'ailleurs dans cette intention que R.Gras a conçu la notion d'Analyse Statistique Implicative afin de rendre compte du préordre partiel sous-jacent aux comportements de réponse des élèves à ce questionnaire. Il ne disposait pas, parmi les méthodes d'analyse de données de moyen de mettre en évidence ce phénomène. L'échelle de Guttman ne répondait que partiellement à sa demande.

orientées spécifiques d'un champ géométrique et d'autres d'un champ algébrico-numérique.

Exemple 3

Dans Régnier et Acioly-Régnier (2007), les auteurs traitent un questionnaire sur les représentations de 320 sujets relativement aux phases de la lune. Une première analyse avait été menée déjà dans Acioly-Régnier et Régnier (2005) mais approfondie en 2007 par un usage du concept de cohésion. Ils ont été conduits, dans ces deux articles, à une interprétation dans le cadre théorique psychologique. Entre autres classes, ils s'intéressent à deux classes formées très tôt par la hiérarchie cohésitive. L'une d'entre elles, renverrait au poids de la culture, l'autre à des niveaux de conceptualisation plus élevés. On retrouve donc, dans ces textes, une démarche interprétative comparable à celle emboîtée ci-dessus.

Dans les ouvrages cités en bibliographie, on trouvera de multiples exemples tant en psychologie, sociologie, qu'en biologie, en art, en médecine, etc.

4.3 Le rapport tout-partie.

Nous insistons sur cette propriété spécifique de l'ASI que la hiérarchie cohésitive satisfait de façon originale à travers l'extension des relations entre variables. C'est par un saut qualitatif, produit généralement par un effet de seuil dans la quantité (ex. psychologie de groupe, vaporisation de l'eau…), que le tout, au prix d'une synthèse, prend son sens. Celui-ci s'extrait de la lecture véritablement dialectique du **rapport non-linéaire tout/partie**[8] (non-proportionnalité et non-additivité de la cause sur l'effet). Avec l'ASI, il y a alors **paradoxe** entre des contraires, lien et absence de lien, car le tout, constitué (on devrait dire *organisé*) de parties en un **système dynamique**, possède des propriétés que ne possèdent pas les parties et qui sont généralement de **niveau supérieur**. De la même façon, et métaphoriquement, en linguistique, la signification d'une phrase ne se fait pas uniquement par l'analyse du sens de chacun de ses mots mais par le sens de chacun de ces mots <u>inscrit</u> donné par l'interaction de ceux-ci[9]. La logique qui sous-tend ce rapport tout/parties est **dialectique** (et non pas dichotomique) car elle concilie interactivement des contradictions : règle et non-règle, **instabilité d'un système dynamique et stabilité structurelle**. Elle se fonde en règle sur l'inexistence importante du contre-exemple et en méta-règles sur la négation de l'entropie, du désordre comme on le constate dans les relations d'un fleuve et de ses affluents. En

[8].. comme le montrent les équations différentielles non-linéaires de l'indice fondamental par rapport aux paramètres cardinaux des observations, contrairement à ce qui est observable avec d'autres indices concurrents.. *« La société n'est pas constituée d'individus, mais exprime la somme des relations, des rapports où ces individus se situent les uns par rapport aux autres* » (K. Marx, « Manuscrits de 1857-1858). Une rue n'est pas la somme des maisons qui y figurent. De même une ville n'est pas somme de ses rues, etc.

[9] « *Il n'y a ni additivité ni proportionnalité entre le sens des unités (mots) et celui de la phrase. On voit se dessiner une topologie du sens* » (F. Gaudin dans « Emergence, complexité et dialectique »). « *…le mot isolé de la langue chinoise n'a en vérité ni signification ni existence à part, chacun ne reçoit sa signification que du parler même (de l'intonation, etc.…), pris isolément il a dix, voire quarante significations,… ; si nous soustrayons ce mot à la totalité, il se perd dans une creuse infinité.* « (F.-W. Schelling, « Philosophie de la mythologie », p.361).

cela, la logique dialectique s'oppose à la logique stricte (du mathématicien) sans, bien sûr, être un sophisme, c'est-à-dire un raisonnement faux. La fécondité et l'originalité de l'ASI tiennent à ce caractère, particulièrement dans l'analyse hiérarchique manifestement non-linéaire où le tout fonde son sens, non par addition des propriétés de ses parties (sous-classes) mais par la synthèse, voire par une reconstruction, des interactions inférentielles. C'est par la notion de **niveau significatif** que nous pouvons mettre en évidence le phénomène de **propriété émergente**. En ce sens, l'A.S.I. apparaît comme une sorte d'avatar de l'apprentissage non linéaire des connaissances, apprentissage fait de dépassements, de ruptures et de reconstruction dialectique (cf. l'épistémologie de G. Bachelard ou de Lev Vygotsky). A l'opposé de l'A.S.I., le rapport tout-parties serait **linéaire** dans le cas d'emboîtements de classes comme en classification fondée sur la similarité jusqu'à son nœud terminal (cf. § 3.2). Et ce linéaire infécond ne peut pas surprendre. Dans l'ouvrage « Dans la lumière et les ombres, Darwin et le bouleversement du monde » (2011), Jean-Claude Ameisen cite François Jacob : « On mesure l'importance d'une découverte à la surprise qu'elle cause » et il ajoute ; « A son caractère profondément inattendu. A ce qui nous manquait pour simplement nous y attendre » (p. 468).

5 Conclusion

A la suite de ces réflexions reliant psychologie de l'apprentissage et représentation hiérarchique de données, on peut se poser la question que se pose Gérard Vergnaud relativement à la correspondance entre signifiés et signifiants qu'il qualifie d'homomorphisme de structure : y aurait-il, de façon semblable, entre classes et sous-classes, au moins morphisme de structures préordonnées : celle de la pensée conceptualisante et celle de la hiérarchie des règles généralisées telle que la présente une hiérarchie cohésitive ? Celle-ci serait-elle une métaphore graphique de la pensée ? *Comme l'avance Thom R. (1980, p.142) au sujet de l'analogie : " ou bien elle est vraie et alors elle s'avère sterile, ou bien elle est audacieuse et alors elle peut être féconde"*. Les deux exemples présentés ici semblent soutenir cette conjecture. D'autres situations expérimentales l'ont confortée. La comparaison entre ces deux structures nous a conduits à aller au delà de la seule construction (spéculative ?) d'un outil classifiant d'analyse de données et, par suite, à donner un sens spécifique à ce mode de représentation par rapport aux modes classiques de représentation hiérarchique. *Mais, ce faisant, ne satisfait-on pas au voeu de Thom R. (1980, p.58) qui écrit encore : "…je localise l'effort théorique de la science dans sa capacité d'organiser les données de l'expérience selon des schémas imposés par des structures théoriques" ?*

Références

[1] Acioly-Régnier, N. M, Régnier, J-C, (2005), *Repérage d'obstacles didactiques et socioculturels au travers de l'A.S.I. des données issues d'un questionnaire*. Gras R., Spagnolo F., David J.(coord). Proceedings Third International Conference A.S.I. Implicative Statistic Analysis Palerme 6-8 octobre 2005, p.63-87.

[2] Agrawal R., Imielinsky T. et Swami A,(1993), Mining association rules between sets of items in large databases, *Proc. of the ACM SIGMOD' 93*, p. 207-216

[3] Bachelard G., (1967), *La Formation de l'esprit scientifique*, Paris, 5e édition, Librairie philosophique J. Vrin.

[4] Cadot M., (2009), Graphe de règles d'implication statistique pour le raisonnement courant. Comparaison avec les réseaux bayesiens et les treillis de Galois, Analyse Statistique Implicative, sous la direction de Gras R., réd, invités Gras R., Régnier J-C., Guillet F. *Une méthode d'analyse de données pour la recherche de causalités*, Cépaduès Ed. Toulouse, p.223-250

[5] Couturier R. et Ag Almouloud S., (2013), Historique et fonctionnalités de CHIC, L'analyse statistique implicative, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F *Méthode exploratoire et confirmatoire à la recherche de causalités*,., Cépaduès Editions,, p.313-325.

[6] Fayyad U., Piatetsky-Shapiro G. and Smyth P. From Data Mining to Knowledge Discovery. In., Piatetsky-Shapiro G., Smyth P, and Uthurusamy R. (eds), *Advances In Knowledge Discovery and Data Mining*, Fayyad U AAAI/MIT Press, 1-31,.

[7] Gras R., (1979), *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'Etat, Université de Rennes 1.

[8] Gras R., Kuntz P., Briand H., Couturier R. (2002), Hiérarchie de règles généralisées et notion de variable supplémentaire en analyse statistique implicative, Actes des IXèmes *Rencontres de la Société Francophone de Classification*, Université de Toulouse, 2002, p. 211-214.

[9] Gras R., Kuntz P. et Briand H. (2003), Hiérarchie orientée de règles généralisées en analyse implicative, *Extraction des Connaissances et apprentissage*, Hermès, 145-157.

[10] Gras R., Couturier R.., Blanchard J., Briand H., Kuntz P., Peter P., (2004), Quelques critères pour une mesure de qualité de règles d'association. Un exemple : l'implication statistique, *Mesures de qualité pour la fouille de données*, RNTI-E-1, Cépaduès –Editions, 3-32.

[11] Gras R., Kuntz P. et Régnier J.C., (2004), Significativité des niveaux d'une hiérarchie orientée en analyse statistique implicative, *Classification et fouille de données*, M. Chavent et M. Langlais Eds, RNTI-C-1, Cépaduès, 39-50.

[12] Gras R. et Kuntz P., (2005), *Discovering R-rules with a directed hierarchy, Soft Computing, A Fusion of Foundations, Methodologies and Applications*, Volume 1, Springer Verlag, 46-58..

[13] Guillet, F. et Hamilton H.J. (2007) (Eds.). *Quality Measures in Data Mining*. Springer.

[14] Hipp J., Guntzer U., Nakhaeizadeh, J.(2000), Mining association rules: Deriving a superior algorithm by analyzing today's approach , Proc. of 4th Eur. Conf. on *Principles of Data Mining and Knowledge Discovery*, Lect. N. in Art. Int. 1910, 160-168.

[15] Lahanier-Reuter D. (1999), Conceptions du hasard et enseignement des probabilités et statistiques, *Éducation et Formation*, PUF, Paris.

[16] Lenca P., Vaillant B., Meyer P., et Lallich S. (2007), Association Rule Interestingness Measures : Experimental and Theorical Studies, Guillet F. and Hamilton H.J. eds, *Studies in Computational Intelligence* 43, Springer, p. 51-76.

[17] Pellois, C., (2002). Apprentissage et développement de la personne dans l'enseignement et la formation, Tome 1 : *du rationnel au complexe*. l'harmattan, Collection : « Recherches et innovations sur et pour des enseignants et des formateurs », p. 105 et suivantes

[18] Pellois, C., (2008). Contrainte et liberté du sujet : entre incertitude et prévisibilité ? In Cadet B. Chasseigne G., Foliot G., *Cognition, incertitude et prévisibilité*, Editions Publiboock, Sciences Humaines et Sociales, Coll. « Psychologie cognitive », Paris, 77-100

[19] Pellois, C., (2010), Sens et incertitude, une forme de complexité en psychologie : des contraintes aux parts de liberté, le développement et ses contextes, in Cadet, C., Chasseigne, G. *Traitement de la complexité dans les sciences humaines*, Editions Publibook Université, Coll. « Psychologie Scientifique », 177-207

[20] Pellois, C., (2013). La psychologie et l'usage du traitement mathématique des données statistiques. Nouvelles perspectives conceptuelles, *Revista Brasiliera de Ensino de Ciencia e technologia*, Vol. 6, n° 1, 230-259

[21] Jean Piaget, (1970) *L'épistémologie génétique* PUF Paris

[22] Régnier, J.-C. et Acioly-Régnier, N.M. (2007) Analyse cohésitive et interprétations des données dans le champ de l'éducation. In Régis Gras et al. *Nouveaux apports théoriques à l'analyse statistique implicative et applications.* Castellón : Innovació Digital Castelló, p.329-343

[23] Thom, R (1980). *Paraboles et catastrophes*, Flammarion, Paris.

[24] Vergnaud, G. (1981). Quelques orientations théoriques et méthodologiques des recherches françaises en didactique des mathématiques – *Recherches en Didactique des Mathématiques*, 2.2, 215-232.

[25] Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactiques des Mathématiques*, 10 (23), 133-170.

[26] Vygotsky L. (1997). *Pensée et langage* (1933) (traduction de Françoise Sève, avant-propos de Lucien Sève), suivi de « Commentaires sur les remarques critiques de Vygotski » de Jean Piaget,(Collection « Terrains», Éditions Sociales, Paris, 1985) ; Rééditions : La Dispute, Paris.

[27] Von Bertalanffy, (1973, 1980) *Théorie générale des systèmes*. Paris : Dunod

Ouvrages de référence

L'implication statistique. Nouvelle méthode exploratoire de donnée, sous la direction de R.Gras et la collaboration de S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn, A.Totohasina, La Pensée Sauvage, Grenoble (1996)

Mesures de Qualité pour la Fouille de Données, H.Briand, M.Sebag, R.Gras er F.Guillet eds, RNTI-E-1, Cépaduès, 2004

Quality Measures in Data Mining , F.Guillet et H.Hamilton eds, Springer, 2007,

Statistical Implicative Analysis, Theory and Applications, R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, eds, Springer, 2008.

Analyse Statistique implicative.Une méthode d'analyse de données pour la recherche de causalités, sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, 2009.

Teoria y Aplicaciones del Analisis Estadistico Implicativo, Eds : P.Orus, L.Zemora, P.Gregori, Universitat Jaume-1, Castellon (Espagne), ISBN : 978-84-692-3925-4, 2009..

L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire. Eds : J.C. Régnier, Marc Bailleul, Régis Gras, Université de Caen, ISBN : 978-2-7466-5256-9, 2012

L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, 2013.

ANNEXE- QUESTIONNAIRE

L'APMEP conduit une réflexion sur l'enseignement des mathématiques au lycée. Elle souhaite recueillir l'opinion du plus grand nombre possible de professeurs de mathématiques. Les résultats nous aideront aussi dans la préparation de l'opération EVAPM Terminale.

Les conclusions seronty publiées dans un prochain BGV et sur le site INTERNET de l'APMEP :

http:/www.univ-lyon1.fr/apmep

An nom de quelle série répondez-vous :.....

(Vous pouvez, bien sûr, répondre pour plusieurs séries, mais utilisez un questionnaire par série)

I Objectifs de la formation mathématique

A votre avis, quels sont les objectifs essentiels de la mission d'un professeur de mathématiques dans la série pour laquelle vous répondez. Pour répondre à cette question, classez par ordre préférentiel décroissant de 1 à 6 (1 : le plus important,...) six des objectifs majeurs de cette formation en les choisissant parmi les objectifs proposés ci-dessous :

A- acquisition de connaissances
B- préparation à la vie professionnelle
C- préparation à la vie civique et sociale
D- préparation aux examens, concours, au passage dans l'enseignement supérieur
E- développement de l'imagination et la créativité
F- développement de la capacité à prouver et valider sa preuve
G- développement de la capacité d'accepter des points de vue différents
H- développement de la volonté et la persévérance
I- développement de l'esprit critique
J- développement de la capacité à communiquer avec objectivité, clarté et précision par des modes de représentation divers
K- développement de compétences utiles dans les autres disciplines
L- développement de la pratique de calculs formels, donc sans nécessité de signification
M- développement de la capacité à mathématiser et à formaliser
N- acquisition de savoir-faire
O- participation au développement d'une culture générale
Réponse (par exemple : 1 : I, 2 : G, 3 : M, 4 : D, 5 : A, 6 : J)

1	2	3	4	5	6

II Votre opinion sur des....opinions

Voici quelques opinions entendues dans la salle des profs. Vous pouvez être d'accord, ou un peu d'accord ou pas d'accord avec l'une ou l'autre. Entourez votre choix :

1-(OP1) C'est vrai que les math constituent un instrument de sélection excessif.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

2-(OP2) An bac, je préfère qu'il y ait un grand problème avec plusieurs parties plutôt qu'un ensemble de petits problèmes indépendants.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

3-(OP3) Dans ma notation, j'attache plus d'importance à la démarche qu'au résultat.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

4-(OP4) Quand je corrige, j'aime bien un barème très détaillé sur les résultats à obtenir.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

5-(OP5) La démonstration est la seule façon rigoureuse de faire des mathématiques.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

6-(OP6) Je préfère des programmes bien définis indiquant ce que je dois et ce que je ne dois pas faire.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

A la sortie de la terminale de la série sur laquelle vous répondez, un élève devrait....

7-(OP7) ....pouvoir reconnaître si un nombre entier écrit dans la base 10 est divisible par 4.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

8-(OP8) ....pouvoir donner un exemple ou un contre-exemple personnels à l'affirmation :
"si deux applications *f* et g sont strictement croissantes sur un intervalle, l'application produit *fxg* y est également croissante".

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

9-(OP9) ...avoir appris à faire un test statistique pour pouvoir réfuter ou accepter l'hypothèse d'adéquation d'une loi théorique à une distribution empirique.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

10-(OPX) ...estimer à vue, à 30% près, le périmètre et l'aire du plancher ainsi que le volume de la salle de classe.

D'ACCORD UN PEU D'ACCORD PAS D'ACCORD

AJOUT DE LA CONFIANCE AU GRAPHE IMPLICATIF

Souhila GHANEM[1] et Raphaël COUTURIER[2]

ADDING CONFIDENCE INTO IMPLICATIVE GRAPH

RÉSUMÉ

L'analyse statistique implicative développée par Régis Gras et ses collaborateurs, porte sur les mesures statistiques de qualité des règles d'associations, plus particulièrement sur celles issues de la modélisation de la loi du nombre de contre-exemples. Cette dernière basée sur le calcul de l'intensité d'implication permet d'extraire les règles les plus étonnantes sans prendre en considération le nombre de contributions des individus à chaque règle. La probabilité conditionnelle représente la fonction la plus classique servant à la confirmation inductive d'une règle. Dans ce papier nous souhaitons ajouter cette mesure probabiliste au graphe d'implication mis en œuvre dans RCHIC, et définir un seuil laissé au choix de l'utilisateur pour sélectionner toutes les règles qui lui semble intéressantes. À travers l'étude du suivi d'une population de malades, nous montrons l'intérêt de la combinaison des deux mesures, intensité d'implication et confiance.

Mots-clés : Analyse Statistique Implicative, RCHIC, Confiance, intensité d'implication

ABSTRACT

Statistical Implicative Analysis (SIA) developed by Régis Gras and his staff is based on a statistical measure that computes the interest of association rules particularly on those that use the counter-examples number law. The latter based on the computation of the implication intensity allows to extract the most surprising rules without considering the number of contributions to each rule. The conditional probability is the most classic function for the inductive confirmation of a rule. In this paper we wish to add this probability measure to the implicative graph implemented in RCHIC, and define a threshold that the user will choose to select all the rules which seem interesting to him/her. Through the study of a population of patients we show the interest of the combination of both measures, intensity of implication and confidence.

Keywords: *Statistical Implicative Analysis (SIA), RCHIC, confidence, implication intensity*

1 Introduction

Dans le cadre de l'exploration de données, la découverte de règles d'association a comme but l'extraction de règles implicatives potentiellement utiles à partir de données initialement stimulées par des recherches. La probabilité conditionnelle est la fonction la plus classique servant à la confirmation inductive d'une règle d'association. La recherche de règles d'associations a été introduite pour une application populaire qui est l'analyse de panier des clients de supermarché. L'objectif est d'étudier les habitudes d'achats des clients en cherchant les ensembles d'articles qui sont fréquemment achetés ensemble. Par exemple pour dire qu'un client qui achète du *pain* tend à acheter du *lait*

[1] Laboratoire LIMED, Université de Bejaia, souhila.ghanem@gmail.com

[2] Institut Femto-ST, Université de Bourgogne Franche-Comté, Belfort, France, raphael.couturier@univ-fcomte.fr

nous écrivons la règle d'association suivante : *pain=>lait [support=2%. Confiance=60%]*, avec le *support* et la *confiance* qui reflètent respectivement l'utilité (2% de toutes les transactions dans lesquelles le client achète le *pain* et *lait* ensembles) et la certitude (60% des clients qui achètent le *pain* achètent aussi du *lait*) de la règle découverte. Ainsi, une règle est dite intéressante ou forte si elle satisfait les seuils minimums de support et confiance. La confiance et le calcul du support fondent la stratégie utilisée dans les principaux algorithmes d'extraction, à la suite d'Apriori, Agrawal *et al.* (1994). Ces algorithmes prennent en considération le support et le seuil de confiance pour sélectionner les règles. Ces dernières sont trop nombreuses et la majorité d'entre elles n'ont aucun intérêt Freitas, A. (2000) et Gras, R. (2000). Ces algorithmes sélectionnent l'ensemble des règles dont le support et la confiance dépassent les seuils de support et de confiance préalablement fixés. Ce type d'algorithme, exhaustif et déterministe Freitas, A. (2000), produit des règles trop nombreuses dont l'intérêt n'est pas toujours assuré Gras, R. (2000). L'Analyse Statistique Implicative est une méthode d'analyse de données dédiée à l'extraction et à la structuration des règles de quasi implication, elle se distingue des autres méthodes statistiques qui permettent de générer des règles d'association, par le fait qu'elle utilise une mesure non linéaire qui satisfait des critères importants, Couturier, R. (2007, 2008). Tout d'abord, cette mesure est basée sur l'intensité d'implication qui mesure le degré de surprise inhérent à une règle. L'Analyse Statistique Implicative est utilisable par le logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive). Ce dernier a été utilisé dans plusieurs domaines, dans le cadre de la hiérarchisation des compétences tel qu'en science de l'éducation, voir par exemple Malaise, S. (2010), Montpied, P. et al (2010), Leroyer, L. (2013), Gras, R. (2008) et Demeuse, M. (2005). Certaines personnes ne sont pas satisfaites par l'ASI car la valeur de confiance de chaque règle n'est pas connu, en effet, seul le degré d'étonnement de la règle est connu.

Dans ce papier nous proposons de rassembler les deux informations : intensité d'implication et confiance dans le graphe d'implication implémenté dans RCHIC[3], pour avoir en même temps les règles les plus étonnantes et dont le pourcentage de participation est très élevé. Pour montrer l'intérêt de l'ajout de la mesure de confiance nous avons étudié les causes qui conduisent à un état de stress en utilisant des données conçues à partir des échocardiographies de stress. L'utilisation de ces deux mesures aidera les médecins à sélectionner les règles qui sont fortes en intensité d'implication et en confiance. Le choix d'un seuil de confiance élevé permet d'extraire les causes qui induisent avec une grande probabilité à avoir un état de stress. En diminuant le seuil de confiance cette probabilité d'avoir un état de stress diminue, de ce fait ce seuil permet d'avoir un graphe plus lisible

Dans la suite de ce papier nous allons calculer l'intensité d'implication et la confiance pour des données expérimentales. A travers les résultats obtenus, nous montrons comment l'expert peut utiliser ces deux informations ensemble pour extraire les règles qui l'intéressent le plus. Après nous allons appliquer notre approche avec des données issues d'une étude sur l'échocardiographie de stress des malades. Nous

[3]RCHIC est la version écrite dans R du logiciel CHIC (elle est encore en cours de développement)

présentons les expérimentations et les différents résultats obtenus avec des interprétations. Enfin nous terminons par une conclusion.

2 Motivations

Dans le but d'extraire la connaissance à partir des grandes bases de données, plusieurs algorithmes ont été conçus. Parmi, les algorithmes de type apriori, ces derniers se basent sur la recherche des règles d'association intéressantes. L'inconvénient de ces algorithmes et qu'ils produisent beaucoup de règles qui ne sont pas toutes intéressantes, et ils ignorent d'autres règles intéressantes Lallich, S et Teytaud, O. (2004). Pour pallier ces problèmes une autre méthode qui permet de structurer les données en individus et variables a été développée. Il s'agit de l'analyse statistique implicative. La méthode implicative se développe au fil de ces problèmes rencontrés et de ces questions posées. Son objectif majeur vise la structuration de données croisant individus et variables. Dans ce papier nous utilisons le graphe d'implication mis en œuvre dans RCHIC qui permet d'extraire les implications les plus étonnantes basées sur le calcul de l'intensité d'implication. Le graphe d'implication permet de donner toutes les implications étonnantes.

La question qui se pose est : est ce que toutes les règles obtenues avec l'ASI avec une intensité d'implication forte sont intéressantes ? La réponse à cette question dépend du domaine étudié et revient en principe à l'expert du domaine. Dans certains cas le nombre de règles fortes en termes d'intensité d'implication est très élevé et le graphe devient moins lisible. Ceci nous a motivé à intégrer une autre mesure qui permet de classer les règles selon leur importance. Il s'agit de calculer la valeur de confiance pour chaque règle. RCHIC calcul cette valeur et l'enregistre dans un fichier appelé **transaction.out**. En récupérant cette valeur pour chaque règle et en l'ajoutant au graphe d'implication nous pourrons déterminer les règles qui sont fortes en termes d'intensité d'implication et qui ont une confiance élevée. Ces dernières pourront être considérées comme les plus intéressantes. Pour cela nous avons effectué une expérimentation sur des données avec des valeurs représentatives, ensuite nous avons calculé l'indice d'implication et la confiance. La variable *n* correspond à la population totale, les variables *na* et *nb* correspondent respectivement aux nombres d'apparition des propriétés a et b et la variable *nabb* correspond au nombre d'apparition de *a* et de *non b*. Dans les tableaux ci-dessous nous présentons les résultats obtenus.

n	na	nb	nabb	confiance	Intensité d'implication
100	20	40	10	50%	71,81
150	20	40	10	50%	88,84
200	20	40	10	50%	93,31
250	20	40	10	50%	95,14
400	20	40	10	50%	97,03

Tableau 1 – Variation de la taille de l'échantillon n

Dans le Tableau 1, nous avons initialement pris un échantillon de *n* variables parmi lesquelles les variables *na*, *nb*, *nabb* apparaissent respectivement 20, 40 et 10 fois. La

valeur de confiance calculée est égale à 50 % et l'intensité d'implication à 71,81, par la suite nous avons fait varier la taille de l'échantillon. Nous remarquons que la confiance ne dépend pas de la taille de l'échantillon, elle est restée constante. Ceci s'explique par le fait que la confiance de la règle a=>b est égale a la probabilité conditionnelle d'avoir b sachant a, cette dernière est égale au nombre d'apparition de la variable a union b divisé par le nombre d'apparition de la variable a. Ainsi, dans le tableau 1,la confiance indépendante de la taille de l'échantillon. Par contre l'intensité d'implication augmente avec la taille de l'échantillon et montre ainsi la croissance de la surprise lorsque *n* croit.

n	na	nb	nabb	confiance	Intensité d'implication
400	20	40	2	90%	99,99
400	20	40	4	80%	99,95
400	20	40	6	70%	99,76
400	20	40	8	60%	99,07
400	20	40	10	50%	97,03
400	20	40	12	40%	92,13
400	20	40	14	30%	82,71
400	20	40	16	20%	68,13
400	20	40	18	10%	50,00

Tableau 2 – Variation de la variable nabb

Dans le Tableau 2, nous avons fixé les valeurs *n*, *na* et nb et nous avons fait varier le nombre d'apparitions de a et non b (*nabb*). Nous remarquons que les valeurs de confiance et d'intensité d'implication sont très fortes lorsque le nombre de *nabb* est faible et elles diminuent avec sa diminution. Avec l'augmentation du nombre de *nabb*les valeurs de confiance et d'intensité d'implication diminue considérablement, cela apparait logique car plus on a de contre-exemples d'une règle plus le nombre d'exemples qui vérifient la règle diminue ce qui implique la diminution de la valeur de confiance et d'intensité d'implication.

n	na	nb	nabb	confiance	Intensité d'implication
400	20	40	10	50%	97,03
400	25	40	10	60%	99,58
400	28	40	10	64,29%	99,88
400	35	40	10	71,43%	99,99

Tableau 3 – Variation de la variable na

n	na	nb	nabb	confiance	Intensité d'implication
400	20	40	10	50%	97,03
400	20	50	10	50%	96,35
400	20	55	10	50%	95,95
400	20	100	10	50%	90,16

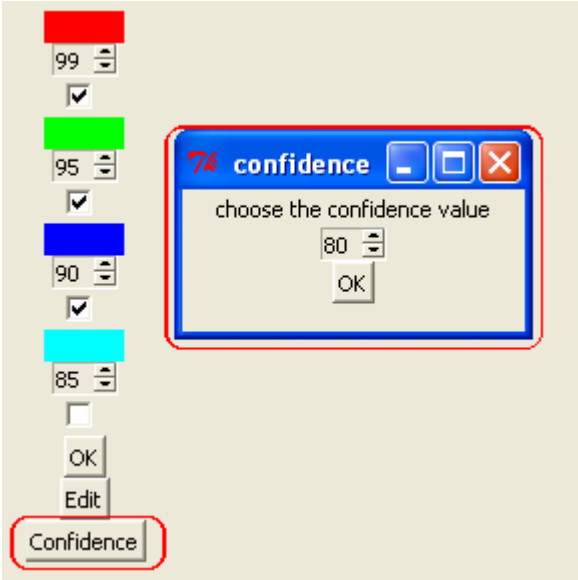
Tableau 4 – Variation de la variable nb

Nous avons fait varier la variable *na* dans le Tableaux 3 et la variable *nb* dans le Tableau 4.En augmentant la valeur de *na*, les valeurs de confiance et d'intensité d'implication augmentent. Par contre en augmentant la variable *nb* la confiance reste constante et l'intensité d'implication diminue.

À partir des résultats présentés dans ces tableaux nous constatons que la modification de la taille de l'échantillon et du nombre d'apparition de la variable en conclusion d'une règle d'implication n'ont aucune influence sur la valeur de confiance par contre en augmentant la taille de l'échantillon l'intensité d'implication augmente. En augmentant le nombre d'apparition de la variable en conclusion d'une règle l'intensité diminue. Dans les cas où la valeur de confiance est faible contrairement à la valeur de l'intensité d'implication qui est très élevée, notre approche offre la possibilité à l'expert de prendre ces implications ou de les ignorer et cela en choisissant un seuil de confiance. Ces dernières sont supposées moins importantes que les règles qui ont une valeur de confiance et d'intensité d'implication élevée qui sont supposées les plus importantes. Dans la suite nous allons voir l'impact d'élimination des règles d'implication qui ont une valeur de confiance faible et cela à travers des données issues des échocardiographies de stress.

3 L'approche proposée et son application sur des données issues des échocardiographie de stress

Notre approche consiste à combiner les deux mesures, l'intensité d'implication et la confiance. Pour cela dans RCHIC nous avons rajouté la possibilité d'afficher les valeurs de la confiance pour chaque règle en ajoutant un bouton « confidence » dans le menu du graphe d'implication. Et pour sélectionner que les règles considérées comme intéressantes nous avons rajouté une fenêtre pour choisir la valeur de confiance désirée.



La suite de cette section présente un exemple concret d'utilisation de RCHIC pour étudier les causes qui conduisent à un état de stress. Nous avons récolté des informations à partir d'une base de données sur l'analyse des données issues des échocardiographie de stress obtenue à l'Université de VANDERBILT par Frank Harrell

Harrell, F. (2015). Il a évalué les échocardiographie de stress de 558 patients jusqu'au 24 février 2015. Le jeu de données contient 31 variables.

L'échocardiographie ou bien échographie cardiaque est un test de diagnostic qui utilise des ondes ultrasonores pour créer une image du muscle cardiaque qui désigne les dimensions et le volume de cœur. Il est ainsi possible de voir la taille, la forme, et le mouvement des soupapes et cavités cardiaques ainsi que la circulation du sang à travers le cœur. L'échocardiographie peut montrer des anomalies du mauvais fonctionnement des valves cardiaques ou des dommages aux tissus du cœur à cause d'une crise cardiaque passé. Avant de présenter les résultats obtenus nous définissons les variables utilisées.

3.1 Variables utilisées

Le jeu de données utilisé comporte les variables suivantes :

bhr : Fréquence cardiaque de base
basebp : Tension de base
basepb : Le produit Bhr* basebp
pkhr: Le sommet de la fréquence cardiaque
sbp : La pression systolique[4]
dp : Le produit pkhr*sbp
dose : Dose de dobutamine[5] donnée
mbp : Pression artérielle maximale
dpmaxdo : Doubler la dose du dobutamine au maximum dans le produit
dobdose : La dose du dobutamine lorsque le produit double en maximum
gender : Le sexe, masculin ou bien féminin
baseEF : La fraction d'éjection cardiaque initiale
dobEF : Fraction d'éjection sur dobutamine
chestpain : Une douleur dans la poitrine
restwma : Repos à anomalie de mouvement de la parois sur échocardiogramme
posSE : Le stress positif échocardiogramme
newMI : Nouvelle crise cardiaque
newPTCA : Angioplastie[6] récente
hxofHT : Le malade est sujet à une hypertension
hxofDM : Le malade est sujet au diabète

[4]Correspond à la pression artérielle mesurée lors de la phase de la contraction du cœur.

[5] La dobutamine est un médicament utilisé pour l'augmentation de la contraction cardiaque, notamment en cas d'insuffisance cardiaque grave

[6] Angioplastie: une technique chirurgicale pour rétablir la circulation normale du sang dans une artère rétrécie ou bloquée par l'athérosclérose, soit par l'insertion d'un ballonnet dans la section rétrécie et le gonfler soit en utilisant un faisceau laser

hxofCig: Si le malade fume : il est soit non fumeur, fumeur modéré ou bien un grand fumeur.
hxofMI: Vérifier si le malade aune crise cardiaque aigue et grave.
hxofPTCA : Vérifier si le malade a subit la chirurgie d'angioplastie
ecg : Diagnostic de base de l'électrocardiogramme[7], permet de déduire trois cas : un cas normal, désigné par la variable acg.Normal, un cas indéterminé désigné par la variable ecg.equivocal (on ne peut pas confirmer s'il ya un trouble cardiaque ou pas) et un cas de crise cardiaque désigné par la variable ecgMI.

Pour lancer l'analyse de ces données sous RCHIC nous avons créé un fichier de type «.csv» (un extrait est montré dans la Figure 1) qui contient en lignes les identifiants des malades et en colonnes les variables. Certains variables sont suivies par la lettre « p ». Ceci précise que les variables sont à partitionner en un nombre fixe d'intervalles. Ensuite l'algorithme des nuées dynamiques Diday, E. (1971)constitue automatiquement les intervalles qui ont des limites distinctes. Nous avons choisi le partitionnement par défaut ce qui veut dire que RCHIC partitionne chaque intervalle en trois sous intervalles. Par exemple, la fréquence cardiaque de base bhr a été partitionnée en trois intervalles: élevée, moyenne et faible. Un intervalle est représenté par une variable binaire et un individu a la valeur 1 s'il appartient à cet intervalle et 0 sinon. En utilisant une telle décomposition, un individu appartient à un seul intervalle. Pour les variables qui contiennent plus d'un sens, on a attribué une variable pour chaque sens par exemple à partir de la variable sexe on a créé deux variables « male » et « female »

	bhr p	basebp p	basedp p	pkhr p	sbp p	dp p	dose p	maxhr p	pctMphr p	mbp p	dpmaxdo p	dobdose p	age p	mal	female	baseEF p	dc
1	92	103	9476	114	86	9804	40	100	74	121	12100	40	85	1	0	27	
2	62	139	8618	120	158	18960	40	120	82	158	18960	40	73	1	0	39	
3	62	139	8618	120	157	18840	40	120	82	157	18840	40	73	1	0	39	
4	93	118	10974	118	105	12390	30	118	72	105	12390	30	57	0	1	42	
5	89	103	9167	129	173	22317	40	129	69	176	22704	40	34	1	0	45	
6	58	100	5800	123	140	17220	40	123	83	140	17220	40	71	1	0	46	
7	63	120	7560	98	130	12740	40	98	71	130	12740	40	81	0	1	48	
8	86	161	13846	144	157	22608	40	144	111	157	22608	40	90	0	1	50	
9	69	143	9867	115	118	13570	40	113	81	151	17063	40	81	0	1	52	
10	76	105	7980	126	125	15750	40	126	94	125	15750	40	86	1	0	52	
11	105	134	14070	171	182	31122	40	171	108	182	31122	40	61	0	1	52	
12	72	112	8064	127	95	12065	30	125	80	101	12625	20	63	1	0	53	
13	90	120	10800	169	184	31096	40	169	126	184	31096	40	86	1	0	54	
14	81	110	8910	110	130	14300	40	110	58	130	14300	40	29	0	1	55	
15	84	176	14784	110	194	21340	40	110	74	194	21340	40	71	0	1	55	

Figure 1 – Extrait du jeu de données

RCHIC calcul la valeur de confiance et d'indice d'implication pour chaque règle, et les enregistre dans un fichier appelé **transaction.out**. Le fichier transaction.out

[7]L'électrocardiographie (ECG) désigne l'examen permettant l'enregistrement du rythme cardiaque. L'ECG consiste à étudier précisément l'activité du cœur, grâce à des électrodes posées sur la poitrine, les poignets et les chevilles. Cette activité est mesurée en plusieurs points du cœur, appelés dérivations, et elle est enregistrée sous la forme d'une courbe pour chacune d'entre elles. L'ECG permet de découvrir des troubles du rythme cardiaque, des troubles de la conduction cardiaque, des signes de souffrance cardiaque.L'ECG peut être utilisé pour identifier une population de patients à faible risque de MI (crise cardiaque). Un ECG normal indique moins de 3% pour le risque de MI et moins de 6% pour le risque de décès dans l'année choisit

correspondant au jeu de données issu des données des échocardiographies de stress contient 3986 règles. Voici une partie de ce fichier :

hyp -> con	occurrence(	occurren	suppor	confidence	classical index	entropic index
maxhr.3 -> dose.2	128.000000000	125.000000	22.939068	28.90625000000	79.8263922333717	33.8972340076598540
dose.2 -> maxhr.3	125.000000000	128.000000	22.401433	29.59999999999	80.1877319812774	34.8368000960555620
maxhr.3 -> HxofCig.he	128.000000000	122.000000	22.939068	27.34375000000	75.8468195796012	31.9113793722364250
HxofCig.heavy -> maxh	122.000000000	128.000000	21.863795	28.68852459016	76.5288650989532	33.7130189901168580
maxhr.3 -> bhr.3	128.000000000	108.000000	22.939068	38.28125000000	99.1447148844599	46.7913717142461320
bhr.3 -> maxhr.3	108.000000000	128.000000	19.354838	45.37037037037	99.6040620841085	58.2992773010986060
maxhr.3 -> dobEF.1	128.000000000	100.000000	22.939068	16.40625000000	42.4976348876953	18.6547383807354310
dobEF.1 -> maxhr.3	100.000000000	128.000000	17.921146	21.00000000000	41.2589073181152	24.4447615698129790
maxhr.3 -> dp.3	128.000000000	88.0000000	22.939068	39.06250000000	99.7955969069153	48.0348474151621050
dp.3 -> maxhr.3	88.0000000000	128.000000	15.770609	56.81818181818	99.9942285424292	80.2317373407175440
maxhr.3 -> basedp.3	128.000000000	88.0000000	22.939068	27.34375000000	92.3162840306758	32.0774828447578530
basedp.3 -> maxhr.3	88.0000000000	128.000000	15.770609	39.77272727272	96.3981401175260	49.5371237867437630
maxhr.3 -> hxofCABG	128.000000000	88.0000000	22.939068	15.62500000000	49.2839395999908	17.7773868543899950
hxofCABG -> maxhr.3	88.0000000000	128.000000	15.770609	22.72727272727	49.0971505641937	26.7144851954282670
maxhr.3 -> dpmaxdo.3	128.000000000	84.0000000	22.939068	38.28125000000	99.7822542442008	46.9033240263058960
dpmaxdo.3 -> maxhr.3	84.0000000000	128.000000	15.053763	58.33333333333	99.9961514771431	81.9021790430464590
maxhr.3 -> baseEF.1	128.000000000	80.0000000	22.939068	10.93750000000	33.8873445987701	12.3471268562275060
baseEF.1 -> maxhr.3	80.0000000000	128.000000	14.336917	17.50000000000	28.9727151393890	20.4322975199196330

Figure 2 – Extrait du fichier transaction.out correspondant au jeu de données.

Le graphe d'implication est constitué en ne représentant que les implications dont la valeur d'intensité d'implication est supérieure à un seuil choisi par l'utilisateur. Notre approche rajoute le seuil de confiance. Les règles en rouge dans la figure 2 correspondent à des règles d'implication très forte en termes d'intensité d'implication pour lesquelles la confiance est faible, c'est-à-dire que le nombre d'individus qui les réalisent est relativement faible. C'est ici qu'on peut voir l'intérêt de notre approche, qui donne la possibilité à l'expert premièrement d'avoir l'information sur la confiance de chaque règle et la possibilité de prendre ces règles ou de les laisser.

3.2 Résultats obtenus sans spécifier un seuil de confiance

Dans le graphe implicatif nous avons offert à l'utilisateur la possibilité d'afficher la valeur de confiance de chaque règle. Dans la Figure 3, vu le grand nombre de règles, nous avons montré qu'une partie du graphe d'implication. Dans les graphes d'implications qui suivent la flèche rouge désigne un seuil d'intensité d'implication égale à 99, La flèche bleueun seuil d'intensité d'implication égale à 95 et la flèche verte désigne un seuil d'intensité d'implication de 90.

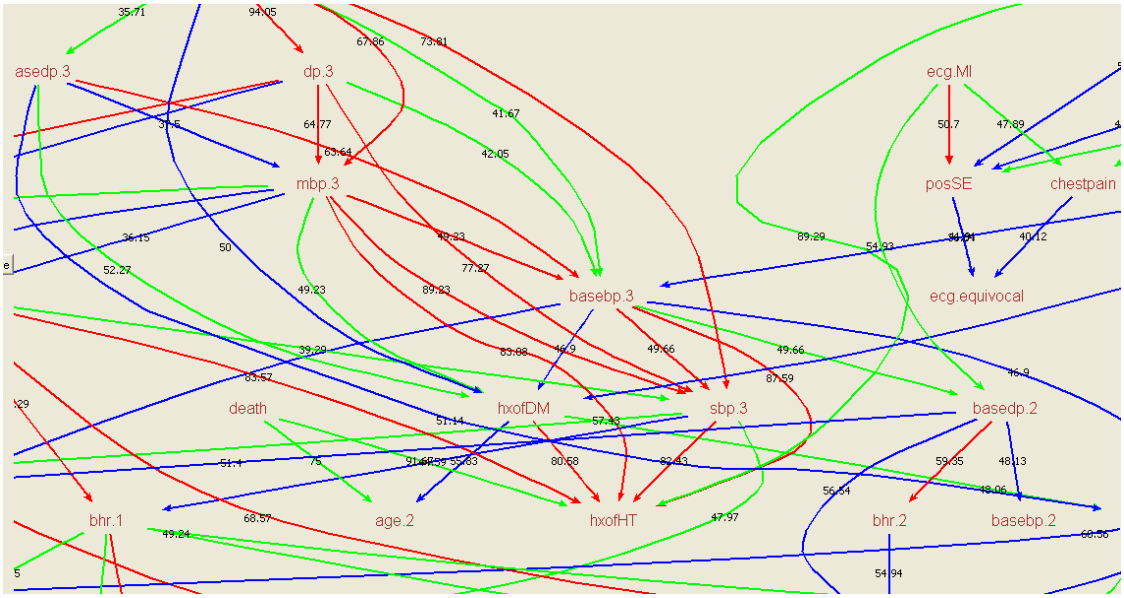


Figure 3 – Extrait du graphe d'implication avec l'affichage des valeurs de confiance

En affichant les valeurs de la confiance, l'utilisateur peut voir l'information sur le nombre de participation des individus à chaque règle, mais vu le grand nombre d'implications l'utilisateur aura des difficultés pour distinguer les règles.On a l'impression que tout est lié. Pour cela nous avons rajouté la possibilité de choisir un seuil de confiance qui nous donne la possibilité de voir le niveau d'importance des règles. Dans ce qui suit nous allons présenter le graphe d'implication en affichant juste les implications qui correspondent au seuil de confiance désiré.

3.3 Résultats obtenus en utilisant un seuil de confiance égale à 80

Le choix d'un indice de confiance élevé va nous permettre d'extraire les règles d'implications dont le taux de participation est très élevé, cela revient à identifier les facteurs les plus importants qui conduisent à un état de stress. Ci-dessous le graphe d'implication correspondant:

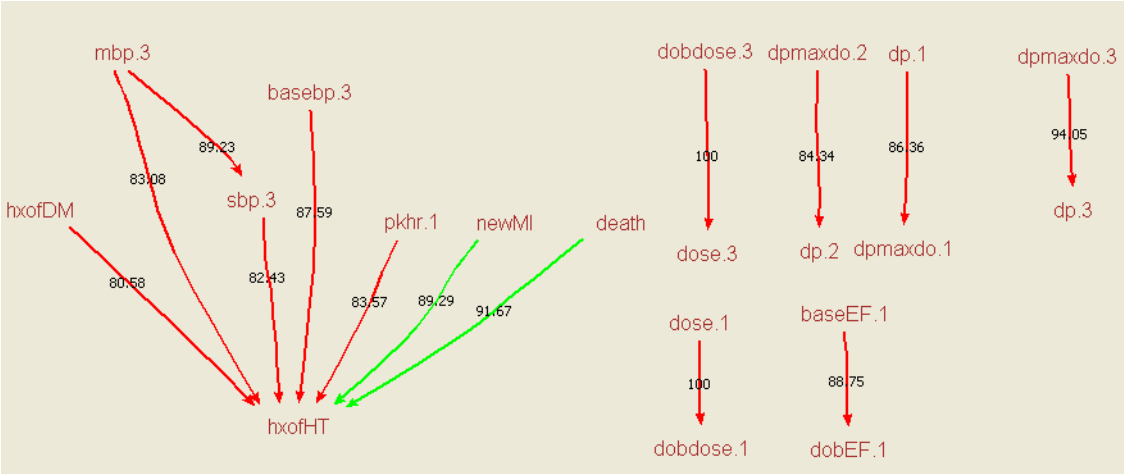


Figure 4 – Graphe implicatif avec un seuil de confiance égale à 80

Avec une valeur de confiance 80 nous remarquons que le graphe est plus lisible et que les résultats sont cohérents car les valeurs faibles impliquent les valeurs faibles et les valeurs élevées impliquent les valeurs élevées. Parmi les implications nous rencontrons les suivantes :

Si la personne est diabétique alors elle a un grand risque de se retrouver en état de stress.

Si la fréquence cardiaque de base est faible alors la personne a tendance à être en état de stress.

Si la pression sanguine maximale et la tension de base sont très élevées alors la personne risque de se retrouver en état de stress.

Si la personne subit une crise cardiaque c'est qu'elle a tendance à se retrouver en un état de stress.

À partir de ces implications nous constatons que les facteurs qui conduisent le plus à un état de stress sont le fait d'avoir:

Le diabète

Une fréquence cardiaque faible

Une pression sanguine maximale très élevée

Une tension de base élevée

Une nouvelle crise cardiaque ;

Car les valeurs de confiance sont très élevées.

3.4 Résultats obtenus en utilisant un seuil de confiance égale à 70

La figure ci-dessous montre le graphe d'implication correspondant aux résultats obtenus avec un seuil de confiance égale à 70

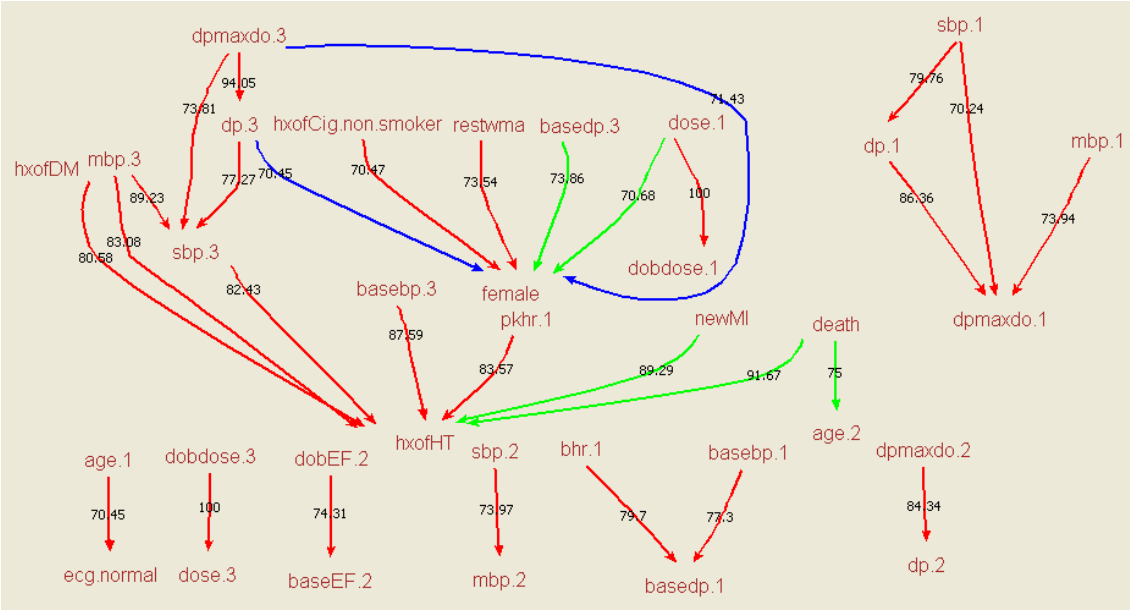


Figure 4 – Graphe implicatif avec un seuil de confiance égale à 70

Avec un seuil de confiance égale à 70, nous remarquons qu'on a, en plus des règles dans le graphe précédent, d'autres règles mais avec un seuil de confiance moins élevé. Ces implications correspondent aux suivantes :

- Les médecins ont utilisé une dose très élevée du double de dobutamine pour 71% des femmes.
- Dans le jeu de données utilisé, plus de 70 % des personnes non fumeurs sont des femmes.
- 73% des personnes qui ont l'anomalie de mouvement de la paroi au repos sont des femmes.
- Les personnes très jeunes ont un diagnostic normal.

3.5 Résultats obtenus avec un seuil de confiance égale à 65

La figure ci-dessous montre le graphe d'implication correspondant aux résultats obtenus avec un seuil de confiance égale à 65

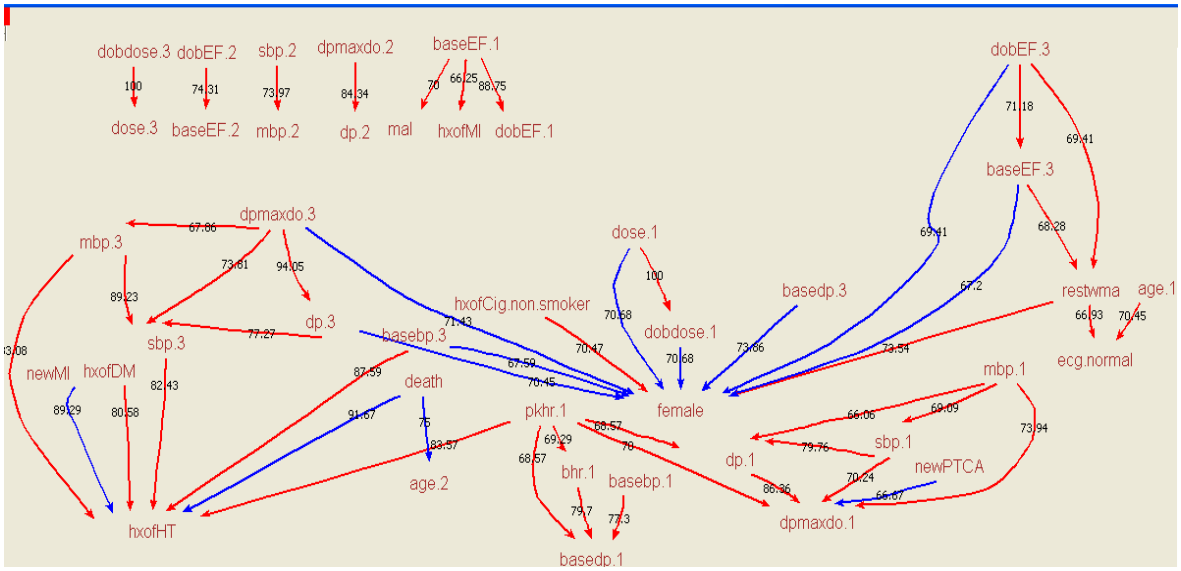


Figure 5 – Graphe implicatif avec un seuil de confiance égale à 65

En diminuant le seuil de confiance à 65, plus d'implications apparaissent parmi lesquelles on voit que:

66% des personnes qui ont l'anomalie de mouvement de la paroi au repos ont tendance à avoir un diagnostic ECG normal.

68 % des personnes qui ont la fraction d'injection cardiaque initiale très élevée et 69% qui ont la fraction d'injection cardiaque élevée sur dobutamine ont tendance à avoir l'anomalie de mouvement de la paroi au repos.

4 Conclusion

Dans ce papier nous avons étudié le couplage de l'indice d'implication avec la confiance afin d'améliorer de faciliter la recherche de règle avec la méthode d'analyse statistique implicative. Notre but est d'aider les experts et utilisateurs à utiliser le graphe d'implication. Pour cela, nous avons rajouté la valeur de la confiance pour chaque règle dans le graphe, ce qui permet de distinguer le niveau d'importance de chaque règle. Nous avons aussi rajouté un seuil de confiance afin de rendre le graphe plus lisible. Après avoir défini l'approche et l'avoir mise en œuvre dans le logiciel de statistique RCHIC, nous avons appliqué notre approche sur différents ensembles de données. Dans ce papier nous avons choisi les données issues des échocardiographies de stress. Les résultats obtenus nous ont permis d'avoir des graphes plus lisibles et d'identifier et de classer les facteurs les plus importants qui conduisent à un état de stress.

Comme perspective à ce travail nous envisageons d'utiliser d'autres mesures que la confiance parmi celles disponibles Guillet, F. et Hamilton, H. J. (2007). De plus, nous pourrions comparer les résultats présentés dans cet article à ceux obtenus en utilisant d'autres mesures telles que l'estimateur de Laplace, DIR Blanchard, J. (2005) ou IPEE Blanchard et al (2005).

Références

[1] Agrawal, R. and Srikant, R. (1994), Fast algorithms for mining association rules, *Proceedings of the 20th VLDB Conference*, Santiago, Chile.

[2] Freitas, A. (2000), Understanding the crucial difference between classification and discovery of association rules - a position paper. SIGKDD *Explorations, vol. 2, 1*, pp.65-69.

[3] Gras, R. (2000), Les Fondements De L'analyse Statistique Implicative, *Quaderni di Ricerca in Didattica*

[4] Couturier, R. (2007). CHIC : utilisation et fonctionnalités, Nouveaux Apports Théoriques à l'Analyse Statistique Implicative et Applications, p. 41-49.

[5] Couturier, R. (2008). CHIC: Cohesive Hierarchical Implicative Classification, In Statistical Implicative Analysis, volume 127 of *Studies in Computational Intelligence*, pages 41-52. Springer.

[6] Malaise, S. (2010), Classification hiérarchique de compétences par l'intermédiaire du logiciel CHIC fondée sur méthode d'analyse statistique implicative, *Actes du congrès de l'actualité de la recherche en éducation et en formation (AREF), université de Genève.*

[7] Montpied, P. et al (2010), La diversité des désirs d'apprendre : quelles informations pratiques l'Analyse Statistique Implicative fournit-elle en regard des curriculums et de l'enseignement des sciences, *A.S.I. 5 Proceedings* 5-7.

[8] Leroyer, L. (2013), Le rapport au support dans le travail de préparation en mathématiques des enseignants du premier degré. pp7-1.

[9] Gras, R. (2008), *Statistical Implicative Analysis Theory and Application*. Book Springer.

[10] Demeuse, M. (2005). Introduction aux théories et aux méthodes de la mesure en psychologie et en sciences de l'éducation. Liège : Les *éditions de l'Université de Liège*.

[11] Lallich, S et Teytaud, O. (2004), Évaluation et validation de l'intérêt des règlesd'association, *Revue des Nouvelles Technologies de l'Information (RNTI)*, Actes EGC-2003: 193-218.

[12] Diday, E. (1971), La méthode des nuées dynamiques,*Journal Revue de statistique appliquée*, volume 19, numéro 2, pages (19-34).

[13] Harrell, F.(2015),http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/CstressEcho.html

[14] Guillet, F. et Hamilton, H. J. (2007). Quality Measures in Data Mining. Springer.

[15] Blanchard, J. (2005).Un système de visualisation pour l'extraction, l'évaluation, et l'exploration interactives des règles d'association, Human-Computer InteractionUniversité de Nantes

[16] Blanchard, J. Guillet, F. Gras, R et Briand, H. (2005), Using information-theoretic measures to assess association rule interestingness», in Proceedings of the fifth IEEE international conference on data mining ICDM'05, IEEE Computer Society.

EXTENSION DE L'ANALYSE STATISTIQUE IMPLICATIVE AU CAS DES VARIABLES CONTINUES QUELCONQUES

Régis GRAS[1], **Jean-Claude REGNIER**[2]

EXTENSION OF THE IMPLICATIVE STATISTICAL ANALYSIS TO ANY CONTINUOUS VARIABLE

RESUME

Après avoir généralisé l'Analyse Statistique Implicative au cas où l'espace des sujets est continu, nous étendons son champ d'application au cas où cette fois les espaces des variables sont continus sur [0; 1]. Ainsi, les variables seront observées sur des intervalles munis d'une loi de répartition continue. Nous procédons, tout d'abord, à l'extension à partir du traitement connu en ASI des variables-intervalles. Puis, nous envisageons un cas particulier où les distributions sur les espaces des variables suivent une même loi uniforme. Enfin, nous traitons le cas général de l'extension aux espaces de variables munis de lois différentes et quelconques.

Mots-clés *: variables, variable-intervalle, variable continue, densité, intensité d'implication, propension.*

ABSTRACT

After having generalized the implicative statistical analysis to the case in which the subjects' space is continuous, we extend its field of application to the case in which variables' spaces are continuous on [0 ; 1]. Thus, the variables will be observed as interval-variables. We present a specific case where the distributions of the variables' spaces follow the same uniform law. Finally, we treat the general case of the extension to variable spaces with various and undefined laws.

Keywords: *variables, interval-variable, continuous variable, density, intensity of application, propensity.*

1 Introduction

Une matrice de données numériques croisant classiquement sujets et variables étant donnée, l'Analyse Statistique Implicative (ASI ou SIA en anglais) attribue une mesure de qualité à des énoncés du type « si un sujet satisfait la variable a alors, généralement, il satisfait la variable b ». Nous avons traité (Gras et Régnier, 2012) le cas où, en ASI, l'espace des sujets E est continu muni d'une loi de répartition donnée. Nous abordons ici la situation où, cette fois, les variables actives elles-mêmes sont continues. Quel est le sens de cette continuité ? Elle exprime que, dans la population-mère de sujets, les observations de la variable suivent une loi continue, par exemple une loi gaussienne ou

[1] Ecole Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire d'Informatique de Nantes-Atlantique (LINA), UMR 6241, Site de la Chantrerie, rue C.Pauc, BP 44306, Nantes cedex 3, e-mail : regisgra@club-internet.fr,

[2] Laboratoire UMR 5191 ICAR – Université Lumière de Lyon –Lyon2, 86 rue Pasteur, 69635 LYON Cedex 07, e-mail : jean-claude.regnier@univ-lyon2.fr

une loi uniforme. La population E de la matrice des données est censée en être une fidèle réalisation.

Précisons. Le type de variable générique que nous étudierons est une variable numérique dont la distribution est continue, de densité donnée et dont les valeurs finies sont prises sur IR ou sur un intervalle de IR. Si sa nature initiale est qualitative, nous nous limiterons au cas où cette « qualité » est quantifiable et ordonnée. Citons, par exemple, des variables qui expriment des saveurs, des sensations, des phénomènes climatiques,… On rencontre en effet ce type de situation dans le cas de données sensorielles. Mais aussi, comme dans (Gras et al, 2001), à la faveur du recueil des notes attribuées à des élèves dans des disciplines variées et conduisant à des intervalles de IR.

L'ensemble V des variables a, b, c… de l'étude peut être constitué de telles variables continues de distributions souvent différentes, de densités de mesure : f_a, f_b, f_c,… et présenter des espaces de valeurs également différents que l'on peut tous ramener par une homothétie convenable, à l'intervalle [0 ; 1]. Si X est la variable aléatoire représentant la valeur de a inférieure à α, alors Prob [X<α] = $\int_0^{\alpha} f_a(t)dt$ et de même : Prob [X≥α] = $\int_{\alpha}^{1} f_a(t)dt$ représente la probabilité pour que la valeur de a soit supérieure ou égale à α. Si a(s) (resp. b(s)) sont les valeurs observées chez le sujet s selon a (resp . b), nous évaluerons dans quelle mesure on observe « rarement » dans E : a(s) > b(s).

Pour chaque sujet s de E, espace des sujets supposé ici discret et fini, nous disposons de la valeur numérique réelle prise selon chacune des variables soit : a(s), b(s), c(s),…Ces valeurs traduisent, dans la plupart des cas sémantiques, un degré d'intensité d'adhésion du sujet s à la variable ou de satisfaction de celle-ci. Elles sont, rappelons-le, manifestement ordonnées sur IR ou un intervalle réel.

Dans cet article, nous revenons d'abord sur la situation traitée dans le cadre des variables-intervalles où chaque intervalle est pondéré par la fréquence de ses instanciations. Le découpage de l'ensemble des valeurs de chaque variable par une partition optimale permet de définir les implications d'intervalles. Dans la section suivante 2, nous étudions le cas où les variables sont continues et uniformément distribuées sur l'ensemble de leurs valeurs. Nous définissons alors un indice de propension entre ces variables. Dans les deux dernières sections 3 et 4, nous étendons l'étude au cas où les variables ont une distribution quelconque mais, possiblement, différentes entre elles. Dans un premier temps, nous traitons le problème par la méthode générale de l'implication statistique. Ensuite, nous envisageons la même situation mais par l'approche propensive comme dans la section 2.

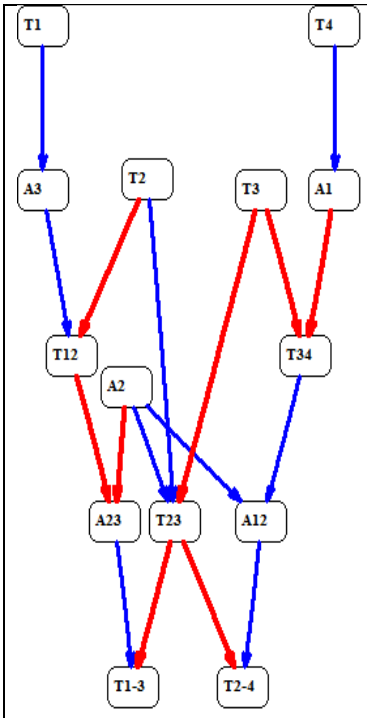
2 Première approche

Pour l'ensemble des variables a, b, c,…, nous examinons, sur leur espace de réalisation ramené à [0, 1], l'ensemble des valeurs prises par les sujets de E selon leurs densités respectives f_a, f_b, $f_{c,\ldots}$non explicitement définies. Notre objectif est de décomposer l'intervalle des valeurs de chaque variable sous forme de sous-intervalles réels. Pour ce faire, sur la base de la répartition des valeurs prises par les n sujets, nous

recherchons la meilleure partition de ces sous-intervalles en un nombre k pour la variable a (l pour b, m pour c, …) maximisant la variance interclasse pour un nombre fixé de sous-intervalles. Cette partition est effectuée selon la méthode inspirée et élargie à partir de celle des « Nuées dynamiques » de E. Diday (1972) et que nous avons utilisée précédemment pour définir l'intensité d'implication entre les variables-intervalles (Gras et al, 2001, 2013). Supposons la partition optimale obtenue constituée des sous-intervalles distincts A_1, A_2,…, A_k (resp. B_1, B_2,…B_l et C_1, C_2, …C_m), eux-mêmes avatars de sous-variables a_1, a_2, …, a_k (resp. b_1, b_2, …,b_l et c_1, c_2, …,c_m.) L'ensemble des valeurs prises par l'ensemble des sujets sur ces sous-intervalles est fini et est représentable par un intervalle de IR, réunion des A_i, que nous pouvons ramener à [0 ; 1] par une homothétie bijective. Nous convenons qu'à A_1 (resp. à A_k), par exemple, s'agrègeront les valeurs non observées « en bout » respectivement à gauche et à droite. Le tableau de données original est remplacé par un nouveau tableau à valeurs binaires. Sur la ligne du sujet x qui prend sa valeur en a selon la modalité a_i (ou de façon équivalente dans A_i) par exemple, on note la nouvelle valeur $1(A_i)$. (mesure de A_i). Ainsi, pour tout x qui satisfait a_i, alors $a_i(x) = 1$ et $a_j(x) = 0$ pour les autres sous-intervalles. Les différentes intensités d'implication des a_i sur les modalités b_j et vice-versa, sont calculées comme pour des variables binaires sous-intervalles par sous-intervalles. Un algorithme permet également de conjoindre, des sous-intervalles contigus d'une variable et d'évaluer les implications optimales d'une variable segmentée vers une autre. Puis les règles sont représentées par un graphe implicatif.

Exemple

On cherche à valider l'hypothèse (certes très vraisemblable) que, généralement, « être de grande taille » en 2013 implique « être jeune ». On dispose pour cela d'un échantillon, pris au hasard dans la population de français, de 77 sujets dont les âges varient de 19 ans à 60 ans et les tailles de 168 cm à 198 cm.



On obtient par C.H.I.C., logiciel de traitement de l'ASI, le graphe implicatif ci-contre lorsque l'on demande que les données numériques de la variable A (âges) soient partitionnées en 3 sous-intervalles optimaux et celles de la variable T (tailles) en 4 sous-intervalles optimaux :

Partitions optimales des données de la variable A :

A1 de 19 à 31 (soit moins de 31)

A2 de 32 à 43

A3 de 46 à 60 (soit plus de 46)

Partitions optimales des données de la variable T:

T1 de 168 à 172 (soit moins de 172 cm)

T2 de 173 à 176

T3 de 178 à 184

T4 de 185 à 198 (soit plus de 185 cm)

On observe que les grandes tailles impliquent (intensité d'implication 0.95) plutôt les jeunes : T4 => A1 (ce qui ne surprendra pas le lecteur) et que ces mêmes jeunes impliquent les tailles de 178 à 198 : A1 => T34

Remarques

1° La prise en compte des mesures des sous-intervalles permet de limiter l'intérêt aux implications à la hauteur de leur importance pondérale.

2° Par la finesse du découpage (k sous-intervalles) certes d'une part on multiplie le temps de calcul, mais d'autre part, on améliore la représentation des nuances de la distribution des instanciations.

3° Une autre approche du problème des variables continues, comparable à l'extension de l'espace des sujets au cas continu, sera à l'étude dans les trois sections suivantes.

4° Ainsi, la distribution a posteriori définie pour chaque variable telle que a à partir de la partition de A en sous-intervalles est explicitement donnée sous forme d'histogramme sur l'intervalle [0 ; 1]. Elle est une des réalisations contingentes de la distribution a priori de a soit f_a. De ce fait, l'implication entre intervalles pourrait être désignée comme implication entre histogrammes.

3 Cas où la variable est uniformément distribuée sur [0 ; 1].

Les extensions de traitements de données dans une approche ASI, développées à partir des premiers travaux de R. Gras, par M. Bailleul (1994) et J.B. Lagrange (1998),

considèrent les variables à valeurs sur [0 ; 1] comme des variables discrètes. Ce que nous souhaitons introduire maintenant, est la prise en considération de variables continues à valeurs sur l'intervalle [0 ; 1] ou s'y ramenant. Ici, nous allons, dans un premier temps, étudier le cas des variables continues sur [0 ; 1] dont la distribution de probabilité est uniforme sur cet intervalle et identique pour toutes les variables.

3.1 Propriétés de la variable Z, produit de deux variables aléatoires continues de loi uniforme sur [0 ; 1]

La problématique de l'étude présente s'exprime ainsi :

Exprimer par un nombre compris entre 0 et 1, dans quelle mesure l'observation, sur un ensemble de sujets E, des valeurs (numériques ou numérisées) prises par une variable continue a s'accompagne généralement de l'observation de valeurs plus grandes prises par une variable continue b.

Nous satisfaisons bien ainsi la philosophie de l'Analyse Statistique Implicative puisque nous voulons quantifier la tendance de la variable a à prendre des valeurs sur E inférieures aux valeurs prises selon la variable b. Cette quantification porte le nom de **propension**. Nous suivons la même procédure que celle suivie par J.-B. Lagrange à propos des variables modales discrètes. Soient X et Y variables aléatoires continues à valeurs dans [0 ; 1] de loi uniforme continue. Elles représentent les valeurs aléatoires des variables a et b qui se réaliseront dans l'observation. Comme il est coutumier en Analyse Statistique Implicative, les deux variables X et Y sont supposées indépendantes. Nous savons alors que la densité de probabilité de X ou de Y est donnée par la fonction $f(u)=1_{[0\,;\,1]}(u)$ de laquelle nous pouvons calculer la fonction de répartition :

$$F(u) = \begin{cases} 1 & u > 1 \\ u & 0 \le u \le 1 \\ 0 & u < 0 \end{cases}$$

Par ailleurs sans difficulté on montre que les espérances respectives de X et Y valent $E(X) = E(Y) = \frac{1}{2}$ et les variances $V(X) = V(Y) = \frac{1}{12}$.

De manière évidente, la variable $\overline{Y} = 1 - Y$ est encore une variable aléatoire à valeur dans [0 ; 1] de loi uniforme continue dont l'espérance et la variance sont celles de Y.

Poursuivant le chemin réalisé dans la construction de l'indice de propension établi avec des variables modales, nous posons Z = X(1-Y). Z est encore une variable aléatoire continue à valeurs dans [0 ; 1] mais qui n'admet plus une densité uniforme. On montre que la densité de probabilité de Z est $g(u) = -\ln(u)\, 1_{]0\,;\,1]}(u)$ ou encore que sa fonction de répartition est :

$$G(u) = \begin{cases} 1 & u > 1 \\ -u\ln(u) + u & 0 < u \leq 1 \\ 0 & u \leq 0 \end{cases}$$

À partir de ces informations, nous pouvons calculer l'espérance puis la variance de la variable aléatoire continue à valeurs sur [0; 1] de la manière suivante:

E(Z) = $\int_{0+}^{1} uf(u)du = \left[-\frac{1}{2}u^2\ln(u) + \frac{1}{4}u^2\right]_{0+}^{1} = \frac{1}{4}$ mais comme Z est le produit de deux variables indépendantes, nous pouvons obtenir directement : E(Z) = E(X) E(1-Y) = $\frac{1}{4}$

Pour ce qui est de la variance, nous savons que :

V(Z) = E[(Z-E(Z))^2] = E(Z^2)-[E(Z)]^2

Or E(Z^2) = $\int_{0+}^{1} u^2 f(u)du = \left[-\frac{1}{3}u^3\ln(u) + \frac{1}{9}u^3\right]_{0+}^{1} = \frac{1}{9}$

donc la variance de Z vaut V(Z) = $\left(\frac{\sqrt{7}}{12}\right)^2$

3.2 Indice de propension et intensité de propension pour des variables aléatoires continues de loi uniforme sur [0 ; 1]

Pour suivre la procédure de construction qui mène à la formule de l'indice de propension, nous considérons le n-uplet de variables iid (Z_1, Z_2,…, Z_n).

Posons $T_n = \frac{1}{n}\sum_{k=1}^{k=n} Z_k$. Il en résulte que $E(T_n) = \frac{1}{n}\sum_{k=1}^{k=n} E(Z_k) = \frac{1}{4}$ ainsi que $V(T_n) = \left(\frac{\sqrt{7}}{12\sqrt{n}}\right)^2$ puisque les variables Z_k sont indépendantes deux à deux. Nous sommes en mesure de fournir l'expression algébrique exacte de l'indice de propension entre les deux variables X et Y.

$$Q_u(X,\overline{Y}) = \frac{T_n - \frac{1}{4}}{\frac{\sqrt{7}}{12\sqrt{n}}} = \sqrt{n}\,\frac{T_n - \frac{1}{4}}{\frac{\sqrt{7}}{12}} = \sqrt{n}\,\frac{\left[\frac{1}{n}\sum_{k=1}^{k=n} X_k(1-Y_k)\right] - \frac{1}{4}}{\frac{\sqrt{7}}{12}}$$

On peut considérer que la variable aléatoire $Q_u(X,\overline{Y})$ indice de propension entre deux variables continues de loi uniforme sur [0 ; 1] suit approximativement une loi de Laplace-Gauss centrée réduite dans la mesure où sa construction même nous place dans les conditions d'un des théorèmes central-limite.

L'intensité d'implication se calcule donc de la façon suivante :

$$\varphi_u(X,Y) = 1 - \Pr ob\left[Q_u(X,\bar{Y}) \le q_u(X,\bar{Y})\right] = \frac{1}{\sqrt{2\pi}} \int_{q_u(X,\bar{Y})}^{\infty} e^{-\frac{t^2}{2}} dt$$

4 Cas général traité par la méthode ASI classique

4.1 Problématisation

Pour chaque variable observée, nous avons jusqu'alors examiné, sur son espace de réalisation, l'ensemble des valeurs prises sur l'ensemble des sujets E. Nos travaux sur les variables numériques et modales ont porté, comme dans la section 2, sur la définition d'une mesure de propension dont nous venons de rappeler une modélisation. A l'instar de J.B. Lagrange (1998) et de S. Guillaume (2000), nous utilisons encore l'expression : **propension** (ou tendance) de a vers b, dès lors que l'on rencontre généralement dans E peu de transactions s∈ E dans lesquelles a(s) > b(s)) pour la relation d'ordre sur [0 ; 1] où a(s) et b(s) sont respectivement les valeurs de a et b observées en s. Notons que si a(s) et b(s) sont des valeurs de rang ou d'intensité préférentielle, on peut exprimer l'inégalité a(s) > b(s) par : « a est préférée à b en s »

Notre exploration implicative, notée encore a=>b, dans le cas des variables continues est ici plus générale et s'exprime de la façon suivante :

Dans quelle mesure peut-on considérer que, pour x∈ E « ***si a(x) ≥ α, alors on peut affirmer que b(x) ≥ β*** » (F1) où α et β sont des réels de l'intervalle [0 ; 1] choisis dans l'analyse. Autrement dit, observe-t-on généralement que : « *si une réalisation pour x selon a dépasse la valeur α alors sa réalisation selon b dépasse aussi la valeur β* ? » (F2). Cette fois, contrairement à ce qui est envisagé dans la section 2, ce sont toutes les valeurs de deux intervalles, pondérés par des densités, qui sont mobilisées. De plus, les bornes de ces intervalles peuvent être indépendantes et de sémantique différentes. La réponse confère à la mesure associée une valeur prédictive.

Notons que si $\alpha = \beta$ égalité à laquelle on peut toujours ramener les deux ensembles de valeurs des deux variables, nous retrouvons la formulation de l'ensemble des contre-exemples, en tout point comparable à celle de la propension conceptualisée précédemment pour les variables numériques (Gras et al, 2009, 2013). En effet, un contre-exemple aux énoncés des règles (F1) et (F2) apparaît dès lors que a(s) > b(s). Comparable certes, mais pas identique épistémologiquement car le caractère continu demeure par la prise en compte des intervalles. Or, rappelons que dans le cadre des variables numériques a et b (Gras et Régnier, 2009, 2013), pour tout i∈ E, nous notons $\bar{b}_i$ le complément à 1 de bi : $\bar{b}_i$ = 1-bi, valeur de la variable b qui récuse b_i. On choisit comme pour l'implication entre variables binaires, l'indice $\sum_{i \in E} a_i \bar{b}_i$ - qui prend la valeur $n_{a \wedge \bar{b}}$ dans le cas binaire - comme indice de non-propension (ou de non-tendance) de a vers b. Ainsi, intuitivement et grossièrement, plus cet indice sera petit, plus on pourra s'attendre à une propension de a vers b (Gras et Régnier, 2009). Toutefois, cet indice ne tient plus dans le cas qui nous intéresse maintenant où les variables sont continues.

4.2 Formalisation de l'intensité d'implication entre variables continues

La formalisation va s'inspirer de celle retenue dans le cas où l'espace des sujets est continu (Gras et Régnier, 2013, p. 165-175). Considérons les sous-ensembles de sujets A et B de E admettant les valeurs respectives selon les variables a et b et satisfaisant les inégalités suivantes : $A=\{s\in E\ /a(s) \geq \alpha\}$ et $B=\{s\in E\ /b(s) \geq \beta\}$ où α et β sont deux valeurs quelconques appartenant à l'intervalle [0 ; 1]. Les cardinaux de A et B sont respectivement n_a et n_b (avec $n_a \leq n_b$). Par conséquent, les sujets s du sous-ensemble $A \cap \overline{B}$ vérifient $a(s) \geq \alpha$ et $b(s) < \beta$ ou encore $1-\beta \leq 1-b(s)<1$

Comme nous le faisons dans le cas binaire, nous comparons le nombre de contre-exemples à l'implication a => b (au sens de la relation F1 indiquée dans la section 3.1) observés dans la contingence et celui que l'on obtiendrait si les variables a et b étaient indépendantes. Pour cela, choisissons respectivement au hasard, de façon indépendante, deux parties $\hat{X}$ et $\hat{Y}$ de E de mêmes cardinaux respectifs que A et B et vérifiant les mêmes inégalités que ces deux sous-ensembles. Cependant la probabilité associée au tirage de $\hat{X}$ et $\hat{Y}$ parmi les parties satisfaisant les propriétés cardinales doit être affectée de celle qui traduit les contraintes pesant sur les valeurs prises selon a et b par les sujets des sous-ensembles $\hat{X}$ et $\hat{Y}$. Cette seconde probabilité est calculée à partir de la distribution produit entre les lois respectives de a et de b. Or, sous l'hypothèse d'indépendance *a priori* de a et b, la loi produit est égale au produit des lois respectives de a et b de densités f_a et f_b nulles à l'extérieur de [0 ; 1]. Les variables a et $\overline{b}=1-b$ sont également indépendantes et $f_{\overline{b}}=1-f_b$ est la densité de probabilité de $\overline{b}$. Les fonctions de répartition associées sont respectivement F_a, F_b et $F_{\overline{b}}$.

Ainsi la probabilité qu'un sujet pris au hasard ait un comportement vis-à-vis des modalités continues de a et de celles de b qui soit un contre-exemple à a => b (c'est à dire que $a(s) \geq \alpha$ tandis que $0 \leq b(s) < \beta$ ou $1-\beta \leq \overline{b}(s) <1$) est

$$\left[\int_{\alpha}^{1} f_a(t)dt\right]\left[\int_{0}^{\beta} f_b(u)du\right]=\left[\int_{\alpha}^{1} f_a(t)dt\right]\left[\int_{1-\beta}^{1} f_{\overline{b}}(u)du\right]=\left[1-F_a(\alpha)\right]\left[1-F_{\overline{b}}(1-\beta)\right]$$

A chaque sujet s de $\hat{X}\cap\hat{\overline{Y}}$, nous associons une variable de Bernoulli de paramètre p, qui est la probabilité qu'il soit à la fois dans l'ensemble des contre-exemples et qu'il vérifie la double inégalité $a(s)\geq \alpha$ et $b(s)<\beta$. Ces quatre événements étant indépendants par hypothèse, la probabilité p s'écrit comme produit de leurs probabilités, à savoir :

$$p=\frac{n_a n_{\overline{b}}}{n^2}\left[\int_{\alpha}^{1} f_a(t)dt\right]\left[\int_{0}^{\beta} f_b(u)du\right]$$

De là, Card($\hat{X}\cap\hat{\overline{Y}}$), somme des variables de Bernoulli indépendantes associées à chaque sujet de $\hat{X}\cap\hat{\overline{Y}}$ peut être considérée comme une variable binomiale de paramètres n et p.

$$\Pr\left[\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}})=k\right]=C_n^k p^k(1-p)^{n-k}$$

Espérance et variance de cette variable aléatoire sont donc :

E[Card ($\hat{X}\cap\hat{\overline{Y}}$] = np et Var [Card ($\hat{X}\cap\hat{\overline{Y}}$)]= np(1-p)

Par suite,

$$\Pr[\mathrm{Card}\,(\hat{X}\cap\hat{\overline{Y}}) \le \mathrm{Card}(A\cap\overline{B})] = \sum_{k=0}^{k=\mathrm{Card}(A\cap\overline{B})} C_n^k p^k (1-p)^{n-k}$$

Comme dans le cas de variables discrètes, nous définissons l'intensité d'implication de a sur b par :

$$\varphi\,(a,b) = 1-\Pr\left[\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}}) \le \mathrm{Card}(A\cap\overline{B})\right]$$

Remarque 1

Cette approche a-t-elle un sens dans sa restriction au cas de variable binaires ? La réponse est oui. En effet si les variables a et b sont binaires, la relation d'implication a => b n'est pas satisfaite lorsque pour $s\in \hat{X}\cap\hat{\overline{Y}}$, a(s) = 1 et b(s) = 0 ou $\overline{b}(s)=1$. Les densités de probabilité sont dégénérées. Pour que l'on ait Pr[{$s\in X$ | a(s)=1]=1, il est suffisant que α= 0. De même, pour que ait Pr[{$s\in \overline{Y}$| b(s)=0]=1, il est également suffisant que β=1. Les deux intégrales définissant p sont alors égales à 1 et p devient égale à $\frac{n_a n_{\overline{b}}}{n^2}$, expression effectivement conforme à ce que nous avons dans la théorie de l'ASI dans le cas binaire.

Remarque 2

Comme nous le montrons dans (Gras et al, 2009, 2013), nous pouvons modéliser le cardinal aléatoire de $X\cap\overline{Y}$ par une variable de Poisson de paramètre

$$\lambda = \frac{n_a n_{\overline{b}}}{n}\left[\int_\alpha^1 f_a(t)dt\right]\left[\int_0^\beta f_b(u)du\right]$$

Par suite $\forall s\in\{0,1,2,\ldots,n\}$ Pr [Card($\hat{X}\cap\hat{\overline{Y}}$) = s]= $\frac{\lambda^s}{s!}e^{-\lambda}$

et $\varphi\,(a,b) = 1-\Pr\left[\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}}) \le \mathrm{Card}(A\cap\overline{B})\right]$=1- $\sum_{s=0}^{s=\mathrm{Card}(A\cap\overline{B})}\frac{\lambda^s}{s!}e^{-\lambda}$

Pour $\lambda\ge 5$, la variable « indice d'implication empirique », notée :

$$Q(a,\overline{b}) = \frac{\mathrm{Card}(X\cap\overline{Y}) - E\left[\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}})\right]}{\mathrm{Var}\left[\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}})\right]} = \frac{\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}})-\lambda}{\sqrt{\lambda}}$$

qui résulte du centrage-réduction de la variable de Poisson, $\mathrm{Card}(\hat{X}\cap\hat{\overline{Y}})$, peut être approchée par la variable gaussienne centrée réduite N(0 ;1).

Si nous considérons la valeur empirique $q(a,\overline{b}) = \frac{n_{a\wedge\overline{b}}-\lambda}{\sqrt{\lambda}}$, alors l'intensité d'implication estimée de la quasi-règle $a\Rightarrow b$, est approximativement :

$$\varphi(a,b) = 1 - \Pr\left[Q(a,\bar{b}) \le q(a,\bar{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{q(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

Insistons sur le sens de cette intégrale. Elle représente la probabilité gaussienne pour que le nombre de transactions observées satisfaisant la quasi-règle $a \Rightarrow b$, soit supérieur à celui qui serait observable sous l'hypothèse d'indépendance de a et b. Autrement dit, $\Pr[Q(a,\bar{b}) \le q(a,\bar{b})]$ est la p-value du test visant à réfuter l'hypothèse de l'indépendance de a et b au profit d'une relation de type quasi-implication.

5 Cas général traité par la méthode ASI propensive

5.1 Formalisation de la relation de propension entre variables modales

J.B. Lagrange (1998) a construit dans le cas des variables modales, un indice de **propension** entre variables modales qui généralise l'indice d'implication entre variables binaires. En posant les conditions suivantes :

- si $a(x)$ et $\bar{b}(x)$ sont les valeurs prises en x par les variables modales a et $\bar{b}$, où $\bar{b}(x) = 1 - b(x)$

- si s_a^2 et $s_{\bar{b}}^2$ sont les variances empiriques des variables a et $\bar{b}$

Définition 1: L'indice de propension de variables modales est :

$$\tilde{q}(a,\bar{b}) = \frac{\sum_{x \in E} a(x)\bar{b}(x) - \frac{n_a n_{\bar{b}}}{n}}{\sqrt{\frac{(n^2 s_a^2 + n_a^2)(n^2 s_{\bar{b}}^2 + n_{\bar{b}}^2)}{n^3}}}$$

Cette solution apportée au cas modal est aussi applicable au cas des *variables fréquentielles*, voire *des variables numériques positives*, à condition d'avoir normalisé les valeurs observées sur les variables, telles que a et b, la normalisation dans [0 ; 1] étant faite à partir du maximum de la valeur prise respectivement par a et b sur l'ensemble E.

Tout en suivant une modélisation comparable à celle de la section 2.1 où la loi de chaque variable est celle de la variable continue uniforme sur [0 ; 1], nous envisageons le cas où les variable sont continues, mais de lois données différentes et à valeurs ramenées par normalisation sur l'intervalle [0 ; 1]. Par ailleurs, notre recherche porte, comme dans la section 3, sur la relation F 1 :

Dans quelle mesure peut-on considérer que, pour $x \in E$ « ***si a(x) ≥ α, alors on peut affirmer que b(x) ≥ β*** » ?

Reprenons les notations de la section 2 : X (resp. Y, resp. $\bar{Y}$) est une variable aléatoire de [0 ; 1], ensemble normalisé des valeurs a_i (resp. b_j et 1- b_j) de la variable a

(resp. b, resp. $\overline{b}$) mais de loi de densité de probabilité f_a (resp. f_b et f_{1-b}) quelconque. De cette manière nous avons Prob $[X \geq \alpha] = \left[\int_{\alpha}^{1} f_a(t)dt\right]$ et Prob$[Y < \beta] = \left[\int_{0}^{\beta} f_b(u)du\right]$.

Les variables aléatoires X et Y (donc également $\overline{Y}$) sont, par hypothèse commune en ASI, indépendantes. Leurs paramètres, espérance et variance, sont calculables et donnés par les formules habituelles.

Introduisons la variable aléatoire $Z = X\overline{Y} = X.(1-Y)$. Elle représente la conjonction d'une réalisation des variables a et 1-b. La loi de Z est le produit des lois de X et de 1-Y et sa densité de probabilité est le produit des celles des lois de probabilité de X et 1-Y. L'espérance de Z est alors connue par : $E(Z) = E(X)E(1-Y) = M_1$ et sa variance est : $V(Z) = E(Z^2) - [E(Z)]^2 = M_2 - M_1^2$

Nous considérons le n-uplet $(Z_1, Z_2, \ldots, Z_n)$ composé de n variables aléatoires indépendantes de même loi que Z et dont la réalisation est celle des n sujets de E. Nous définissons alors comme dans la section 2, la variable somme : $T_n = \frac{1}{n}\sum_{k=1}^{k=n} Z_k$. Il en résulte que l'espérance $E(T_n) = \frac{1}{n}\sum_{k=1}^{k=n} E(Z_k) = M_1$ ainsi que la variance $V(T_n) = \frac{1}{n}\sum_{k=1}^{k=n} V(Z_k) = \frac{1}{n}(M_2 - M_1^2)$ puisque les variables Z_k sont mutuellement indépendantes.

En utilisant le théorème de la limite centrale dit de Lindeberg-Lévy puisque l'indice de propension est une moyenne de variables aléatoires indépendantes identiquement distribuées, on démontre que T_n suit approximativement, pour n grand, la loi de Laplace-Gauss d'espérance $E(Z) = M_1$ et de variance $\frac{1}{n}(M_2 - M_1^2)$

Autrement dit, la loi de la variable « indice de propension empirique »

$$\tilde{Q}(a, \overline{b}) = \frac{T_n - M_1}{\sqrt{\frac{1}{n}(M_2 - M_1^2)}}$$

est approximativement la loi gaussienne centrée réduite N(0 ; 1).

Dans une mise en pratique, la calcul d'une réalisation de la variable « indice de propension empirique », nécessite d'estimer l'espérance et la variance de la variable T_n. Pour ce faire, nous utilisons m_a la moyenne des observations a_i et m_b la moyenne des observations b_i ; qui nous permet d'obtenir une estimation de la moyenne M_1 avec $m_a(1-m_b)$. Ensuite nous réalisons une estimation de la variance de T_n à partir de la variance v_a des a_i et v_b celles des b_i, une estimation du moment M_2 d'ordre 2 est obtenue à partir des observations a_i et b_i par $(v_a + m_a^2)(v_b + m_{\overline{b}}^2)$ puisque les observations b_i et $\overline{b_i}$ ont la même variance.

L'indice empirique d'implication devient :

$$\tilde{q}(a,\bar{b}) = \frac{\frac{1}{n}\sum_{i\in E} a_i \bar{b}_i - m_a m_{\bar{b}}}{\sqrt{\frac{v_a(v_b + m_{\bar{b}}^2) + m_a^2 v_b}{n}}}$$

Quant à l'estimation de l'intensité de propension, elle est encore obtenue par :

$$\varphi(a,b) = 1 - \Pr\left[\tilde{Q}(a,\bar{b}) \le \tilde{q}(a,\bar{b})\right] = \frac{1}{\sqrt{2\pi}} \int_{\tilde{q}(a,\bar{b})}^{\infty} e^{-\frac{t^2}{2}} dt$$

Nous avons démontré (Gras, Régnier, 2013) que la réduction de ces formules exprimées dans le cas continu à l'instar des variables numériques au cas binaire était parfaitement valide.

6 Conclusion

Nous avons examiné différentes situations relatives aux variables principales continues dans le cadre de l'Analyse Statistique implicative de l'A.S.I. Partant de la situation la plus élémentaire où ces variables sont uniformément distribuées, nous avons considéré les variables continues à distribution quelconque sur [0 ; 1]. Après un retour simplificateur par les variables continues par intervalle, nous avons établi des lois permettant de calculer le critère fondamental en A.S.I., l'intensité d'implication qui évalue la qualité implicative d'une variable sur une autre, leurs distributions pouvant être différentes. Nous pouvons retenir la méthodologie adaptée pour traiter le cas continu dans d'autres types d'analyse de données. De plus, nous retenons la capacité dont nous bénéficierons dans des applications où il s'agit de nuancer spécifiquement la connaissance des variables à prendre en compte dans l'analyse, comme par exemple nuancer l'information sur la complexité *a priori* de ces variables pour une population de sujets donnée.

Références

[1] Bailleul, M., (1994). *Analyse statistique implicative : variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique*. Thèse Université de Rennes 1

[2] Diday E, (1972), *Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes*, Thèse d'Etat, Université de Paris VI.

[3] Gras R., Régnier J.C. et Guillet F. (2009). *L'Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*, RNTI-E-16, Cépaduès Editions

[4] Gras R., Diday E., Kuntz P et Couturier R. (2001), Variables sur intervalles et variables-intervalles en analyse statistique implicative, *Actes du 8ème Congrès de la Société Francophone de Classification, Université des Antilles-Guyane, 17-21 décembre 2000*,166-173

[5] Gras R. et Régnier J.-C. (2013). Extension de l'A.S.I. aux variables non binaires in *L'Analyse Statistique Implicative, Méthode exploratoire et confirmatoire à la recherche de causalités*, Cépaduès Editions, 70-77

[6] Lagrange J.-B., (1998), Analyse implicative d'un ensemble de variables numériques ; application au traitement d'un questionnaire à modalités modales ordonnées, *Revue de Statistique Appliquée*, XLVI, 71-93.

[7] Régnier, J.-C., & Gras, R. (2005) Statistique de rangs et Analyse Statistique Implicative. *Revue de Statistique Appliquée*. 53(1) p.5-38

Ouvrages de référence

[1] *L'implication statistique. Nouvelle méthode exploratoire de donnée*, sous la direction de R.Gras, et la collaboration de S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn, A.Totohasina, La Pensée Sauvage, Grenoble (1996)

[2] *Mesures de Qualité pour la Fouille de Données*, H.Briand, M.Sebag, R.Gras er F.Guillet eds, RNTI-E-1, Cépaduès, 2004

[3] *Quality Measures in Data Mining* , F.Guillet et H.Hamilton eds, Springer, 2007,

[4] *Statistical Implicative Analysis, Theory and Applications*, R.Gras, E. Suzuki, F. Guillet, F. Spagnolo, eds, Springer, 2008.

[5] *Analyse Statistique implicative. Une méthode d'analyse de données pour la recherche de causalités*, sous la direction de Régis Gras, réd. invités R. Gras, J.C. Régnier, F. Guillet, Cépaduès Ed. Toulouse, 2009.

[6] *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, Eds : P.Orus, L.Zamora, P.Gregori, Universitat Jaume-1, Castellon (Espagne), ISBN : 978-84-692-3925-4, 2009..

[7] *L'Analyse Statistique Implicative : de l'exploratoire au confirmatoire*. Eds : J.C. Régnier, Marc Bailleul, Régis Gras, Université de Caen, ISBN : 978-2-7466-5256-9, 2012

[8] *L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités*, sous la direction de Gras R., eds Gras R., Régnier J.-C., Marinica C., Guillet F., Cépaduès Editions, 522 pages, ISBN 978.2.36493.056.8, 2013.

CLASSIFYING OBJECTIVE INTERESTINGNESS MEASURES BASED ON THE TENDENCY OF VALUE VARIATION

Nghia QUOC PHAN[1], Hiep XUAN HUYNH[2], Fabrice GUILLET[3], Régis GRAS[4]

CLASSIFICATION DES MESURES D'INTÉRÊT OBJECTIVES BASÉE SUR LA TENDANCE DE VARIATION DES VALEUR

RÉSUMÉ

Dans les dernières années, la recherche de découverte de connaissances à partir des données a intéressé à nombreux chercheurs et de nombreux résultats de recherche ont été utilisés efficacement dans de plusieurs domaines de la vie. La mesure d'intérêt joue un rôle important dans le domaine de recherche de découverte des connaissances. C'est pourquoi l'étude sur les mesures d'intérêt qui ont déjà développés est de plus en plus importante. La plupart des études se sont basées sur deux méthodes principales: classification basée sur les propriétés d'une mesure et de classification basée sur les comportements d'une mesure. Dans notre étude, nous allons étudier et examiner la tendance de variation de la valeur des mesures d'intérêt objectives qui répondent à la nature asymétrique en prenant la dérivée partielle de la fonction qui calcule la valeur des mesures d'intérêt selon le tableau de contingence 2x2. Nos résultats montrent que les mesures d'intérêt objectives asymétriques sont classées par la considération de la variation (l'augmentation, la réduction ou la stabilité) à partir des formules dérivées de chaque mesure.

***Mots-clés**: Mesures d'intérêt objectives, classification, dérivées partielles, tendance des valeurs de variation, intensité d'implication.*

ABSTRACT

In recent years, research in Knowledge Discovery in Databases has been a topic of increasing interest for many researchers. Currently, a variety of such research has effectively been used in many areas of life. Interestingness measures play an important role in the field of research in knowledge discovery. Therefore, the study of the classification of interestingness measures is also an important topic for researchers. Classification of measures is mostly based on two main methods: classification based on the properties of measures and classification based on the behavior of measures. In this study, we propose a new classification method focusing on the research and study of the value variation of the objective interestingness measures that satisfy the asymmetric nature, by taking the partial derivative of the function that calculates the value of interestingness measures according to the 2×2 contingency table. Our results show that asymmetrical objective interestingness measures are classified by considering the increasing, decreasing or stable derivative formula for each measures.

***Keywords**: objective interestingness measures, classification, partial derivative, tendency of value variation, implication intensity.*

[1] TraVinh University, 126 National Road 53, Tra Vinh City, Vietnam, nghiatvnt@gmail.com
[2] Can Tho University, 3/2 Street Ninh, Kieu District, Can Tho City, Vietnam, hxhiep@ctu.edu.vn
[3]University of Nantes, Polytechnic School of Nantes University, France, fabrice.guillet@univ-nantes.fr
[4]University of Nantes, Polytechnic School of Nantes University, France, regisgra@club-internet.fr

1 Introduction

Interestingness measures (Piatetsky-Shapiro and Matheus, 1994; Guillet and Hamilton, 2007), both subjective and objective measures (Silberschatz and Tuzhilin 1995), play an important role in term of evaluating the quality of knowledge according to association rules. In the past ten years, the study of interestingness measures (Jalali-Heravi and Zaïane, 2010; Jon Hills *et al.*, 2012; David Glass, 2013; Martínez-Ballesteros *et al.* 2014) has developed significantly. Researchers in this field only focus on two main directions: (1) proposing new measures (Liu, 2008; David Glass, 2013; Selvarangam and Ramesh Kumar, 2014); and (2) studying the properties, behaviors and trends of the variation of the measures to rank, cluster and classify these measures (Tan *et al.*, 2002; Lenca *et al.*, 2004; Blanchard *et al.*, 2009; Huynh *et al.*, 2005, 2006, 2007), which assists users to select appropriate measures for their specific application. Since the number of interestingness measures have been increased, the concern of researchers have been increased, especially for the classification of these measures as:

- Classification based on objective and subjective criteria (Silberschatz and Tuzhilin, 1995, 1996);
- Evaluation of a good measures based on three principles (Piatetsky-Shapiro, 1991) based on the key properties in order to evaluate the measures to conclude that each measures only satisfies a number of properties and possibly appiesl in some specific areas (Tan *et al.*, 2002, 2004);
- Classification of measures based on the development a multi-criteria system (Lenca *et al.*, 2004, 2008);
- Classification of measures based on three criteria: the subject, the scope and the nature of the measures (Blanchard *et al.*, 2009);
- Classification of 62 measures based on a definition of 12 properties and an assigniment values for first nine properties that supported users for selecting appropriate measures in their specific application (Maddouri and Gammoudi, 2007);
- Classification for 40 objective interestingness measures by examining the value of 6 properties: independence, balance, symmetry, variation, description and statistic (Huynh *et al.*, 2007, 2011);
- Sixty two measures are classified into seven classes by basing on 19 properties and applying classification techniques (Guillaume *et al.*, 2012);
- Classification of 35 measures by the examining of behavior (Huynh *et al.*, 2005, 2006, 2007);
- Classification of 61 measures by studying of over 110 data sets (Tew *et al.*, 2013);
- Examining of variable values of the individual measures by taking the partial derivative of the calculated function of measures (Gras and Kuntz, 2008).

The determination of the variability of the interestingness measures is one of the most important criteria in assessing the objective interestingness measures (Huynh *et al.*, 2011). Based on examining the variation of the value by taking the partial derivative of the calculated function of Implication index (Gras and Kuntz, 2008), a new method is proposed for the classification of objective interestingness measures. The proposed method has ability to determine the increasing, the decreasing or the independence of the interestingness measures for each input parameter. Accordingly, the measures are

classified into different classes, i.e., increasing measures, decreasing measures, and independent measures, for each input parameter of the measures calculated function. Results of this classification will help users select the appropriate measures for their application when they know the input parameters of the actual data set.

This paper is organized into five sections. Section 1 introduces the general objective interestingness measures and measures classification methods. Section 2 reviews three methods for measures classification: classification based on a study of the properties of measures; classification based on a examining of the behavior of measures; and classification based on an examining of the variable values of individual measures. Section 3 describes the details of the method for the behavior evaluation of the measures according to the tendency of value variation. Section 4 conducts a classification of the interestingness measures which satisfy the asymmetric with additional reviews and comments based on partial derivatives formula. The final section summarizes the important results achieved in this paper.

2 The classification methods of interestingness measures

Recently, the classification usually focuses on three main approaches, i.e., classification based on examining of the properties of the measures, classification based on examining of the behavior of the measures, and classification based on a examining of the variable values of individual measures.

2.1 Classification based on examining of measures properties

The classification method based on theoretical attributes is first used popularly to classify the measures. In the general level of classification, interestingness measures are divided into two separate classes, such as subjective interestingness measures and objective interestingness measures. The subjective interestingness measures evaluate knowledge models based on the target, the knowledge and the belief of users. The objective measures valuate knowledge models based on the structure of the statistical data (Silberschatz and Tuzhilin, 1996). In the detailed level of classification, there are three key principles for a good measures: (1) the law $a \rightarrow b$ is equal to 0 if a and b independent; (2) the monotonically increasing with $a \cap b$; (3) the monotonically decreasing with $a\ or\ b$ is set to a criterion to evaluate the interestingness measures (Piatetsky-Shapiro, 1991). In the experimental level of classification, several researchers have been proposed the number of measures by using different criteria and classes to categorize and to evaluate measures. For example, using nine criteria for evaluation, i.e, conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility, and actionability, are examined in order to classify 38 objective measures (Geng and Hamilton, 2006). Building a system of multi-criteria decision to classify 20 measures based on the theoretical properties and the experimental results on 10 datasets (Lenca *et al.*, 2004, 2008). Studying theoretical properties, such as independence, balance, symmetry, variability, description and statistic of the measures is used to divide 40 objective interestingness measures into five classes: balance, independence, description, statistic and others (Huynh *et al.*, 2007, 2011). Expanding with 19 properties is used to divide 61 measures into 7 separate classes (Guillaume *et al.*, 2012).

2.2 Classification based on the behavior of interestingness measures

The classification method based on a study of the behaviors of the measures is based on the experimental investigation of the interestingness measures on specific data sets. From the experimental results, the behaviors of the measures can be predicted for each specific application area for the classification of the measures. By doing so, users can select the appropriate interestingness measures for their specific application. Method of assessing the interestingness measures based on behavioral examination was first proposed to investigate the behavior of 35 measures based on calculating of the distance between interestingness measures by using two clustering methods, Agglomerative Hierarchical Clustering (AHC) and Partitioning Around Medoids (PAM). As a result, measures are classified into 16 separate clusters to help users choose measures that can discover the best knowledge (Huynh *et al.*, 2005). The behavior of the measures was studied by correlation graph method. The approach was conducted on two prototypical datasets with opposite characteristics: a strongly correlated one (mushroom dataset) and a lowly correlated one (synthetic dataset). The results showed that the correlation between the interestingness measures depended on the type of data and the ranking rules to partition 40 measures into 6 stable clusters (Huynh *et al.*, 2006). In their experimental results, the study examined the behavior of 61 objective interestingness measures are conducted on 110 different datasets. The findings of the study classified 61 measures into 21 clusters and confirmed that the selection of the measures depends on the dataset. However, this research has not given the criteria of how to select a good dataset for inclusion in their experiments (Tew *et al.*, 2013).

2.3 Classification based on value variation of each individual measures

The examining of value variation of each individual interestingness measures allows us to know the increasing or the decreasing tendency of them according to the parameters of the calculated function (Gras and Kuntz, 2008). This method is carried out by taking the partial derivative of the calculated function of measures for each parameter. From the obtained result of taking the partial derivative, the behavior of each measures can be considered. Specifically, based on the partial derivative formula of each measures, the relationship of the measures with each parameter and the variable correlation between parameters of the calculated function of measures both can be considered.

3 Evaluation of the behavior of objective interestingness measures according to variable value trends

3.1 Objective interestingness measures

Let an association rule $a \rightarrow b$ where a and b are two disjoint sets of items (called item set). Item set a (resp. b) is associated with a subset of transactions $A = T(a)$ (resp. $B = T(b)$) with $T(a) = \{t \in T, a \subseteq t\}$, and T is the set of all transactions. The rule can be described by four cardinalities $(n, n_A, n_B, n_{A\bar{B}})$ where: $n = |T|, n_A = |A|, n_B = |B|, n_{A\bar{B}} = |A \cap \bar{B}|$. The cardinal $n_{A\bar{B}}$ corresponds to the effective number of negative examples of the rule (Figure 1).

The interestingness value of an association rule based on an objective interestingness measures will then be calculated by using the cardinality of a rule $m(a \to b) = f(n, n_A, n_B, n_{A\bar{B}}) \in \mathrm{R}$. To calculate easily, the following equivalent transformations should be used: $\mathrm{n_{AB}} = \mathrm{n_A} - \mathrm{n_{A\bar{B}}}$, $\mathrm{n_{\bar{A}}} = \mathrm{n} - \mathrm{n_A}$, $\mathrm{n_{\bar{B}}} = \mathrm{n} - \mathrm{n_B}$, $\mathrm{n_{\bar{A}B}} = \mathrm{n_B} - \mathrm{n_A} + \mathrm{n_{A\bar{B}}}$, $\mathrm{n_{\bar{A}\bar{B}}} = \mathrm{n} - \mathrm{n_B} - \mathrm{n_{A\bar{B}}}$.

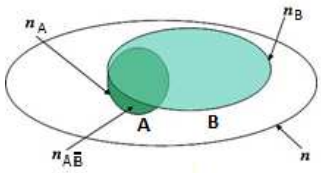


Figure 1.The cardinality of an association rule a ⟶ b

For example: two given sets A and B where A has 1 element and B has 3 elements. An association rule is in the form a ⟶ b. A = {Bread}, B = {Milk, Diappers, Beer} where $n = 100$, $n_A = 50$, $n_B = 80$ and $n_{A\bar{B}} = 10$.

The objective interestingness measures, Support Expectation, is identified by the formula:

$$m(a \to b) = f(n, n_A, n_B, n_{A\bar{B}}) = \frac{n_A(n_B - n_A + n_{A\bar{B}})}{n(n - n_A)}$$

Therefore, the interestingness value is:

$$m(a \to b) = \frac{50(80 - 50 + 10)}{100(100 - 50)} = 0.4$$

3.2 Statistical implicative analysis

The theory of statistical implicative analysis was developed to evaluate the learning behavior of pupils in the process of teaching algebra and geometry (Gras, 1979). This theory affects many areas such as pedagogy, psychology, and computer science. It has developed a unifying methodology and created synergy of scientific disciplines such as mathematics, statistics, psychology, education and data mining. It has also created a framework allowing for the evaluation of the strength of implications, formed through the acquisition of technical knowledge by natural or artificial processes. Especially in data mining, the extraction of knowledge from a large set of association rules is not conducted by humans to serve decision-making processes. Therefore, the development of interestingness measures plays an important role in the evaluation of association rules, and this is a clear success of statistical implicative analysis theory.

3.3 Examining the behavior of measures according to variable value trends

Stability analysis of implication index is to observe small variations of this measures in the surrounding space of parameters n, n_A, n_B and $n_{A\overline{B}}$ (Gras and Kuntz, 2008). To do this, taking the partial derivative for each parameter of the implication index formula is carried out. This method shows some abilities in supporting the research of the interestingness measures, i.e., the examining of increasing or decreasing variability of the measures, the relationship between the dependent variable parameters, n, n_A, n_B and $n_{A\overline{B}}$. However, the method also showed a limited role of parameters, n, n_A, n_B and $n_{A\overline{B}}$ to denote the increasing or decreasing variability of the measures.

Example 1. Let us examine the Added value measures (Sahar, 2003) with the following formula: $f = \frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_B}{n}$

We have the partial derivative formula for each parameters, $n, n_A, n_B, n_{A\bar{B}}$, of the function f as follows:

$$\frac{\partial f}{\partial n} = \frac{n_B}{n^2}; \quad \frac{\partial f}{\partial n_A} = \frac{n_{A\bar{B}}}{n_A^2}; \quad \frac{\partial f}{\partial n_B} = -\frac{1}{n}; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = -\frac{1}{n_A}$$

where n = 100, $n_A = 20$, $n_B = 40$, $n_{A\bar{B}} = 4$, we calculate the relative values as follows:

$$f = 0.4; \quad \frac{\partial f}{\partial n} = 0.004; \quad \frac{\partial f}{\partial n_A} = 0.01; \quad \frac{\partial f}{\partial n_B} = -0.01; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = -0.05$$

Thus, the variation of the Added value measures depends on two parameters, n and n_A and independent of two parameters, n_B and $n_{A\bar{B}}$. The interestingness value will increase as the number of n and n_A increases, whereas the interestingness value decreases when the number of parameters, n_B and $n_{A\bar{B}}$, increases. However, the declining rate of interestingness value does not depend on the increased rate of these parameters.

Example 2. Let us examine the Conviction measures (Brin et al., 1997) with the following formula: $f = \frac{n_A(n - n_B)}{n n_{A\bar{B}}}$

We have the partial derivative formula for each parameters, $n, n_A, n_B, n_{A\bar{B}}$ of the function f as follows:

$$\frac{\partial f}{\partial n} = \frac{n_A n_B}{n_{A\bar{B}} n^2}; \quad \frac{\partial f}{\partial n_A} = \frac{(n - n_B)}{n n_{A\bar{B}}}; \quad \frac{\partial f}{\partial n_B} = -\frac{n_A}{n n_{A\bar{B}}}; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = -\frac{n_A(n - n_B)}{n n_{A\bar{B}}^2}$$

where n = 100, $n_A = 20$, $n_B = 40$, $n_{A\bar{B}} = 4$, we calculate the relative values as follows:

$$f = 3; \quad \frac{\partial f}{\partial n} = 0.02; \quad \frac{\partial f}{\partial n_A} = 0.15; \quad \frac{\partial f}{\partial n_B} = -0.05; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = -0.75$$

Different from Added value measures, the variability of the Conviction measures depends on two parameters, n and $n_{A\bar{B}}$, and independent of two parameters, n_A $and n_B$. However, interestingness value will increase when the value of n, n_A rises and the increased rate of the value does not depend on the increased rate of the parameter n_A. Conversely, the interestingness value will decrease when the value of parameters, n_B and $n_{A\bar{B}}$ increases and the decreased rate of the interestingness value does not depend on the increase rate of the parameter n_B.

Example 3. Let us examine the Coverage measures (Geng and Hamilton, 2006) with the following formula: $f = \frac{n_A}{n}$

We have the partial derivative formula for each of the four parameters, $n, n_A, n_B, n_{A\bar{B}}$ of the function f as follows:

$$\frac{\partial f}{\partial n} = -\frac{n_A}{n^2}; \quad \frac{\partial f}{\partial n_A} = \frac{1}{n}; \quad \frac{\partial f}{\partial n_B} = 0; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = 0$$

where n = 100, $n_A = 20$, $n_B = 40$, $n_{A\bar{B}} = 4$, we calculate the relative values as follows:

$$f = 0.2; \quad \frac{\partial f}{\partial n} = -0.002; \quad \frac{\partial f}{\partial n_A} = 0.01; \quad \frac{\partial f}{\partial n_B} = 0; \quad \frac{\partial f}{\partial n_{A\bar{B}}} = 0$$

Unlike these two above mentioned measures, i.e., Conviction measures and Added measures, the variation of the Coverage measures depends only on the parameter n and has a tendency of decreasing when the parameter n increases. In contrast, its variation has tendency of increasing and is independent of the parameter n_A. Two parameters, $n_B\ and n_{A\overline{B}}$, are not constributed completely on the demonstration of the variation in this measures. This is a good example to illustrate the limitation of the parameters in indicating the variation of the measures as fore-mentioned.

4 Classification based on examining the tendency of value variation

Based on a synthesic list of objective interestingness measures, which is applied to evaluate the quality of the knowledge as the form of the association rule $a \rightarrow b$ (Agrawal and Srikant, 1994), we have found that the group of measures satisfied the symmetry property (m($a \rightarrow b$) = m($b \rightarrow a$)) and the group of measures satisfied the asymmetry property (m($a \rightarrow b$) ≠ m($b \rightarrow a$)), they offered an almost equal proportion of quantities. According to the five criteria are used to evaluate objective interestingness measures based on the distribution of the probability with the size of 2×2, a good measures must satisfy the symmetry property of the permutation of two variables (Tan *et al.*, 2002). It means that two association rules, $a \rightarrow b$ and $b \rightarrow a$, must have the same interestingness value. However, this is not true for many applications. The Confidence measures is an example. It is evaluated as a good measures by many researchers and is applied in many practical applications but it is an asymmetric measures. Therefore, in this paper, we focus on the study of the group of objective interestingness measures satisfying the asymmetric property. We calculate the partial derivative of the measures function according to 4 parameters $\boldsymbol{f}(n, n_A, n_B, n_{A\overline{B}})$ of this group with the purpose of analyzing the value variability of each measures according to the distribution of the probability with the size of 2×2. Accordingly, the group of asymmetric measures is divided into four different classes for each parameter. This classification helps researchers and users see the relationship between the value of each parameter and the value variation of the asymmetric measures from which they can select suitable measures for their study and application. The formula for calculating the asymmetric measures and the partial derivatives of measures based on 4 parameters are presented in the appendices.

In the next step, an examination of the derivative formula of each measures is conducted to classify them into groups with tendency of to increasing, decreasing or to determine that they are independent of parameters n, n_A, n_B, $n_{A\overline{B}}$.

Table 1: Findings of the variation of the measures based on partial derivatives under 4 parameters (1: increasing Variability; -1: decreasing Variability; 0: independent).

No	Name of interestingness measures	n	n_A	n_B	$n_{A\overline{B}}$
1	1-way Support	1	1	-1	-1
2	Added value, Pavillon, Centred Confidence, Dependency	1	1	-1	-1
3	Bayes factor, Odd multiplier	1	1	-1	-1
4	Causal-Confidence	1	1	-1	-1
5	Causal-Confirmed confidence	1	1	-1	-1
6	Loevinger, Certainty Factor, Satisfaction	1	1	-1	-1
7	Relative Risk, Class correlation ratio	1	1, 0, -1	-1	-1
8	Collective strength	1	1, 0, -1	-1	-1
9	Confidence	0	1	0	-1
10	Causal Confirm	1	1	-1	-1
11	Conviction	1	1	-1	-1
12	Coverage	-1	1	0	0
13	Descriptive Confirmed-Confidence, Ganascia Index	0	1	0	-1
14	Descriptive-Confirm	-1	1	0	-1
15	Entropic Implication Intensity 1	1	1	-1	-1
16	Entropic Implication Intensity 2	1	1	-1	-1
17	Examples and counter-examples rate	0	1	0	-1
18	Gain, Fukuda	-1	1	0	-1
19	Gini index	1	1	-1	-1, 0, 1
20	Goodman–Kruskal	1	1	-1	-1, 0, 1
21	Implication index	-1	-1	1	1
22	Implication Intensity (II)	1	1	-1	-1
23	Probabilistic measures of deviation from equilibrium (IPEE), Indice Probabiliste d'Ecart d'Equilibre	0	-1	0	0
24	Directed Information ratio (DIR)	0, 1, -1	1, 0, -1	0, 1, -1	-1, 0, -1
25	MGK, Ion	-1,1	1	-1,1	-1
26	J-measures	1	1	-1	-1, 0, 1
27	Klosgen	1	1	-1	-1
28	K-measures	1	1	-1	-1
29	Kulczynski index	0	1	-1	-1
30	Laplace	0	1	0	-1
31	Least contradiction	0	1	-1	-1
32	Leverage, Leverage 1	1	-1	-1	-1
33	Mutual Information MI, 2-way Support Variation	1	1	-1	-1, 0, 1
34	Prevalence	-1	0	1	0
35	Putative Causal Dependency	-1	1	-1	-1
36	Recall, Completeness	0	1	-1	-1
37	Sebag and Schoenauer	0	1	0	-1
38	Specificity 1, Negative Reliability	1	1	-1	-1
39	Zhang Zhang	1	1	-1	-1

Table 1 shows that most of the measures are variable to the tendency of increasing or decreasing according to four parameters n, n_A, n_B, $n_{A\overline{B}}$. However, the majority of the measures are of increasing variability with the parameter n_A. This reflects the general rule of the objective interestingness measures: the more examples do appear, the more reliability do increase. On the other hand, the decreasingly variable measures with the parameter $n_{A\overline{B}}$ are also prominent in Table 1. This also follows the general rule of interestingness measures. If the number of counter-examples increases, then the value of the interestingness measures decreases.

In the list of the analyzed measures, we have found that some measures having the variation is quite interesting. They do not follow the general variation rule of the measures is always increasing, decreasing and independent under parameters n,n_A,n_B,$n_{A\overline{B}}$. For example, Relative risk and Collective strength are two measures whose variation depends on the value of the parameter n_A. When examining these two measures, we see that the values of their partial derivative with respect to n_A is as follows: initial positive value, then proceed to 0, finally reaching negative values. This means that when the parameter n_A increases, the variability of those measures increases, decreases, and reaches extreme values. When we examine the Gini index, Goodman-Kruskal, J-measures, and Mutual information under the parameter $n_{A\overline{B}}$, we also see cases similar to Relative risk and Collective strength variation depend on the value of parameter n_A. In particular, Directed Information Ratio (DIR) and MGK measures have variability that depend on specific condition of $\frac{n_A - n_{A\overline{B}}}{n_A}$, $\frac{n - n_{A\overline{B}}}{n_A}$and $\frac{n_B}{n}$.

Table 2: Classification of measures based on partial derivative under the parameter *n*

Decreases with n	Independent with n	Increases with n	Others
Coverage Descriptive-Confirm Gain, Fukuda Implication index Prevalence Putative Causal Dependency	Confidence Descriptive Confirmed-Confidence , Ganascia Index Examples and counter-examples rate Probabilistic measures of deviation from equilibrium (IPEE), Indice Probabiliste d'Ecart d'Equilibre Kulczynski index Laplace Least contradiction Recall, Completeness Sebag and Schoenauer	1-way Support Added value, Pavillon, Centred Confidence, Dependency Bayes factor, Odd multiplier Causal-Confidence Causal-Confirmed confidence Loevinger, Certainty Factor, Satisfaction Relative Risk , Class correlation ratio Collective strength Causal Confirm Conviction Entropic Implication Intensity 1 Entropic Implication Intensity 2 Implication Intensity Gini index Goodman–Kruskal J-measures Klosgen K-measures Leverage, Leverage 1	Directed Information ratio (DIR) MGK, Ion

		Mutual Information MI, 2-way Support Variation Specificity 1, Negative Reliability Zhang Zhang	

Table 2 shows that the number of measures are variable increasingly on the parameter n that gets over 50 percents of the studied measures. This reflects that the interestingnes value of measures depend on the size of the data used to examine them. It means that their rate of the variation depends on the rate of change of the parameter n. In contrast, the number of measures are variable decreasingly on the parameter n quite small. these measures are Coverage, Descriptive-Confirm, Gain, Fukuda, Implication index, Prevalence, and Putative Causal Dependency. Group of the independent measures on the parameter n is an especial class because most of them agrees with the descriptive property. If the measures satisfies the descriptive property, its interestingness value will not depend on the value of the parameter n (the size of the data). This result shows that the method of the study of the variation trend of measures based on partial derivative achieved results as accurate as other methods. DIR measures and MGK measures have the variation depending on the specific values of the parameters n.

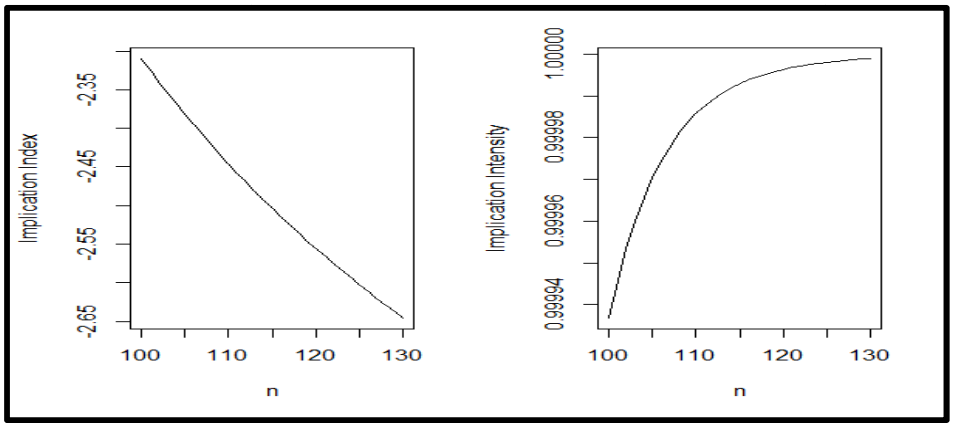


Figure 2: Compares the variation between Implication index and Implication intensity under the parameter n

Figure 2 shows the decreasing variability of Implication index and the increasing variability of Implication intensity under the parameter *n* on ARQAT tool implemented in R (Nghia and Hiep, 2014). They are two measures representing for two measures classes which varies on the parameter n (Implication index for increasing class and Implication intensity for decreasing class).

Table 3: Classification of measures based on partial derivative under the parameter n_A

Decreases with n_A	**Independent with n_A**	**Increases with n_A**	**Others**
Implication index Probabilistic measures of deviation from equilibrium (IPEE), Indice Probabiliste d'Ecart d'Equilibre Leverage, Leverage 1	Prevalence	1-way Support Added value, Pavillon, Centred Confidence, Dependency Bayes factor, Odd multiplier Causal-Confidence Causal-Confirmed confidence Loevinger, Certainty Factor, Satisfaction Confidence Causal Confirm Conviction Coverage Descriptive Confirmed-Confidence, Ganascia Index Descriptive-Confirm Entropic Implication Intensity 1 Entropic Implication Intensity 2 Examples and counter-examples rate Gain, Fukuda Gini index Goodman–Kruskal Implication Intensity MGK, Ion J-measures Klosgen K-measures Kulczynski index Laplace Least contradiction Mutual Information MI, 2-way Support Variation Putative Causal Dependency Recall, Completeness Sebag and Schoenauer Specificity 1, Negative Reliability Zhang Zhang	Relative Risk , Class correlation ratio Collective strength Directed Information ratio (DIR)

The results of Table 3 show that the class of objective interestingness measures varies in the increasing tendency under parameter n_A with relatively high proportion (31/39) and Prevalence measures is the only independent measurement parameter n_A. This result shows that the interestingness measures of the association rule a → b depend on the number of elements of the set A (n_A). As the parameter n_A is greater, the interestingness measures of the association rules reach the maximum value under the variability of the derivative. Particularly, the group of measures that is derived from Confidence measures belongs to the class of the increasing measures by the parameter n_A. This is consistent with the principle for determining the reliability of the association rule a → b. The class of decreasing measures under the parameter n_A shares a small percentage (3 over all measures in our examining). They include measures such as Implication index, IPEE, Leverage. The group consisting of Relative risk, Class correlation ratio, Collective strength, Directed information ratio have variation value depending on the specific value of the parameter n_A.

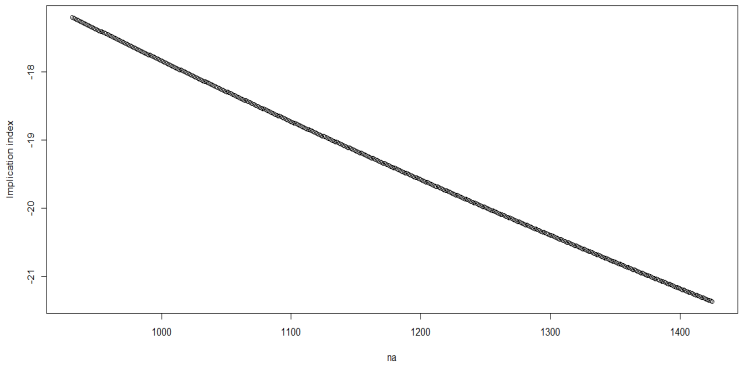


Figure 3: Show the decreasing variability of Implication index under the parameter *1521, ={931, 1425}, =1025, =4*

Implication index measures is the decreasing variability on the parameter . The decreasing variation of this measures is shown in Figure 3. This is a measures representing the class of the decreased measures by the parameter .

Table 4: Classification of measures based on partial derivative under the parameter

Decreases with	**Independent with**	**Increases with**	**Others**
1-way Support Added value, Pavillon, Centred Confidence, Dependency Bayes factor, Odd multiplier Causal-Confidence Causal-Confirmed confidence Loevinger, Certainty Factor, Satisfaction Relative Risk , Class correlation ratio Collective strength Causal Confirm Conviction Entropic Implication Intensity 1 Entropic Implication Intensity 2 Gini index Goodman–Kruskal Implication Intensity J-measures Klosgen K-measures Kulczynski index Least contradiction Leverage, Leverage 1 Mutual Information MI, 2-way Support Variation Putative Causal Dependency Recall, Completeness Specificity 1, Negative Reliability Zhang Zhang	Confidence Coverage Descriptive Confirmed-Confidence , Ganascia Index Descriptive-Confirm Examples and counter-examples rate Gain, Fukuda Probabilistic measures of deviation from equilibrium (IPEE), Indice Probabiliste d'Ecart d'Equilibre Laplace Sebag and Schoenauer	Implication index Prevalence	Directed Information ratio (DIR) MGK, Ion

Table 4 shows that most of the examining measures in groups of variable inverse measures with the parameter . According to this table, it can be confirmed that if the value of increases, the interestingness value will decrease. This is entirely consistent with the principles of the theory to determine the interestingness value of the association

rules because the studied measures are the asymmetry measures. Similarly, when we consider the measures with the parameters n, most of the independent variable measures with the parameter are the descriptive measures as Confidence described, Coverage, Laplace, Sebag and Schoenauer. The class of measures with variable covariates has only two measures: Implication index and Prevalence. It means that the interestingness value of those measures will increase when the value of increases. Like the examining results with parameters n, Directed Information Ratio, MGK have variation value depending on the specific value of the parameter .

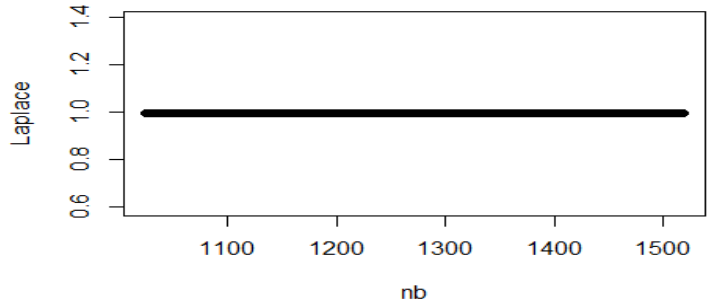


Figure 4: Show independence of Laplace measures under the parameter

Figure 4 shows the independence of Laplace measures under parameter on ARQAT tool implemented in R (Nghia and Hiep, 2014). This is a measures of the class of independent measures of the parameter .

Table 5: Classification of measures based on partial derivative under parameter –

Decreases with –	**Independent with –**	**Increases with –**	**Others**
1-way Support Added value, Pavillon, Centred Confidence, Dependency Bayes factor, Odd multiplier Causal-Confidence Causal-Confirmed confidence Loevinger, Certainty Factor, Satisfaction Relative Risk , Class correlation ratio Collective strength Confidence Causal Confirm Conviction Descriptive Confirmed-Confidence , Ganascia Index Descriptive-Confirm Entropic Implication Intensity 1 Entropic Implication Intensity 2 Examples and counter-examples rate Gain, Fukuda Implication Intensity MGK, Ion Klosgen K-measures Kulczynski index Laplace Least contradiction Leverage, Leverage 1	Coverage Probabilistic measures of deviation from equilibrium (IPEE), Indice Probabiliste d'Ecart d'Equilibre Prevalence	Implication index	Directed Information ratio (DIR) Gini index Goodman–Kruskal J-measures Mutual Information MI, 2-way Support Variation

Putative Causal Dependency Recall, Completeness Sebag and Schoenauer Specificity 1, Negative Reliability Zhang Zhang			

Table 5 shows that the class of objective interestingness measures in the decreasing variation with the parameter accounts for 71%. This reflects accurately the role of the parameter when determining the interestingness measures of the association rule . If the number of counter-examples increases, the interestingness value of the association rule decreases. In this class, most of measures are derived from Confidence measures. This result reflects exactly the formula of determining the reliability of an association rule as defined in the formula of Confidence measures (———). Therefore, the parameter is always inversely related to the reliability. The class of independent measures on the parameter has relatively small proportion of total of studied measures. It includes the three measures Coverage, IPEE, and Prevalence. The group of measures having increasing variation on the parameter is only one measures (Implication index). This is a paradoxical situation in determining the interestingness measures of the association rules because as the number of counter-examples increases interestingness value should decrease, but in this situation it has increased. The last subclass includes five measures, in which the variability of Directed Information ratio measures depends on the constraint of two expressions: ——— and —. The remaining measures have variation value depending on the specific value of the parameter .

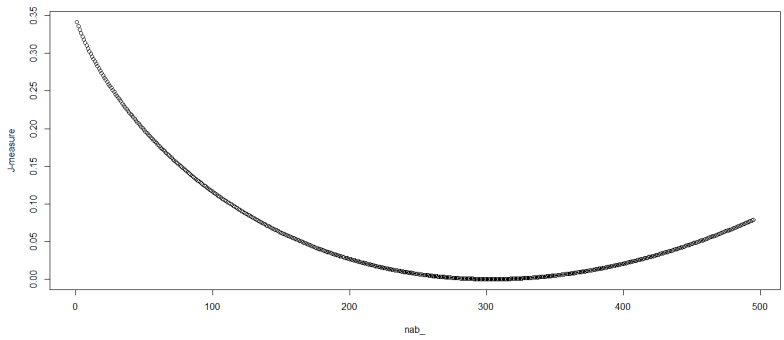


Figure 5: Show variable value depending on the specific value of the parameter of J-measures

The variation value of measures J-measures depends on the specific value of the parameter . The variation of this measures is shown in Figure 5. This measures represents the class of measures whose variation value depending on the specific value of the parameter .

5 Conclusion

Classification of the interestingness measures has attracted many researchers in the field of Knowledge Discovery in Databases. The study focuses on two main classification methods: one based on the properties of the measures and the other based

on the behaviors of the measures. In this paper, we examined the value variation of the asymmetric objective interestingness measures by taking their partial derivatives using the four parameters, $n, n_A, n_B, n_{A\overline{B}}$. The results of classification are considered in the increasing, decreasing or independent variation of the derivative formula of each measures, thus evaluating the variable tendency of the asymmetric objective interestingness measures. This classification helps researchers and users to obtain more insight into the group of asymmetric objective interestingness measures such as the increasing, decreasing variability of each measures, the relationship between the value variability and the statistical parameter values $n, n_A, n_B, n_{A\overline{B}}$, and the interdependence between these parameters in the formula calculating the interestingness value of the measures. This information is the basis for defining an appropriate measures in their specific application.

References

[1] A. Silberschatz and A. Tuzhilin (1995), On subjective measures of interestingness in knowledge discovery, *KDD'95 - Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, 275-281.

[2] A. Silberschatz and A. Tuzhilin (1996), What makes patterns interesting in knowledge discovery systems, *IEEE Transactions on Knowledge and Data Engineering 8(6)*, 970-974.

[3] C. Tew *et al.* (2013), Behavior-based clustering and analysis of interestingness measures for association rule mining, *Journal of Data Mining and Knowledge Discovery 28*, Springer-Verlag, 1004-1045.

[4] David H. Glass (2013), Confirmation measures of association rule interestingness, *Knowledge-Based Systems 44*, 65–77.

[5] F. Guillet and H. J. Hamilton (2007), Quality Measures in Data Mining - Series in Computational Intelligence 43, Springer-Verlag.

[6] G. Piatetsky-Shapiro (1991), Discovery, analysis, and presentation of strong rules, *Knowledge Discovery in Databases*, 229-248.

[7] G. Piatetsky-Shapiro and C. J. Matheus (1994), The interestingness of deviations, *AAAI'94 - Knowledge Discovery in Databases Workshop*, 25-36.

[8] H. X. Huynh *et al.* (2006), Evaluating Interestingness Measures with Linear Correlation Graph, *Advances in Applied Artificial Intelligence (LNCS 4031)*, Springer-Verlag, 312 – 321.

[9] H. X. Huynh *et al.* (2005), Data Analysis Approach for Evaluating the Behavior of Interestingness Measures, *Discovery Science (LNCS 3735)*, Springer-Verlag, 130-137.

[10] H. X. Huynh *et al.* (2006), Extracting representative measures for thepost-processing of association rules, *The 2006 IEEE International Conference on Research, Innovation and Vision for the Future (RIVF'08)*, 100-106.

[11] H. X. Huynh *et al.* (2007), A graph-based clustering approach to evaluate interestingness measures: a tool and a comparative study (Chapter 2), *Quality Measures in Data Mining*, Springer-Verlag, 25-50.

[12] H. X. Huynh *et al.* (2012), Classification of objective interestingness measures, *Journal of Can Tho University (2011:20a)*, 147 - 158.

[13] J. Blanchard *et al.* (2009), Semantics-based classification of rule interestingness measures, *Post-Mining of Association Rules: techniques for effective knowledge extraction" (Y. Zhao, C. Zhang, L. Cao editors), IGI Global*, 56-79.

[14] J. Liu et al. (2008), A New Interestingness Measures of Association Rules, The Second International Conference on Genetic and Evolutionary Computing (WGEC'08), 393 – 397.

[15] Jon Hills *et al.* (2012), Interestingness Measures for Fixed Consequent Rules, *Intelligent Data Engineering and Automated Learning - IDEAL 2012 (LNCS Volume 7435)*, Springer-Verlag, 68–75.

[16] K. Selvarangam and K. Ramesh Kumar (2014), Selecting Perfect Interestingness Measures By Coefficient Of Variation Based Ranking Algorithm, *Journal of Computer Science (Volume 10, Issue 9)*, 1672-1679.

[17] L. Geng and H. J. Hamilton (2006), Interestingness measures for data mining: A survey, *ACM Computing Surveys (Volume 38)*, 1-32.

[18] M. Jalali-Heravi and O. Zaïane (2010), A study on interestingness measures for associative classifiers, *SAC'10 - Proceedings of the 25th ACM Symposium on Applied Computing*, 1039 -1046.

[19] M. Martínez-Ballesteros *et al.* (2014), Selecting the best measures to discover quantitative association rule, *Neurocomputing (Volume 126)*, 3–14.

[20] Maddouri and Gammoudi (2007), On Semantic Properties of Interestingness Measures for Extracting Rules from Data, *ICANNGA (1) 2007, Springer-Verlag Berlin Heidelberg*, 148-158

[21] P. Lenca *et al.* (2004), A multicriteria decision aid for interestingness measures selection, LUSSI-TR-2004-01-EN, 1–27.

[22] P. Lenca *et al.* (2008), On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid, *European Journal of Operational Research (Volume 184, Issue 2)*, 610–626.

[23] P. N. Tan *et al.* (2002), Selecting the Right Interestingness Measures for Association Patterns, *SIGKDD '02 Edmonton, Alberta, Canada ACM 1-58113-567-X/02/0007*, 1-10.

[24] P. N. Tan *et al.* (2004), Selecting the right objective measures for association analysis, *Journal of Information Systems (Volume 29, Issue 4)*, 293-313.

[25] R. Agrawal and R. Srikant (1994), Fast algorithms for mining association rules, *VLDB'94 - Proceedings of the 20th International Conference on Very Large Data Bases*, 487-499.

[26] R. Gras (1979), Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs en didactique des mathématiques, *PhD thesis, Université de Rennes 1*.

[27] R. Gras and P. Kuntz (2008), An overview of the Statistical Implicative Analysis (SIA) development, *Statistical Implicative Analysis - Studies in Computational Intelligence (Volume 127)*, Springer-Verlag, 11-40.

[28] S. Brin *et al.* (1997), Dynamic itemset counting and implication rules for market basket data, *Proceedings of the ACM SIGMOD international conference on management of data*, 255–264.

[29] S. Guillaume et al. (2012), Categorization of interestingness measures for knowledge extraction, journals/corr/abs-1206-6741, 1 – 34.

[30] S. Sahar (2003), What is interesting: studies on interestingness in knowledge discovery. *PhD thesis, School of Computer Science, Tel-Aviv University*.

Appendix 1

Formula of asymmetric objective interestingness measures

No	Name of interestingness measures	$f(n, n_A, n_B, n_{A\bar{B}})$
1.	1-way Support	$\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n(n_A - n_{A\bar{B}})}{n_A n_B}$
2.	Added value, Pavillon, Centred Confidence, Dependency	$\frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_B}{n}$
3.	Bayes factor, Odd multiplier	$\frac{n n_A - n_A n_B - n n_{A\bar{B}} + n_B n_{A\bar{B}}}{n_B n_{A\bar{B}}}$
4.	Causal-Confidence	$1 - \frac{1}{2}\left(\frac{1}{n_A} + \frac{1}{n - n_B}\right) n_{A\bar{B}}$
5.	Causal-Confirmed confidence	$1 - \frac{1}{2}\left(\frac{3}{n_A} + \frac{1}{n - n_B}\right) n_{A\bar{B}}$
6.	Loevinger, Certainty Factor, Satisfaction	$1 - \frac{n n_{A\bar{B}}}{n_A (n - n_B)}$
7.	Relative Risk , Class correlation ratio	$\frac{(n_A - n_{A\bar{B}})(n - n_A)}{n_A (n_B - n_A + n_{A\bar{B}})}$
8.	Collective strength	$\frac{(n_A - n_{A\bar{B}})(n - n_B - n_{A\bar{B}})(n_A (n - n_B) + n_B (n - n_A))}{((n - n_A)(n - n_B) + n_A n_B)(n_B - n_A + 2 n_{A\bar{B}})}$
9.	Confidence	$\frac{n_A - n_{A\bar{B}}}{n_A}$
10.	Causal Confirm	$\frac{n + n_A - n_B - 4 n_{A\bar{B}}}{n}$
11.	Conviction	$\frac{n_A (n - n_B)}{n n_{A\bar{B}}}$
12.	Coverage	$\frac{n_A}{n}$
13.	Descriptive Confirmed-Confidence, Ganascia Index	$1 - \frac{2 n_{A\bar{B}}}{n_A}$
14.	Descriptive-Confirm	$\frac{n_A - 2 n_{A\bar{B}}}{n}$
15.	Entropic Implication Intensity 1	$\sqrt{II\left(\left(1 - H^{\propto}_{B\|A}\right)\left(1 - H^{\propto}_{\bar{A}\|\bar{B}}\right)\right)^{\frac{1}{2\propto}}}$ with (α=1) and $H_{A\|B} = \frac{n_A - n_{A\bar{B}}}{n_B} log_2 \frac{n_{A\bar{B}} - n_A}{n_B} + \frac{n_A - n_B - n_{A\bar{B}}}{n_B} log_2 \frac{n_B - n_A + n_{A\bar{B}}}{n_B}$ Where II is Inplication intensity
16.	Entropic Implication Intensity 2	$\sqrt{II\left(\left(1 - H^{\propto}_{B\|A}\right)\left(1 - H^{\propto}_{\bar{A}\|\bar{B}}\right)\right)^{\frac{1}{2\propto}}}$ with (α=2) and $H_{A\|B} = \frac{n_A - n_{A\bar{B}}}{n_B} log_2 \frac{n_{A\bar{B}} - n_A}{n_B} + \frac{n_A - n_B - n_{A\bar{B}}}{n_B} log_2 \frac{n_B - n_A + n_{A\bar{B}}}{n_B}$

17.	Examples and counter-examples rate (Exam-Cex-rate, Excounterex rate)	$\frac{n_A - 2n_{A\bar{B}}}{n_A - n_{A\bar{B}}}$
18.	Gain, Fukuda	$\frac{n_A(1-\theta) - n_{A\bar{B}}}{n}$
19.	Gini index	$\frac{(n_A - n_{A\bar{B}})^2 + {n_{A\bar{B}}}^2}{nn_A} + \frac{(n_B - n_A + n_{A\bar{B}})^2 + (n - n_B - n_{A\bar{B}})^2}{n(n-n_A)} - \frac{{n_B}^2 + (n-n_B)^2}{n^2}$
20.	Goodman–Kruskal	$\frac{\alpha}{\beta}$ where $\alpha = max\left(\frac{n_A - n_{A\bar{B}}}{n}, \frac{n_{A\bar{B}}}{n}\right) + max\left(\frac{n_B - n_A + n_{A\bar{B}}}{n}, \frac{n - n_B - n_{A\bar{B}}}{n}\right) + max\left(\frac{n_A - n_{A\bar{B}}}{n}, \frac{n_B - n_A + n_{A\bar{B}}}{n}\right) + max\left(\frac{n_{A\bar{B}}}{n}, \frac{n - n_B - n_{A\bar{B}}}{n}\right) - max\left(\frac{n_A}{n}, \frac{n-n_A}{n}\right) - max\left(\frac{n_B}{n}, \frac{n-n_B}{n}\right)$ $\beta = 2 - max\left(\frac{n_A}{n}, \frac{n-n_A}{n}\right) - max\left(\frac{n_B}{n}, \frac{n-n_B}{n}\right)$
21.	Implication index	$q(A,\bar{B}) = \frac{n_{A\bar{B}} - \frac{n_A(n-n_B)}{n}}{\sqrt{\frac{n_A(n-n_B)}{n}}}$
22.	Implication Intensity (II)	$\frac{1}{\sqrt{2\pi}} \int_{q(A,\bar{B})}^{+\infty} e^{-\frac{t^2}{2}} dt$ **Or** $1 - \sum_{k=max\,(0,n_A-n_B)}^{n_{A\bar{B}}} \frac{C_{n_B}^{n_A-k} C_{(n-n_B)}^{k}}{C_n^{n_A}}$
23.	Probabilistic measures of deviation from equilibrium (IPEE),Indice Probabiliste d'Ecart d'Equilibre	$1 - \frac{2}{2^{n_A}} \sum_{k=0}^{n_{A\bar{B}}} C_{n_A}^{k}$
24.	Directed Information ratio(DIR)	$\begin{cases} -\infty & if\ \frac{n_B}{n} = 1 \\ 0 & if\ \frac{n_B}{n} \le \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} \le \frac{1}{2} \\ 1 + \frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} + \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A} & if\ \frac{n_B}{n} \le \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} > \frac{1}{2} \\ \frac{1}{1 + \frac{n_B}{n} log_2 \frac{n_B}{n} + \frac{n-n_B}{n} log_2 \frac{n-n_B}{n}} & if\ \frac{n_B}{n} > \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} \le \frac{1}{2} \\ 1 - \frac{\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} + \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A}}{\frac{n_B}{n} log_2 \frac{n_B}{n} + \frac{n-n_B}{n} log_2 \frac{n-n_B}{n}} & if\ \frac{n_B}{n} > \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} > \frac{1}{2} \end{cases}$
25.	MGK, Ion	$\begin{cases} 1 - \frac{nn_{A\bar{B}}}{n_A(n-n_B)} & if\ \frac{n - n_{A\bar{B}}}{n_A} > \frac{n_B}{n} \\ \frac{n(n_A - n_{A\bar{B}})}{n_A n_B} - 1 & otherwise \end{cases}$
26.	J-measures	$\frac{n_A - n_{A\bar{B}}}{n} log_2 \frac{n(n_A - n_{A\bar{B}})}{n_A n_B} + \frac{n_{A\bar{B}}}{n} log_2 \frac{nn_{A\bar{B}}}{n_A(n-n_B)}$

27.	Klosgen	$\sqrt{\frac{n_A - n_{A\bar{B}}}{n}}\left(\frac{n - n_B}{n} - \frac{n_{A\bar{B}}}{n_A}\right)$
28.	K-measures	$\left(\frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n - n_B - n_{A\bar{B}}}{n - n_A}\right)(log_2(n - n_B) - log_2 n_B)$
29.	Kulczynski index	$\frac{(n_A - n_{A\bar{B}})}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)$
30.	Laplace	$\frac{n_A - n_{A\bar{B}} + 1}{n_A + 2}$
31.	Least contradiction	$\frac{n_A - 2n_{A\bar{B}}}{n_B}$
32.	Leverage, Leverage 1	$1 - \frac{n_{A\bar{B}}}{n_A} - \frac{n_A n_B}{n^2}$
33.	Mutual Information **MI,** 2-way Support Variation	$\frac{n_A - n_{A\bar{B}}}{n} log_2 \frac{n(n - n_{A\bar{B}})}{n_A n_B} + \frac{n_{A\bar{B}}}{n} log_2 \frac{n n_{A\bar{B}}}{n_A(n - n_B)} + \frac{n_B - n_A + n_{A\bar{B}}}{n} log_2 \frac{n(n_B - n_A + n_{A\bar{B}})}{(n - n_A)n_B} + \frac{n - n_B - n_{A\bar{B}}}{n} log_2 \frac{n(n - n_B - n_{A\bar{B}})}{(n - n_A)(n - n_B)}$
34.	Prevalence	$\frac{n_B}{n}$
35.	Putative Causal Dependency	$\frac{3}{2} + \frac{4n_A - 3n_B}{2n} - \left(\frac{3}{2n_A} + \frac{2}{n - n_B}\right) n_{A\bar{B}}$
36.	Recall, Completeness	$\frac{n_A - n_{A\bar{B}}}{n_B}$
37.	Sebag and Schoenauer	$\frac{n_A}{n_{A\bar{B}}} - 1$
38.	Specificity 1, Negative Reliability	$\frac{n - n_B - n_{A\bar{B}}}{n - n_A}$
39.	Zhang Zhang	$\frac{n n_A - n_A n_B - n n_{A\bar{B}}}{max\ ((n_A - n_{A\bar{B}})(n - n_B), n_B n_{A\bar{B}})}$

Appendix 2

Formula of partial derivative under the parameter n

No	$\frac{\partial Z}{\partial n}$
1	$\frac{n_A - n_{A\bar{B}}}{n n_A ln2}$
2	$\frac{n_B}{n^2}$
3	$\frac{n_A - n_{A\bar{B}}}{n_B n_{A\bar{B}}}$
4	$\frac{n_{A\bar{B}}}{2(n - n_B)^2}$
5	$\frac{n_{A\bar{B}}}{2(n - n_B)^2}$
6	$\frac{n_B n_{A\bar{B}}}{n_A(n - n_B)^2}$

7	$\frac{(n_A - n_{A\bar{B}})}{n_A(n_B - n_A + n_{A\bar{B}})}$
8	$\frac{(n_A - n_{A\bar{B}})\left(n_A(n-n_B) + n_B\big((n-n_A)\big) + (n - n_B - n_{A\bar{B}})(n_A + n_B)\right)\big((n-n_A)(n-n_B) + n_A n_B\big)(n_B - n_A + 2n_{A\bar{B}}) - k}{\left(\big((n-n_A)(n-n_B) + n_A n_B\big)(n_B - n_A + 2n_{A\bar{B}})\right)^2}$ $k = (n_A - n_{A\bar{B}})\left(n_A(n-n_B) + n_B\big((n-n_A)\big) + (n - n_B - n_{A\bar{B}})(n_A + n_B)\right)((n-n_B) + (n-n_B)(n_B - n_A + 2n_{A\bar{B}})$
9	0
10	$\frac{-n_A + n_B + 4n_{A\bar{B}}}{n^2}$
11	$\frac{n_A n_B}{n_{A\bar{B}} n^2}$
12	$-\frac{n_A}{n^2}$
13	0
14	$-\frac{n_A - 2n_{A\bar{B}}}{n^2}$
15	$\frac{II'_n(h)^{\frac{1}{2}} + II\frac{1}{2}(h)^{-\frac{1}{2}}\left(1 - \left(-\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A}\right)\left(1 - \left(-\frac{1}{n_B}(log_2 \frac{n - n_B - n_{A\bar{B}}}{n_B} - \frac{1}{ln2})\right)\right)\right)}{2\sqrt{II(h)^{\frac{1}{2}}}}$ $h = 1 - \left(-\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A}\right)\left(1 - \left(-\frac{n_{A\bar{B}}}{n_B} log_2 \frac{n_{A\bar{B}}}{n_B} - \frac{n - n_B - n_{A\bar{B}}}{n_B} log_2 \frac{n - n_B - n_{A\bar{B}}}{n_B}\right)\right)$
16	$\frac{II'_n(h)^{\frac{1}{4}} + II\frac{1}{4}(h)^{-\frac{3}{4}}\left(1 - \left(-\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A}\right)^2\right)k}{2\sqrt{II(h)^{\frac{1}{4}}}}$ $h = 1 - \left(-\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A}\right)^2\left(1 - \left(-\frac{n_{A\bar{B}}}{n_B} log_2 \frac{n_{A\bar{B}}}{n_B} - \frac{n - n_B - n_{A\bar{B}}}{n_B} log_2 \frac{n - n_B - n_{A\bar{B}}}{n_B}\right)^2\right)$ $k = \left(-2\left(-\frac{n_{A\bar{B}}}{n_B} log_2 \frac{n_{A\bar{B}}}{n_B} - \frac{n - n_B - n_{A\bar{B}}}{n_B} log_2 \frac{n - n_B - n_{A\bar{B}}}{n_B}\right)(-\frac{1}{n_B}(log_2 \frac{n - n_B - n_{A\bar{B}}}{n_B} + \frac{1}{ln2})\right)$
17	0
18	$-\frac{n_A(1-\theta) - n_{A\bar{B}}}{n^2}$
19	$-\frac{(n_A - n_{A\bar{B}})^2 + {n_{A\bar{B}}}^2}{n^2 {n^2}_A}$ $+\frac{(n_B - n_A + n_{A\bar{B}})^2(2n - n_A) + 2(n - n_B - n_{A\bar{B}}) - (n - n_B - n_{A\bar{B}})^2(2n - n_A)}{(n(n-n_A))^2}$ $-\frac{2n{n_B}^2 + 2(n - n_A) - 2n(n - n_B)^2}{n^4}$
20	$\frac{\alpha}{\beta}$ where

	$\alpha = \left(\frac{n_A - n_{A\bar{B}}}{n^2} + \frac{n - n_B - n_{A\bar{B}}}{n^2} + \frac{n_A - n_{A\bar{B}}}{n^2} + \frac{n - n_B - n_{A\bar{B}}}{n^2} - \frac{n - n_A}{n^2} - \frac{n - n_B}{n^2}\right)\left(2 - max\left(\frac{n_A}{n}, \frac{n - n_A}{n}\right) - max\left(\frac{n_B}{n}, \frac{n - n_B}{n}\right)\right) - \left(max\left(\frac{n_A - n_{A\bar{B}}}{n}, \frac{n_{A\bar{B}}}{n}\right) + max\left(\frac{n_B - n_A + n_{A\bar{B}}}{n}, \frac{n - n_B - n_{A\bar{B}}}{n}\right) + max\left(\frac{n_A - n_{A\bar{B}}}{n}, \frac{n_B - n_A + n_{A\bar{B}}}{n}\right) + max\left(\frac{n_{A\bar{B}}}{n}, \frac{n - n_B - n_{A\bar{B}}}{n}\right) - max\left(\frac{n_A}{n}, \frac{n - n_A}{n}\right) - max\left(\frac{n_B}{n}, \frac{n - n_B}{n}\right)\right)\left(-\frac{n - n_A}{n^2} - \frac{n - n_B}{n^2}\right)$ $\beta = (2 - max\left(\frac{n_A}{n}, \frac{n - n_A}{n}\right) - max\left(\frac{n_B}{n}, \frac{n - n_B}{n}\right))^2$
21	$\frac{1}{2\sqrt{n}}\left(n_{A\bar{B}} + \frac{n_A(n - n_B)}{n}\right)$
22	$-\sum_{k=max\,(1, n_A - n_B)}^{n_{A\bar{B}}} \frac{n_B!}{((n_A - k)!\,(n_B - n_A + k)!\,k!\,(n - n_B - k - 1)!} \frac{(n - n_B - 1)!}{} \frac{n_A!\,(n - n_A - 1)!}{(n - 1)!}$
23	0
24	$\begin{cases} 0 & if(\frac{n_B}{n} \leq \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} > \frac{1}{2} \\ \frac{\frac{1}{n^2}(log_2 \frac{n_B}{n} + \frac{n_B}{ln2}) + (\frac{n_B}{n^2}\ (log_2 \frac{n - n_B}{n}) + \frac{1}{ln2})}{(\frac{n_B}{n} log_2 \frac{n_B}{n} + \frac{n - n_B}{n} log_2 \frac{n - n_B}{n})^2} & if(\frac{n_B}{n} > \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} \leq \frac{1}{2} \\ -\frac{(\frac{n_A - n_{A\bar{B}}}{n_A} log_2 \frac{n_A - n_{A\bar{B}}}{n_A} + \frac{n_{A\bar{B}}}{n_A} log_2 \frac{n_{A\bar{B}}}{n_A})(\frac{1}{n^2}(log_2 \frac{n_B}{n} + \frac{n_B}{ln2}) + (\frac{n_B}{n^2}\ (log_2 \frac{n - n_B}{n}) + \frac{1}{ln2}))}{(\frac{n_B}{n} log_2 \frac{n_B}{n} + \frac{n - n_B}{n} log_2 \frac{n - n_B}{n})^2} & if(\frac{n_B}{n} > \frac{1}{2}\ and\ \frac{n_A - n_{A\bar{B}}}{n_A} > \frac{1}{2} \end{cases}$
25	$\begin{cases} -\frac{n_{A\bar{B}}(nn_A - n_A n_B) - nn_A n_{A\overline{B}}}{(nn_A - n_A n_B)^2} & if\ \frac{n - n_{A\bar{B}}}{n_A} > \frac{n_B}{n} \\ \frac{(n_A - n_{A\bar{B}})}{n_A n_B} & otherwise \end{cases}$
26	$\frac{n_A - n_{A\bar{B}}}{n^2}(-log_2 \frac{n(n_A - n_{A\bar{B}})}{n_A n_B} + \frac{1}{ln2}) + \frac{n_{A\bar{B}}}{n^2}\left(log_2 \frac{nn_{A\bar{B}}}{n_A(n - n_B)} - \frac{n_B}{(n - n_B)ln2}\right)$
27	$\left(\frac{\frac{n_A - n_{A\bar{B}}}{n^2}}{2\sqrt{\frac{n_A - n_{A\bar{B}}}{n}}}\left(\frac{n - n_B}{n} - \frac{n_{A\bar{B}}}{n_A}\right)\right) + \left(\sqrt{\frac{n_A - n_{A\bar{B}}}{n}}\right)\left(\frac{n_B}{n^2}\right)$
28	$\left(-\frac{(n - n_A) - (n - n_B - n_{A\bar{B}})}{(n - n_A)^2}\right)(log_2(n - n_B) - log_2 n_B) + \left(\frac{n_A - n_{A\bar{B}}}{n_A} - \frac{n - n_B - n_{A\bar{B}}}{n - n_A}\right)\left(\frac{1}{(n - n_B)ln2}\right)$
29	0
30	0
31	0
32	$\frac{2nn_A n_B}{n^4}$

33	$\frac{n_A - n_{A\bar{B}}}{n^2}(-log_2 \frac{n(n-n_{A\bar{B}})}{n_A n_B} + \frac{(2n - n_{A\bar{B}})}{n(n-n_{A\bar{B}})ln2}) + \frac{n_{A\bar{B}}}{n^2}(-log_2 \frac{nn_{A\bar{B}}}{n_A(n-n_B)} - \frac{n_B}{(n-n_B)ln2}) + \frac{n_B - n_A + n_{A\bar{B}}}{n^2}(-log_2 \frac{n(n_B - n_A + n_{A\bar{B}})}{(n-n_A)n_B} + \frac{(n_B - n_A + n_{A\bar{B}})(n-n_A) - n(n_B - n_A + n_{A\bar{B}})}{(n_B - n_A + n_{A\bar{B}})ln2}) + \frac{n - n_B - n_{A\bar{B}}}{n^2}(log_2 \frac{n(n-n_B-n_{A\bar{B}})}{(n-n_A)(n-n_B)} + \frac{(2n - n_B - n_{A\bar{B}})(n-n_A)(n-n_B) - n(n-n_B-n_{A\bar{B}})(2n-n_B-n_A)}{(n-n_A)(n-n_B)(n-n_B-n_{A\bar{B}})ln2}$
34	$-\frac{n_B}{n^2}$
35	$-\frac{4n_A - 3n_B}{2n^2} + \frac{2n_{A\bar{B}}}{(n-n_B)^2}$
36	0
37	0
38	$\frac{-n_A + n_B + n_{A\bar{B}}}{(n-n_A)^2}$
39	$\frac{(n_A - n_{A\bar{B}})\, max\big((n_A - n_{A\bar{B}})(n-n_B), n_B n_{A\bar{B}}\big) - (nn_A - n_A n_B - nn_{A\bar{B}})(n_A - n_{A\bar{B}})}{(max\big((n_A - n_{A\bar{B}})(n-n_B), n_B n_{A\bar{B}}\big))^2}$

Appendix 3

Formula of partial derivative under the parameter n_A

No	$\frac{\partial Z}{\partial n_A}$
1	$\frac{n_{A\bar{B}}}{n^2{}_A}(log_2 \frac{n(n_A - n_{A\bar{B}})}{n_A n_B} + \frac{1}{nln2})$
2	$\frac{n_{A\bar{B}}}{n_A^2}$
3	$\frac{n - n_B}{n_B n_{A\bar{B}}}$
4	$\frac{n_{A\bar{B}}}{2n_A{}^2}$
5	$\frac{3n_{A\bar{B}}}{2n_A{}^2}$
6	$\frac{1}{n_A^2} \cdot \frac{nn_{A\bar{B}}}{(n-n_B)}$
7	$\frac{nn_A^2 - n_A^2 n_B + nn_B n_{A\bar{B}} - 2nn_A n_{A\bar{B}} + nn_{A\bar{B}}^2}{(n_A(n_B - n_A + n_{A\bar{B}}))^2}$
8	$\frac{h + (n_A - n_{A\bar{B}})(n - n_B - n_{A\bar{B}})(n_A(n-n_B) + n_B(n-n_A)\big((n-n_A)(n-n_B)\big)(n_B - n_A + 2n_{A\bar{B}})((n-n_A)(n-n_B) + n_A n_B)}{\big(((n-n_A)(n-n_B) + n_A n_B)(n_B - n_A + 2n_{A\bar{B}})\big)^2}$ $h = \big((n - n_B - n_{A\bar{B}})(n_A(n-n_B)(n_A + n_B)\big)\big(n_B\big((n-n_A)\big)(n_A + n_{A\bar{B}})(n-n_B) - n_B\big)\big((n-n_A)(n-n_B) + n_A n_B\big)(n_B - n_A + 2n_{A\bar{B}})$
9	$\frac{n_{A\bar{B}}}{n_A^2}$
10	$\frac{1}{n}$

11	$\frac{(n-n_B)}{nn_{A\bar{B}}}$
12	$\frac{1}{n}$
13	$\frac{2n_{A\bar{B}}}{n_A^2}$
14	$\frac{1}{n}$
15	$\frac{II'_{n_A}(h)^{\frac{1}{2}} + II\frac{1}{2}(h)^{-\frac{1}{2}}\left(\left(\frac{n_{A\bar{B}}}{n_A^2}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{1}{ln2})+\frac{n_{A\bar{B}}}{n_A^2}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{1}{ln2})\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)\right)}{2\sqrt{II\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)\right)^{\frac{1}{2}}}}$ $h = 1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)$
16	$\frac{II'_n(h)^{\frac{1}{4}} + II\frac{1}{4}(h)^{-\frac{3}{4}}\left(\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)-k\right)}{2\sqrt{II(h)^{\frac{1}{4}}}}$ $h = 1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)$ $k = 2\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(-\frac{n_{A\bar{B}}}{n_A^2}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{1}{ln2})-\frac{1}{n_A^2}(log_2\frac{n_{A\bar{B}}}{n_A}-+\frac{1}{ln2})\right)$
17	$\frac{n_{A\bar{B}}}{(n_A-n_{A\bar{B}})^2}$
18	$\frac{(1-\theta)}{n}$
19	$\frac{2nn_A(n_A-n_{A\bar{B}})-n_A((n_A-n_{A\bar{B}})^2+n_{A\bar{B}}{}^2)}{n^2n^2{}_A} - \frac{2(n_B-n_A+n_{A\bar{B}})+((n_B-n_A+n_{A\bar{B}})^2+(n-n_B-n_{A\bar{B}})^2)(2n-n_A}{(n(n-n_A))^2}$
20	$\frac{\alpha}{\beta}$ where $\alpha = \left(\frac{3}{n}\right)\left(2-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right)\right) - \left(max\left(\frac{n_A-n_{A\bar{B}}}{n},\frac{n_{A\bar{B}}}{n}\right)+max\left(\frac{n_B-n_A+n_{A\bar{B}}}{n},\frac{n-n_B-n_{A\bar{B}}}{n}\right) + max\left(\frac{n_A-n_{A\bar{B}}}{n},\frac{n_B-n_A+n_{A\bar{B}}}{n}\right)+max\left(\frac{n_{A\bar{B}}}{n},\frac{n-n_B-n_{A\bar{B}}}{n}\right) - max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right)\right)(\frac{1}{n})$ $\beta = (2-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right))^2$

21	$-\frac{1}{2\sqrt{\frac{n-n_B}{n}}}\frac{n_{A\bar{B}}}{}\left(\frac{n}{n_A}\right)^{\frac{3}{2}}-\frac{1}{2}\sqrt{\frac{n-n_B}{n_A}}$
22	$-\sum_{k=max\,(1,n_A-n_B)}^{n_{A\bar{B}}}\frac{n_B!}{((n_A-k-1)!\,(n_B-n_A+k)!}\frac{(n-n_B)!}{k!\,(n-n_B-k)!}\frac{(n_A-1)!\,(n-n_A)!}{n!}$
23	$-\frac{2}{2^{n_A}ln2}\sum_{k=1}^{n_{A\bar{B}}}\frac{(n_A-1)!}{k!\,(n_A-k-1)!}$
24	$\begin{cases}\frac{n_{A\bar{B}}}{n^2{}_A}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{n_{A\bar{B}}}{ln2})+\frac{1}{n^2{}_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{n_{A\bar{B}}}{ln2}) & if(\frac{n_B}{n}\leq\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2}\\ 0 & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}\leq\frac{1}{2}\\ -\frac{\frac{n_{A\bar{B}}}{n^2{}_A}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{n_{A\bar{B}}}{ln2})+\frac{1}{n^2{}_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{n_{A\bar{B}}}{ln2})}{\frac{n_B}{n}log_2\frac{n_B}{n}+\frac{n-n_B}{n}log_2\frac{n-n_B}{n}} & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2}\end{cases}$
25	$\begin{cases}\frac{nn_{A\bar{B}}}{(n-n_B)n^2{}_A} & if\ \frac{n-n_{A\bar{B}}}{n_A}>\frac{n_B}{n}\\ \frac{(n_An_B(n_A-n_{A\bar{B}}))-nn_B(n_A-n_{A\bar{B}})}{n_Bn^2{}_A} & otherwise\end{cases}$
26	$\frac{1}{n}(log_2\frac{n(n_A-n_{A\bar{B}})}{n_An_B}-\frac{n_Bn_{A\bar{B}}}{ln2})+\frac{n_{A\bar{B}}}{nn_Aln2}$
27	$\left(\frac{\frac{1}{n}}{2\sqrt{\frac{n_A-n_{A\bar{B}}}{n}}}\left(\frac{n-n_B}{n}-\frac{n_{A\bar{B}}}{n_A}\right)\right)-\sqrt{\frac{n_A-n_{A\bar{B}}}{n}}\left(\frac{n_{A\bar{B}}}{n^2{}_A}\right)$
28	$\left(\frac{n_{A\bar{B}})}{n^2{}_A}-\frac{n-n_B-n_{A\bar{B}}}{(n-n_A)^2}\right)(log_2(n-n_B)-log_2\,n_B)$
29	$\frac{1}{2}\left(\frac{1}{n_B}+\frac{n_{A\bar{B}}}{n_A^2}\right)$
30	$\frac{3+n_{A\bar{B}}}{(n_A+2)^2}$
31	$\frac{1}{n_B}$
32	$\frac{n_{A\bar{B}}}{n_A^2}-\frac{n_B}{n^2}$
33	$\frac{1}{n}(log_2\frac{n(n-n_{A\bar{B}})}{n_An_B}+\frac{n_A-n_{A\bar{B}}}{n_Aln2})$ $+\frac{n_{A\bar{B}}}{nn_Aln2}-\frac{1}{n}log_2\frac{n(n_B-n_A+n_{A\bar{B}})}{(n-n_A)n_B}$ $+\frac{n_B-n_A+n_{A\bar{B}}}{n}(\frac{(n_B-n_A+n_{A\bar{B}})(n-n_A)}{(n-n_A)(n_B-n_A+n_{A\bar{B}})ln2})$ $+\frac{n-n_B-n_{A\bar{B}}}{n}\frac{(n(n-n_B-n_{A\bar{B}}))(n-n_B)}{((n-n_A)(n-n_B))^2ln2}$
34	0
35	$\frac{2}{n}+\frac{3n_{A\bar{B}}}{2n_A^2}$
36	$\frac{1}{n_B}$
37	$\frac{1}{n_{A\bar{B}}}$

38	$\frac{n-n_B-n_{A\bar{B}}}{(n-n_A)^2}$
39	$\frac{(n-n_B)\,max\big((n_A-n_{A\bar{B}})(n-n_B),n_Bn_{A\bar{B}}\big)-(nn_A-n_An_B-nn_{A\bar{B}})(n-n_B)}{(max\big((n_A-n_{A\bar{B}})(n-n_B),n_Bn_{A\bar{B}}\big))^2}$

Appendix 4

Formula of partial derivative under the parameter n_B

No	$\frac{\partial Z}{\partial n_B}$
1	$-\frac{n_A-n_{A\bar{B}}}{n_An_B.\,ln2}$
2	$-\frac{1}{n}$
3	$-\frac{n_A-n_{A\bar{B}}}{n_{A\bar{B}}}.\frac{n}{n_B^2}$
4	$\frac{n_{A\bar{B}}}{2(n-n_B)^2}$
5	$\frac{n_{A\bar{B}}}{2(n-n_B)^2}$
6	$\frac{nn_{A\bar{B}}}{n_A(n-n_B)^2}$
7	$-\frac{(n_A-n_{A\bar{B}})(n-n_A)}{n_A(n_B-n_A+n_{A\bar{B}})^2}$
8	$\frac{-\big((n_A-n_{A\bar{B}})(n_A(n-n_B)\big)+h+k+l}{\Big(\big((n-n_A)(n-n_B)+n_An_B\big)(n_B-n_A+2n_{A\bar{B}})\Big)^2}$ $h=\Big(n_B\big((n-n_A)\big)+(n-n_B-n_{A\bar{B}})\Big)\big((n-2n_A)\big)\big((n-n_A)(n-n_B)+n_An_B\big)(n_B-n_A+2n_{A\bar{B}})$ $k=(n_A-n_{A\bar{B}})(n-n_B-n_{A\bar{B}})\big(n_A(n-n_B)\big)$ $l=n_B(n-n_A)\big((n_B-n_A+2n_{A\bar{B}})(n-n_A)\big)(2n_{A\bar{B}})\big((n-n_A)n-n_B(n-n_B)+n_An_B\big)$
9	0
10	$-\frac{1}{n}$
11	$-\frac{n_A}{nn_{A\bar{B}}}$
12	0
13	0
14	0
15	$\frac{II'_{n_A}(h)^{\frac{1}{2}}+II\frac{1}{2}(h)^{-\frac{1}{2}}\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(\left(\frac{n_{A\bar{B}}}{n_B^2}(log_2\frac{n_{A\bar{B}}}{n_B}+\frac{1}{ln2})-\frac{(n-n_{A\bar{B}})}{n_B^2}(log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}+\frac{1}{ln2})\right)\right)\right)}{2\sqrt{II\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)\right)}}$ $h=1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)$

16	$\frac{II'_n(h)^{\frac{1}{4}} + II\frac{1}{4}(h)^{-\frac{3}{4}}\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2 l\right)}{2\sqrt{II\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)\right)}}^{\frac{1}{4}}$ $h = 1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)$ $l = \left(-2\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\left(-\frac{n_{A\bar{B}}}{n_B^2}(log_2\frac{n_{A\bar{B}}}{n_B}+\frac{1}{ln2})-\frac{(n-n_{A\bar{B}})}{n_B^2}(log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}+\frac{1}{ln2})\right)\right)$
17	0
18	0
19	$\frac{2(n_B-n_A+n_{A\bar{B}})-2(n-n_B-n_{A\bar{B}})}{n(n-n_A)}-\frac{2n_B+2(n-n_B)}{n^2}$
20	$\frac{\alpha}{\beta}$ *where* $\alpha = \left(\frac{2}{n}\right)\left(2-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right)\right)(max\left(\frac{n_A-n_{A\bar{B}}}{n},\frac{n_{A\bar{B}}}{n}\right)+max\left(\frac{n_B-n_A+n_{A\bar{B}}}{n},\frac{n-n_B-n_{A\bar{B}}}{n}\right)+max\left(\frac{n_A-n_{A\bar{B}}}{n},\frac{n_B-n_A+n_{A\bar{B}}}{n}\right)+max\left(\frac{n_{A\bar{B}}}{n},\frac{n-n_B-n_{A\bar{B}}}{n}\right)-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right))(\frac{1}{n})$ $\beta = (2-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-max\left(\frac{n_B}{n},\frac{n-n_B}{n}\right))^2$
21	$\frac{1}{2}n_{A\bar{B}}(\frac{n_A}{n})^{-\frac{1}{2}}(n-n_B)^{-\frac{3}{2}}+\frac{1}{2}(\frac{n_A}{n})^{\frac{1}{2}}(n-n_B)^{-\frac{1}{2}}$
22	$-\sum_{k=max\,(1,n_A-n_B)}^{n_{A\bar{B}}}\frac{(n_B-1)!}{((n_A-k)!\,(n_B-n_A+k-1)!}\frac{(n-n_B)!}{k!\,(n-n_B-k)!}\frac{n_A!\,(n-n_A)!}{n!}$
23	0
24	$\begin{cases} 0 & if(\frac{n_B}{n}\le\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2} \\ \frac{\frac{1}{n}(log_2\frac{n_B}{n}+\frac{1}{ln2})-\frac{1}{n}log_2\frac{n-n_B}{n}-\frac{1}{(n-n_B)ln2}}{(\frac{n_B}{n}log_2\frac{n_B}{n}+\frac{n-n_B}{n}log_2\frac{n-n_B}{n})^2} & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}\le\frac{1}{2} \\ -\frac{(\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A})(\frac{1}{n}(log_2\frac{n_B}{n}+\frac{1}{ln2})-\frac{1}{n}log_2\frac{n-n_B}{n}-\frac{1}{(n-n_B)ln2})}{(\frac{n_B}{n}log_2\frac{n_B}{n}+\frac{n-n_B}{n}log_2\frac{n-n_B}{n})^2} & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2} \end{cases}$
25	$\begin{cases} -\frac{nn_An_{A\bar{B}}}{(nn_A-n_An_B)^2} & if\ \frac{n-n_{A\bar{B}}}{n_A}>\frac{n_B}{n} \\ \frac{n(n_A-n_{A\bar{B}})}{n_An^2{}_B} & otherwise \end{cases}$
26	$\frac{n_A-n_{A\bar{B}}}{nn_Bln2}+\frac{n_{A\bar{B}}}{n(n-n_B)ln2}$
27	$-\frac{1}{n}\sqrt{\frac{n_A-n_{A\bar{B}}}{n}}$
28	$\left(\frac{1}{n-n_A}\right)(log_2(n-n_B)-log_2 n_B)+\left(\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n-n_B-n_{A\bar{B}}}{n-n_A}\right)\left(\frac{-1}{(n-n_B)ln2}-\frac{1}{n_Bln2}\right)$
29	$-\left(\frac{n_A-n_{A\bar{B}}}{2n_B^2}\right)$

30	0
31	$-\frac{n_A - 2n_{A\bar{B}}}{n_{\bar{B}}^2}$
32	$-\frac{n_A}{n^2}$
33	$\frac{n_A - n_{A\bar{B}}}{nn_B ln2} + \frac{n_{A\bar{B}}}{n(n-n_B)ln2} + \frac{1}{n} log_2 \frac{n(n_B - n_A + n_{A\bar{B}})}{(n-n_A)n_B} + \frac{n_B - n_A + n_{A\bar{B}}}{n}\left(\frac{n_A - n_{A\bar{B}}}{n_B(n_B - n_A + n_{A\bar{B}})ln2}\right) - \frac{1}{n} log_2 \frac{n(n - n_B - n_{A\bar{B}})}{(n-n_A)(n-n_B)} - \frac{n - n_B - n_{A\bar{B}}}{n}\left(\frac{(n-n_B) + (n - n_B - n_{A\bar{B}})}{(n-n_B) + (n - n_B - n_{A\bar{B}})ln2}\right)$
34	$\frac{1}{n}$
35	$\frac{-3}{2n} - \frac{2n_{A\bar{B}}}{(n-n_B)^2}$
36	$-\frac{n_A - n_{A\bar{B}}}{n_{\bar{B}}^2}$
37	0
38	$\frac{-1}{n - n_A}$
39	$\frac{(nn_A - n_A n_B - nn_{A\bar{B}})n_B}{(max((n_A - n_{A\bar{B}})(n-n_B), n_B n_{A\bar{B}}))^2}$

Appendix 5

Formula of partial derivative under the parameter $n_{A\bar{B}}$

No	$\frac{\partial Z}{\partial n_{A\bar{B}}}$
1	$-\frac{1}{n_A} log_2 \frac{n(n_A - n_{A\bar{B}})}{n_A n_B} - \frac{1}{(n_A - n_{A\bar{B}})ln2}$
2	$-\frac{1}{n_A}$
3	$-n_A \frac{n - n_B}{n_B n_{A\bar{B}}^2}$
4	$-\frac{1}{2}\left(\frac{1}{n_A} + \frac{1}{n - n_B}\right)$
5	$-\frac{1}{2}\left(\frac{3}{n_A} + \frac{1}{n - n_B}\right)$
6	$-\frac{n}{n_A(n-n_B)}$
7	$-\frac{(n-n_A)n_B}{n_A(n_B - n_A + n_{A\bar{B}})^2}$
8	$\frac{h - k + n_B(n - n_A) - l)(2((n-n_A)(n-n_B) + n_A n_B)}{\left(((n-n_A)(n-n_B) + n_A n_B)(n_B - n_A + 2n_{A\bar{B}})\right)^2}$ $h = -((n_A(n-n_B)) + (n_B((n-n_A))((n - n_B - n_{A\bar{B}}))$ $k = (n_A - n_{A\bar{B}})(n_A - n_{A\bar{B}})(n - n_B - n_{A\bar{B}})(n_A(n-n_B))$ $l = (n_A - n_{A\bar{B}})(n - n_B - n_{A\bar{B}})(n_A(n-n_B) + n_B(n - n_A)$
9	$-\frac{1}{n_A}$

10	$-\frac{4}{n}$
11	$-\frac{n_A(n-n_B)}{nn_{A\bar{B}}^2}$
12	0
13	$-\frac{2}{n_A}$
14	$-\frac{2}{n}$
15	$\frac{II\frac{1}{2}(h)^{-\frac{1}{2}}\left(k+\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(\left(\frac{1}{n_B}(log_2\frac{n_{A\bar{B}}}{n_B}+\frac{1}{ln2})-\frac{1}{n_B}(log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}+\frac{1}{ln2})\right)\right)\right)\right)}{2\sqrt{II\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)\right)^{\frac{1}{2}}}}$ $h=1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)$ $k=\left(\left(-\frac{1}{n_A}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}+\frac{1}{ln2})+\frac{1}{n_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{1}{ln2})\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\right)\right)$
16	$\frac{II_n'(h)^{\frac{1}{4}}+II\frac{1}{4}(h)^{-\frac{3}{4}}(k)+\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2 l\right)}{2\sqrt{II\left(1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)\right)^{\frac{1}{4}}}}$ $h=1-\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)^2\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)$ $k=-2\left(-\frac{n_A-n_{A\bar{B}}}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{n_{A\bar{B}}}{n_A}log_2\frac{n_{A\bar{B}}}{n_A}\right)\left(\frac{1}{n_A}(log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{1}{ln2})-\frac{1}{n_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{1}{ln2})\right)\left(1-\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)^2\right)$ $l=\left(-2\left(-\frac{n_{A\bar{B}}}{n_B}log_2\frac{n_{A\bar{B}}}{n_B}-\frac{n-n_B-n_{A\bar{B}}}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}\right)\left(-\frac{1}{n_B}(log_2\frac{n_{A\bar{B}}}{n_B}+\frac{1}{ln2})+\frac{1}{n_B}log_2\frac{n-n_B-n_{A\bar{B}}}{n_B}-\frac{1}{ln2})\right)\right)$
17	$-\frac{n_A}{(n_A-n_{A\bar{B}})^2}$
18	$-\frac{1}{n}$
19	$\frac{-2(n_A-n_{A\bar{B}})+2n_{A\bar{B}}}{nn_A}+\frac{2(n_B-n_A+n_{A\bar{B}})-2(n-n_B-n_{A\bar{B}})}{n(n-n_A)}$
20	$\frac{\alpha}{\beta}\ where$ $\alpha=\left(max\left(\frac{-1}{n},\frac{1}{n}\right)+max\left(\frac{1}{n},\frac{-1}{n}\right)+max\left(\frac{-1}{n},\frac{1}{n}\right)+max\left(\frac{1}{n},\frac{-1}{n}\right)\right)$ $\beta=2-max\left(\frac{n_A}{n},\frac{n-n_A}{n}\right)-\left(\frac{n_B}{n},\frac{n-n_B}{n}\right)$

21	$\frac{1}{\sqrt{\frac{n_A(n-n_B)}{n}}}$
22	$-\sum_{k=max\,(1,n_A-n_B)}^{n_{A\bar{B}}-1} \frac{n_B!}{((n_A-k)!\,(n_B-n_A+k)!}\frac{(n-n_B)!}{k!\,(n-n_B-k)!}\frac{(n_A)!\,(n-n_A)!}{n!}$
23	0
24	$\begin{cases} -\frac{1}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{1}{(n_A-n_{A\bar{B}})ln2}+\frac{1}{n_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{1}{ln2}) & if(\frac{n_B}{n}\leq\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2} \\ 0 & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}\leq\frac{1}{2} \\ -\frac{-\frac{1}{n_A}log_2\frac{n_A-n_{A\bar{B}}}{n_A}-\frac{1}{(n_A-n_{A\bar{B}})ln2}+\frac{1}{n_A}(log_2\frac{n_{A\bar{B}}}{n_A}+\frac{1}{ln2})}{\frac{n_B}{n}log_2\frac{n_B}{n}+\frac{n-n_B}{n}log_2\frac{n-n_B}{n}} & if(\frac{n_B}{n}>\frac{1}{2}\ and\ \frac{n_A-n_{A\bar{B}}}{n_A}>\frac{1}{2} \end{cases}$
25	$\begin{cases} -\frac{n}{n_A(n-n_B)} & if\ \frac{n-n_{A\bar{B}}}{n_A}>\frac{n_B}{n} \\ -\frac{n}{n_An_B} & otherwise \end{cases}$
26	$-\frac{1}{n}(log_2\frac{n(n_A-n_{A\bar{B}})}{n_An_B}+\frac{1}{ln2})\ +\frac{1}{n}(log_2\frac{nn_{A\bar{B}}}{n_A(n-n_B)}+\frac{1}{ln2})$
27	$\frac{-\frac{1}{n}}{2\sqrt{\frac{n_A-n_{A\bar{B}}}{n}}}\left(\frac{n-n_B}{n}-\frac{n_{A\bar{B}}}{n_A}\right)-\frac{1}{n_A}\sqrt{\frac{n_A-n_{A\bar{B}}}{n}}$
28	$-\left(\frac{1}{n_A}-\frac{1}{n-n_A}\right)(log_2(n-n_B)-log_2 n_B)$
29	$-\frac{1}{2}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)$
30	$-\frac{1}{n_A+2}$
31	$-\frac{2}{n_B}$
32	$-\frac{1}{n_A}$
33	$-\frac{1}{n}log_2\frac{n(n-n_{A\bar{B}})}{n_An_B}+\frac{n_A-n_{A\bar{B}}}{(n-n_{A\bar{B}})ln2}$ $+\frac{1}{n}log_2\frac{nn_{A\bar{B}}}{n_A(n-n_B)}+\frac{1}{nln2}+\frac{1}{n}log_2\frac{n(n_B-n_A+n_{A\bar{B}})}{(n-n_A)n_B}$ $+\frac{n_B-n_A+n_{A\bar{B}}}{n(n_B-n_A+n_{A\bar{B}})ln2}-\frac{1}{n}log_2\frac{n(n-n_B-n_{A\bar{B}})}{(n-n_A)(n-n_B)}-\frac{n-n_B-n_{A\bar{B}}}{(n-n_B-n_{A\bar{B}})ln2}$
34	0
35	$-\left(\frac{3}{2n_A}+\frac{2}{n-n_B}\right)$
36	$-\frac{1}{n_B}$
37	$-\frac{n_A}{n_{A\bar{B}}^2}$
38	$\frac{-1}{n-n_A}$
39	$\frac{-n\,max\left((n_A-n_{A\bar{B}})(n-n_B),n_Bn_{A\bar{B}}\right)-(nn_A-n_An_B-nn_{A\bar{B}})max\,(-(n-n_B),n_B)}{(max\left((n_A-n_{A\bar{B}})(n-n_B),n_Bn_{A\bar{B}}\right))^2}$

MEDIDAS ALTERNATIVAS DE LOS GRADOS DE ADHESIÓN DE INDIVIDUOS A LA IMPLICACIÓN Y SIMILARIDAD PARA VARIABLES MODALES

Larisa ZAMORA[1] y Pablo GREGORI[2]

MESURES ALTERNATIVES DES DEGRES D'ADHESION DES INDIVIDUS A L'IMPLICATION ET SIMILARITE POUR DES VARIABLES MODALES

ALTERNATIVE MEASUREMENTS OF AGREEMENT LEVEL OF INDIVIDUALS TO IMPLICATION AND SIMILARITY FOR CATEGORICAL VARIABLES

RESUMEN

Se ofrecen tres propuestas alternativas a las introducidas en Bailleul y Gras (1994), para medir el grado de adhesión de los individuos a las implicaciones y similaridades, que afectan, a su vez, al cálculo de su tipicalidad y contribución a la formación de las clases del árbol jerárquico de cohesión y similaridad. Dos propuestas atienden principalmente al efecto que los individuos muestreados producen sobre los índices calculados, mientras que la tercera se basa en una muestra auxiliar.

Palabras clave : *tipicalidad, contribución, similaridad, cohesión, análisis implicativo.*

RÉSUMÉ

On présente trois propositions alternatives à celles introduites à Bailleul et Gras (1994), afin de mesurer le degré d'adhésion des individus aux implications et similarités, qui modifient, à son tour, le calcul de leur typicalité et contribution à la formation des classes de l'arbre hiérarchique de cohésion et similarité. Deux propositions partent de l'effet que les individus échantillonnés produisent sur les indices calculés, tandis que la troisième est basée sur un échantillon auxiliaire.

Mots-clés : *typicalité, contribution, similarité, cohésion, analyse implicative.*

ABSTRACT

We propose some alternative definitions to the one given in Bailleul and Gras (1994), for measuring the agreement level of individuals to implications and similarities, which alter, at the same time, the computation of their typicality and contribution to the formation of classes in the hierarchical tree. Some proposals are based on the effect of the sampled individuals on the computed index, as well as one of them starts from an auxiliary sample.

Keywords : *typicality, contribution, similarity, cohesion, implicative analysis.*

[1] Departamento de Matemática. Facultad de Matemática y Computación, Universidad de Oriente, Santiago de Cuba, Cuba, larisa@csd.uo.edu.cu

[2] Departamento de Matemáticas. Instituto de Matemáticas y Aplicaciones de Castellón, Universitat Jaume I de Castellón, Campus Riu Sec E-12071, Spain, gregori@uji.es

1 Introducción

El Análisis Estadístico Implicativo (ASI) recopila una serie de procedimientos de la estadística multivariante cuyo objetivo es estructurar un grupo de variables muestreadas sobre un conjunto de individuos. Nació en el contexto de la Didáctica de las Matemáticas, en la tesis doctoral Gras (1979), en su vertiente implicativa. Previamente, Lerman (1970) introdujo un análisis de similaridad, donde se usaban la independencia estadística y la probabilidad para medir lo próximas que estaban dos variables binarias. Y en base a esa forma original de medir la similaridad (diferente de las otras similaridades y distancias utilizadas en los análisis *cluster*, o de conglomerados), se creaba la estructura jerárquica de las variables.Trasponer esta situación simétrica, de la similaridad entre variables, a una no simétrica, como es la relación de implicación, donde hipótesis y tesis no son intercambiables, fue el objetivo de la tesis Gras (1979). En ella se definía la intensidad de una (cuasi-)implicación entre variables binarias, con la filosofía de Lerman (independencia + probabilidad), y se aplicaba a un estudio experimental sobre adquisiciones cognitivas. Aquello supuso el nacimiento del ASI que, por completitud, englobó al análisis de similaridades de Lerman en su repertorio.

A partir de entonces, se fue ampliando el rango de actuación de los procedimientos: nacido en el seno de las variables binarias (Lerman et al., 1981ab), que describían la presencia o ausencia de ciertos rasgos de interés didáctico, se ampliaron las definiciones a las variables modales, frecuenciales y de intervalo (Bailleul y Gras, 1994, Lagrange, 1998). De igual manera, se desarrolló la profundidad de dichos análisis. Por una parte, las intensidades de las implicaciones entre parejas de variables permitían representar un grafo dirigido en el que los investigadores pueden interpretar las relaciones implicativas. Por otro lado, tomando como ejemplo la estructura jerárquica a la que daba lugar el análisis de similaridad (en forma de árbol, en el que cada variable es una clase inicial, y a cada nivel de la jerarquía se reúnen las dos clases más similares del nivel precedente), se intentó encontrar una estructura jerárquica entre las implicaciones. Al igual que en el mundo matemático un teorema implica ciertos corolarios, se buscó la forma de crear esa estructura donde unas implicaciones « conducen » a otras.

Para ello, se definió la cohesión de una pareja (ordenada) de variables, como una transformación concreta de su intensidad implicativa. Y la de una clase (ordenada) de variables, como la media geométrica de las cohesiones de todas las parejas (ordenadas) de variables. Finalmente, la implicación de una clase de variables sobre otra, se definió como una fusión particular entre intensidades implicativas y las cohesiones de las clases. Con este último índice se pudo reproducir una estructura jerárquica, como en las similaridades, pero de relaciones no simétricas, que ordena las clases « implicativas ». A este análisis se le llama análisis de cohesiones, y su aplicación a datos reales conduce a conclusiones como « esta relación implicativa entre este grupo de variables *favorece* esta otra relación implicativa entre este otro grupo de variables ».

El ASI también enriquece la profundidad de estos dos análisis jerárquicos:

- destacando algunos de los niveles de la jerarquía (nodos significativos),
- destacando algunos individuos de la muestra, por el papel que juegan en la formación de cada nuevo nivel de la jerarquía.

Se llama nodo (o nudo o clase o nivel) significativo a aquel que contiene una serie de variables que guardan una especial consonancia respectiva, comparada con la de los

niveles de la jerarquía anterior y posterior. El investigador suele atender a alguno de estos niveles significativos a la hora de las interpretaciones, ya que sus variables constituyentes están más fuertemente ligadas que en otros niveles del árbol jerárquico.

En cada nivel de la jerarquía (sea un nodo significativo o no), se clasifican los individuos de la muestra. Por una parte se clasifican en torno a la « contribución » para la formación de ese nivel (o clase). Es decir, que para cada individuo se calcula un valor que indica lo mucho que sus datos han favorecido la formación de dicha clase. Por ejemplo, si una clase consiste en la similaridad de dos variables, un individuo que tiene valores distintos en dichas variables, no facilita en absoluto que esas dos variables se hayan reunido en una clase, como similaridad. Sin embargo, otro individuo cuyos datos sobre ambas variables coinciden, es quien más fuertemente favorece que dichas variables se hayan reunido en una clase. El reconocimiento de dichos individuos también puede ayudar a la interpretación de resultados, incluso si se usan individuos « ficticios » que representen perfiles fácilmente identificables por el investigador.

La otra clasificación que se realiza sobre los individuos se formaliza en torno a la sintonía que hay entre la clase formada y los datos del individuo sobre las variables de esa clase: si una clase tiene un índice de similaridad medio, un individuo « típico » no será aquel cuyos datos sean muy coincidentes para todas las variables de dicha clase (ya que esa idea lleva al concepto descrito anteriormente). Sería muy típico, más bien, aquel individuo cuyos datos coinciden en las variables muy similares y no tanto en variables menos similares (de ahí el nombre tipicalidad).

Estas clasificaciones se realizan a partir de comparar los índices de las clases con una versión « individualizada » de los mismos (a la que llaman « respeto » o « grado de adhesión »), y el ordenamiento posterior de los individuos, con los pesos (de contribución o tipicalidad) respectivos. En este trabajo repasamos las propuestas de la literatura y aportamos nuevos puntos de vista para su consideración.

2 Definiciones, notaciones y datos de ejemplo

En el ASI, se denota por V al conjunto de variables, entre las que se encuentran a, b, etc., y por E, al conjunto de individuos, entre los que se encuentran x, y, etc., de modo que a(x) es el dato del individuo x sobre la variable a, etc. Las variables originarias del ASI fueron binarias, pero su aplicación se amplió a más tipos de variables. En este trabajo nos centramos en las variables modales ordinales, con n categorías cada una, como las escalas Likert. Para su tratamiento por el ASI, estas variables deben recodificar sus categorías a los valores $0, \frac{1}{m-1}, \frac{2}{m-1}, \ldots, \frac{m-2}{m-1}$ y 1. Por ejemplo, para $m=5$ categorías, se tendrían las recodificaciones 0, 0.25, 0.5, 0.75 y 1.

En este trabajo vamos a usar un conjunto de datos de preferencias de estilos musicales (ver anexo), donde un grupo de individuos valora su nivel de preferencia sobre 15 estilos musicales. El ASI nos permite estructurar dichos estilos musicales (en base a las respuestas de los individuos) de dos modos: por similitud y por implicación. Por ejemplo, si los estilos HIP y PUN resultan muy similares para el ASI, no significará que sean parecidos a nivel armónico, lógicamente, sino que ambos estilos agrupan a audiencias relativamente cercanas. De igual modo, si la intensidad de implicación HIP → PUN es elevada, el número de contraejemplos, es decir, de individuos a los que

les gusta HIP y no les gusta PUN, es extraordinariamente bajo, para lo que ocurriría si esos estilos gustaran al público de manera independiente.

Aunque este trabajo aborda una temática que afecta por igual al análisis de similaridades y de cohesiones, debido a su estrecho paralelismo formal, escribiremos únicamente el caso de cohesiones, como es habitual en la literatura, y por ser ligeramente más abstracto, en razón a su carácter no simétrico. La intensidad de implicación entre las variables binarias a y b se denota por $\varphi(a,\bar{b})$ y mide cuán bajo es el número de contraejemplos a la regla $a \rightarrow b$, al compararlo con una situación de aleatoriedad total (independencia estadística) entre a y b. Se calcula por:

$$\varphi(a,\bar{b}) = P(N_{a\bar{b}} > n_{a\bar{b}})$$

donde $n_{a\bar{b}}$ representa el número de contraejemplos a la regla $a \rightarrow b$ observados en la muestra, y $N_{a\bar{b}}$ es una variable aleatoria que calcula el número de contraejemplos bajo independencia entre *a* y *b*, y su media y varianza depende de los efectivos *n* (tamaño de la muestra), n_a (número de individuos que cumplen *a*) y $n_{\bar{b}}$ (número de individuos que no cumplen *b*). Se puede calcular bajo varios modelos de probabilidad, a escoger según la forma del muestreo (Lerman *et al.*, 1981).

La intensidad de implicación generaliza su definición para variables frecuenciales (con valores en el intervalo [0,1]) en Gras y Larher (1992). Las variables modales, tras recodificar sus categorías a valores numéricos equidistantes en el intervalo [0,1], se pueden tratar como variables frecuenciales. Sin pérdida de generalidad, seguiremos llamando *a* y *b* a las variables recodificadas. La intensidad de implicación entre variables modales y frecuenciales se calcula de forma que mantiene coherencia con la definición sobre variables binarias. Basta con definir el nuevo número de contraejemplos como $n_{a\bar{b}} = \sum_{x\in E} a(x)(1-b(x))$, y los nuevos efectivos

$n_a = \sum_{x\in E} a(x)$ y $n_{\bar{b}} = \sum_{x\in E} (1-b(x))$, que coinciden en el caso binario.

Con las intensidades de implicación se puede representar el llamado grafo implicativo, en el que el investigador visualiza fácilmente las implicaciones que resultan con intensidad alta. En la Figura 1 se muestra el resultado para los datos de la música.

Por otro lado se tiene el análisis jerárquico de similaridades de Lerman y el de cohesiones de Gras. Como su estructura es muy paralela, se describe el de cohesiones por ser algo más abstracto.

Inicialmente, podemos decir que cada variable forma una clase por sí misma.

En el primer nivel del proceso jerárquico, se va a formar una nueva clase ordenada, llamémosla $C_1 = (C_{11}, C_{12})$, como la unión de las dos clases del nivel anterior (es decir, las variables) con mayor intensidad de implicación. Aunque en este nivel se torna trivial el concepto que se explica a continuación, se introduce aquí para facilitar su comprensión en los siguientes niveles:

– Se llama par genérico de la clase $C_1 = (C_{11}, C_{12})$ a la pareja de variables que maximiza la intensidad de implicación, de entre todas las parejas posibles, con una variable en C_{11} y otra en C_{12}. En este nivel el concepto es trivial pues sólo

hay una variable en C_{11} y otra en C_{12}, y por tanto dichas variables forman el par genérico.

- Se llama intensidad de implicación genérica de C_1, al valor de la intensidad implicativa para el par genérico.
- Se valora el « respeto » o « grado de adhesión » de los individuos *x* de *E* frente a dicha implicación genérica, de la siguiente forma :
  - Mínimo respeto: $\varphi_x(a,\bar{b})=0$, si *a(x)=1* y *b(x)=0*, por ser *x* un contraejemplo genuino de la regla $a \rightarrow b$.
  - Máximo respeto: $\varphi_x(a,\bar{b})=1$, si *b(x)=1*, por ser *x* un ejemplo que « respeta » la regla $a \rightarrow b$.
  - Respeto neutro: $\varphi_x(a,\bar{b})=0.5$, si *a(x)=b(x)=0*, por ser *x* un individuo que ni la refuerza ni la contradice.
- Estas valoraciones, del respeto de los individuos sobre la implicación genérica de la clase C_1, sirven para clasificarlos en torno a dos nuevos conceptos: la tipicalidad y la contribución. Se calcula entonces una lista con el valor de tipicalidad para cada individuo *x* respecto de la clase C_1, y otra lista similar con el valor de contribución. No describimos su formulación, pero indicamos su utilidad para enriquecer la información proporcionada por el análisis, pudiendo ser productiva para el investigador.

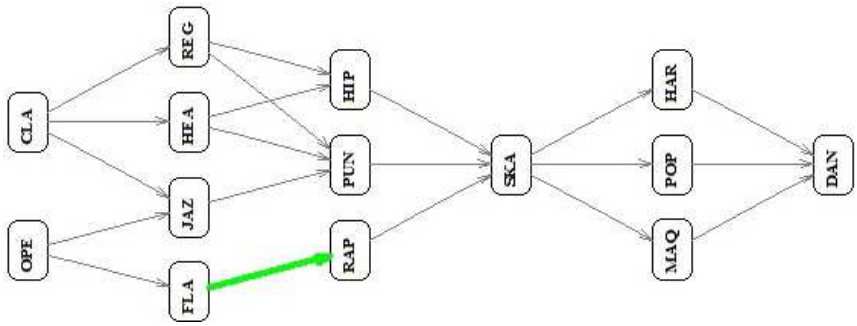


Figura 1 – Grafo implicativo (girado por conveniencia de edición), que expresa una estructura de relaciones de implicación. En este caso las intesidades resultan de muy bajo nivel, pero se representan a modo de ilustración. La única implicación significativa es FLA → RAP.

Para continuar con el nivel 2 (y posteriores) del algoritmo, se precisa definir la implicación entre grupos (o clases) ordenados de variables. Esta definición se proporciona en varias etapas:

1. Cohesión de una pareja (ordenada) de variables *(a,b)*, denotada por *c(a,b)* : se define como una transformación monótona de la intensidad de implicación :
   1. cuyo valor es nulo si $\varphi(a,\bar{b})<0.5$, por ser una implicación demasiado débil,
   2. y crece de forma estrictamente monótona con $\varphi(a,\bar{b})\geq 0.5$. La forma elegida fue $c(a,b)=\sqrt{1-E(\varphi(a,\bar{b}))^2}$, donde se usa la función de entropía $E(p):=-p\log_2 p-(1-p)\log_2(1-p)$.

2. Cohesión de una clase (ordenada) de variables $A=(a_1,\ldots, a_r)$, denotada por $c(A)$: se define como la media geométrica de las cohesiones de todos los pares de variables (a_i,a_j) donde $1 \leq i < j \leq r$.
3. Implicación de la clase (ordenada) A sobre la clase (ordenada) B, denotada por $\psi(A,B)$: se define teniendo en cuenta la intensidad máxima entre un elemento de A y otro de B, además de las cohesiones de ambas clases. En concreto se definió como $\psi(A,B) = \left[\max_{\substack{a\in A \\ b\in B}} \varphi(a,\bar{b}) \right]^{rs} \sqrt{c(A)c(B)}$, donde r y s son las cantidades de variables en las respectivas clases A y B.

Así pues, en el segundo nivel de la jerarquía (y posteriores) se consideran todas las clases del nivel anterior (quitando las dos que se reunieron, y la nueva que se formó con ellas), y se reúnen las dos clases que ahora presenten la mayor implicación $\psi(A,B)$. Denotemos por $C_2 = (C_{21}, C_{22})$ la nueva clase. Entonces:

- los conceptos de par genérico e intensidad de implicación genérica conllevan buscar esa implicación máxima entre parejas de variables, la primera en C_{21} y la segunda en C_{22}.
- C_{21} y C_{22}, por venir del nivel anterior, deben tener identificados (si no son variables sencillas) sus respectivos pares genéricos, intensidades de implicación genéricas, y clasificados a todos los individuos de la muestra según el respeto sobre sí mismas, en el nivel en que se formó cada una. Esta información se debe guardar a cada nivel del algoritmo.
- Para la nueva clase C_2, se consideran todas las implicaciones genéricas de clases englobadas por C_2, definiendo el llamado vector potencia implicativa de C_2.
- Los conceptos de tipicalidad y contribución de los individuos x sobre la nueva clase C_2 se calcularán, como un promedio, a partir de las valoraciones del respeto de los individuos sobre sólo las implicaciones genéricas de las clases que engloba C_2. Y por tanto retoma los datos calculados en aquellos niveles anteriores, que en esta iteración queden contenidos en la clase C_2.

Estando, los conceptos de tipicalidad y contribución, basados en las valoraciones sobre el respeto de las implicaciones genéricas (solo esas) que contiene la clase, nuestro trabajo pretende aportar nuevos puntos de vista a las filosofías presentadas por Ratsimba-Rajohn (1991), sobre variables binarias, y Bailleul (1994), sobre variables modales.

A modo de ilustración, en la Figura 2 se muestran los resultados respectivos de aplicar los análisis de similaridades y cohesiones a los datos de las preferencias musicales.

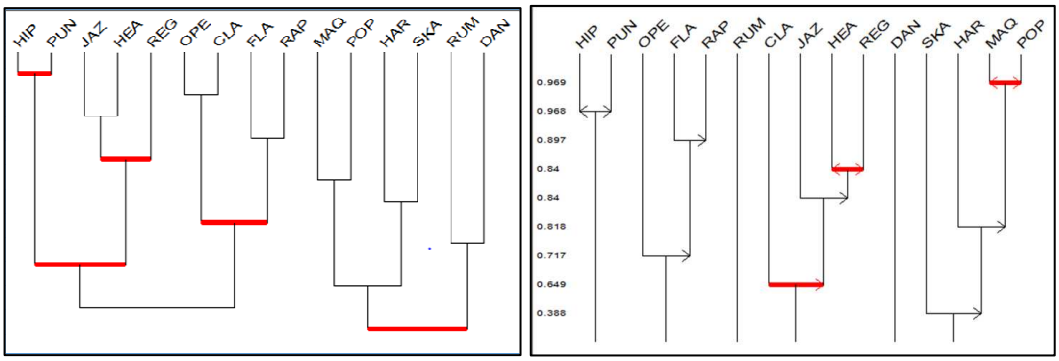


Figura 2 – Árboles jerárquicos, de similaridad (izquierda) y cohesión (derecha), para los datos de música. Los nodos en trazo grueso (y rojo) son los significativos, y se suelen escoger para la interpretación de resultados.

Como se puede ver en las figuras, en los árboles jerárquicos se señalan unos niveles especiales, los nodos significativos, que no trataremos en este trabajo, pero que indican nodos cuyas variables están más ligadas que las del nodo precedente y el posterior, motivo por el cual, el investigador puede indagar con más atención en busca de una interpretación práctica de los resultados, además de centrar en ellos la interpretación de los individuos en cuanto a tipicalidad y contribución.

3 Grado de adhesión de un individuo a una implicación en variable binaria y modal

Como se indicaba en la introducción, la base para las medidas de tipicalidad y contribución de los individuos a las clases formadas, es el « respeto » o « grado de adhesión » de los mismos a las implicaciones genéricas de la clase en cuestión.

En el caso binario, Ratsimba-Rajohn (1991) estableció el « respeto » de un individuo x hacia una implicación $a \rightarrow b$, tal y como hemos descrito en el primer nivel del algoritmo jerárquico de la sección anterior.

En la Tabla 1 presentamos dos cuadros: uno en el que ordenamos a los cuatro tipos de individuos (según sus repuestas a a y b, respectivamente) por orden de menor a mayor respeto, y otro en el que colocamos su « valor de respeto ».

Tabla 1 – A la izquierda, los 4 individuos posibles se ordenan de menor a mayor nivel de « respeto » por la implicacion $a \rightarrow b$. A la derecha, sus valores de « respeto » por la regla $a \rightarrow b$, según la filosofía de Ratsimba-Rajohn (1991).

Ordenamiento		
a / b	0	1
0	2	3
1	1	3

Valor de respeto		
a / b	0	1
0	0.5	1
1	0	1

En Bailleul y Gras (1994) se plantea la ambigüedad de los dos individuos que presentan el mismo orden (rango), tratándose de dos comportamientos distintos, además de plantear la situación paralela en el caso de variables modales, con n categorías cada una de ellas. Para resolverla, propone realizar dos pasos:

1. Ordenar el conjunto de parejas de respuestas posibles,
2. Vincular el valor que mide el « respeto » o la contradicción de la implicación, a la propia intensidad de dicha implicación.

Tomando la noción de implicación, extendida a variables modales, el valor paralelo al número de contraejemplos, es la suma $a(x)(1 - b(x))$ para todos los individuos x (ver definición en sección anterior).

Por ese motivo, los individuos x para los que $a(x) > b(x)$, respetan « menos » la regla $a \rightarrow b$, siendo el menos respetuoso de todos, aquel individuo x tal que $a(x)=1$ y $b(x)=0$. Y por contra, los individuos x para los que $a(x) \leq b(x)$ respetan « más » la regla $a \rightarrow b$, siendo el más respetuoso de todos, aquel individuo x tal que $a(x)=1$ y $b(x)=1$. Según la intuición y justificación, Bailleul y Gras (1994) establecen un ordenamiento, compatible con el de Ratsimba-Rajohn (1991) y que matiza el caso coincidente. En la Tabla 2 mostramos ejemplos de la filosofía de ordenamiento del « grado de adhesion » de Bailleul (1994) para distinto número de categorías de las variables modales.

Tabla 2 – Ordenamientos según el « grado de adhesión », de menor a mayor, de los individuos a la regla $a \rightarrow b$, para el caso de variables modales con $n=2,3$ y 5 categorías, y propuesto por Bailleul y Gras (1994).

a / b	0	1
0	2	3
1	1	4

a / b	0	0.5	1
0	4	5	6
0.5	3	7	8
1	1	2	9

a / b	0	0.25	0.5	0.75	1
0	11	12	13	14	15
0.25	10	16	17	18	19
0.5	6	9	20	21	22
0.75	3	5	8	23	24
1	1	2	4	7	25

Por no abundar en fórmulas innecesarias, indicamos que el ordenamiento se inicia en la pareja de modalidades que ocupa la casilla inferior izquierda (que sería la que menos respeta la implicación), continúa ascendiendo por las paralelas a la antidiagonal (desde abajo hacia arriba), sin llegar a la propia antidiagonal. Y a partir de entonces se ordenan los restantes casos por filas, de izquierda a derecha, y desde la fila superior hacia abajo, hasta llegar a la esquina inferior derecha (que es la que más respeta la implicación).

Una vez establecido el ordenamiento compatible con la filosofía de respeto o grado de adhesión de cada pareja de categorías a la implicación, se calcula un valor numérico para medir dicho grado de respeto, que se calcula a partir del ordenamiento, y del valor del índice de la implicación:

- El ordenamiento realizado se cambia de escala al intervalo [0,1]: restando 1 unidad de las posiciones y dividiendo todas ellas por la máxima resultante (m^2-1).
- Se usa el nuevo valor de la escala como factor multiplicativo de la implicación en cuestión, $\varphi(a,\bar{b})$.

En la Tabla 3 se muestran las tres etapas de la ideas de Bailleul y Gras (1994), para el caso de variables binarias, donde se aprecia la diferencia respecto de la propuesta de Ratsimba-Rajohn (1991), mostrada en la Tabla 1.

Tabla 3 – Proceso de cálculo del grado de adhesión de los individuos a la regla $a \rightarrow b$, para el caso de variables binarias, propuesto por Bailleul y Gras (1994).

Ordenamiento		
a / b	0	1
0	2	3
1	1	4

→

Orden re-escalado		
a / b	0	1
0	1/3	2/3
1	0	1

→

Grado de adhesión		
a / b	0	1
0	$\frac{1}{3}\varphi(a,\bar{b})$	$\frac{2}{3}\varphi(a,\bar{b})$
1	0	$\frac{1}{3}\varphi(a,\bar{b})$

Bailleul y Gras (1994) definen, a partir de este grado de adhesión de cada individuo x a una implicación, las fórmulas de contribución y tipicalidad de cada individuo x a cada nueva clase formada C, usando las implicaciones genéricas de dicha clase C, como se comentaba en la descripción del algoritmo jerárquico de la sección anterior.

4 Propuestas alternativas al grado de adhesión

Debido a que en la literatura no se habían encontrado expresiones para la determinación del índice de similaridad, los nodos significativos, la tipicalidad y contribución de los individuos a las clases para el caso de variables modales, en Zamora (2015) se propone una expresión para calcular el índice de similaridad para estas variables, así como para la determinación de la tipicalidad y contribución de los individuos a las clases que se forman durante la construcción del árbol de similaridad, basándose en los trabajos de Bailleul y Gras (1994) y Lagrange (1998), lo que permitió transponer el concepto de grado de adhesión de los individuos a la similaridad entre dos variables modales. Esto conllevó a los autores del presente trabajo a reformular el papel de este grado de adhesión incluso para la implicación, más consolidada en la literatura.

Por una parte, las propuestas de Ratsimba-Rajohn (1991) y Bailleul y Gras (1994) vienen debidamente justificadas por razonamientos basados en la definición de *regla* y *contraejemplo*. Sin embargo, el concepto de respeto o grado de adhesión a una regla, se puede cuestionar racionalmente desde otros puntos de vista. Por ejemplo:

1. ¿Qué efecto provocaría un nuevo individuo sobre la intensidad de una implicación ya valorada en una muestra? Si se añade, y provoca un aumento de la intensidad, se trataría de un individuo que la respeta (o se adhiere). Si, por el contrario, provoca una disminución de la intensidad, se trataría de un individuo que no la respeta.
2. ¿Qué efecto han provocado los individuos de la muestra sobre la intensidad de una implicación ya valorada? Si se extrae, y provoca un aumento de la intensidad, se trata de un individuo que no la respeta (pues la ha lastrado al incluirse en la muestra). Si, por el contrario, disminuye la intensidad, se trata de un individuo que la respeta (pues su ausencia debilita la implicación).
3. Consideremos una muestra ajena especial, en la que hay un individuo que representa a cada pareja de categorías de las variables *a* y *b*. Para esa muestra, las variables *a* y *b* serían, por lógica, estadísticamente independientes. Calcular

el incremento de cada individuo al índice también informa de su respeto a la implicación.

En los siguientes apartados desarrollamos estas tres ideas y reflexionamos sobre sus posibilidades, usando datos simulados sobre gustos musicales, bajo distintos números de categorías.

4.1 Propuesta 1 : añadir a la muestra un individuo ficticio de cada tipo

Para la muestra considerada, en las variables *a* y *b*, se procede a añadir cada vez un individuo con cada pareja de categorías posible. Se calcula el incremento que dicho individuo produce en el valor de la intensidad de implicación (de no estar a estar), y al final se ordenan estos individuos por el incremento producido en la intensidad de implicación, desde el que da un incremento más negativo (y sería el primero en el ordenamiento), hasta el que produce un incremento más positivo (que sería el último en el ordenamiento). Esta sería la primera alternativa al ordenamiento dado por Bailleul y Gras (1994).

Para visualizar los efectos de esta propuesta, usamos los datos binarios de la música, y consideramos dos implicaciones, que mostramos en la Tabla 4.

Tabla 4 – Aplicación de la propuesta 1 con 2 modalidades para dos implicaciones de la muestra: datos (izquierda), valores de intensidad de implicación al añadir cada individuo posible (centro), y ordenamiento según el incremento provocado en la intensidad de implicación. Las intensidades reales, sin individuos añadidos, son $\varphi(HIP, \overline{O}\,\overline{P}\,\overline{E}) = 0.3970013$ y $\varphi(\mathrm{HIP}, \overline{\mathrm{MAQ}}) = 0.5536308$, respectivamente.

Efectivos		
HIP/OPE	0	1
0	6	3
1	9	2

→

Intensidad al añadir 1 individuo		
HIP/OPE	0	1
0	0.4153317	0.3417477
1	0.3884028	0.4418189

→

Ordenamiento		
HIP/OPE	0	1
0	3	1
1	2	4

Efectivos		
HIP/MAQ	0	1
0	2	7
1	2	9

→

Intensidad al añadir 1 individuo		
HIP/MAQ	0	1
0	0.6489726	0.5262412
1	0.4663390	0.5749585

→

Ordenamiento		
HIP/MAQ	0	1
0	4	2
1	1	3

Se puede ver en la Tabla 4 que los ordenamientos son distintos, y que los individuos *extremos* (según los razonamientos a priori de Ratsimba-Rajohn y Bailleul) pueden dejar de serlo, al nivel práctico de “elevar” o “rebajar” la intensidad de la implicación.

Para visualizar los ordenamientos en el caso de más modalidades, simulamos aleatoriamente (ver Anexo) 1000 individuos en 2 variables, con 5 modalidades por variable, mostrando el resultado de la propuesta en la Tabla 5.

Tabla 5 – Aplicación de la propuesta 1 con 5 modalidades para una implicación de una muestra simulada con 1000 individuos y 5 modalidades por variable: valores de intensidad de implicación al añadir cada individuo posible (izquierda), y ordenamiento según el incremento provocado en la intensidad de implicación. La intensidad real, sin individuos añadidos, es $\varphi(V1,\overline{V}\,\overline{2}) = 0.3923378$. Los valores están redondeados por cuestión de espacio, pero no hay coincidencias (ejecutar código R de anexo para ver valores con más precisión).

Intensidad al añadir 1 individuo					
V1/V2	0	0.25	0.5	0.75	1
0	0.401	0.397	0.392	0.388	0.384
0.25	0.396	0.394	0.392	0.390	0.388
0.5	0.392	0.392	0.392	0.392	0.393
0.75	0.388	0.390	0.392	0.395	0.397
1	0.384	0.388	0.392	0.397	0.401

→

Ordenamiento					
V1/V2	0	0.25	0.5	0.75	1
0	24	21	14	5	2
0.25	20	18	12	8	6
0.5	9	10	11	16	17
0.75	3	7	13	19	23
1	1	4	15	22	25

Usando otras simulaciones se llega a ordenamientos ligeramente distintos. Sorprende en este caso el individuo de datos (V1=0;V2=1), que tiene un ordenamiento muy distinto al que le otorga el enfoque de Bailleul y Gras (1994).

En resumen, se puede concluir que, en el caso general, el ordenamiento de las parejas de modalidades no siempre es el mismo, y depende de los valores que tomen las variables en la muestra original, más cuanto menor sea el tamaño de la muestra.

4.2 Propuesta 2 : suprimir de la muestra un individuo de cada tipo

En este caso, se procede a suprimir de la muestra un individuo. De igual manera que en la primera propuesta, se ordenan estos individuos por el incremento producido en la intensidad de implicación (de no estar a estar), y lo establecemos como ordenamiento alternativo al dado por Bailleul y Gras (1994). Aunque este algoritmo recorre aparentemente toda la muestra de individuos, esencialmente sólo precisa calcularse para cada cruce de categorías de *a* y *b* que estén presentes en la muestra, y luego atribuir su valor a cada individuo de la muestra que corresponda. Es cierto que con esta propuesta no atribuimos orden a los individuos que no están presentes en la muestra, aunque no causa problemas por no ser necesario su grado de adhesión para los cálculos posteriores de tipicalidad y contribución. Para visualizar los efectos de esta propuesta, usamos los datos binarios de la música, y consideramos dos implicaciones, que mostramos en la Tabla 6. Se puede ver en la Tabla 6 que los ordenamientos que provoca esta propuesta coinciden con los de la propuesta 1, a pesar de variar los valores de las intensidades de implicación. Es presumible que haya datos para los que se encuentren diferencias entre las propuestas 1 y 2, aunque se supone serán muy ligeras, menores cuando el tamaño de la muestra aumente, pues los parámetros muestrales (media y varianza) serán más robustos frente al cambio de añadir o suprimir un individuo.

Tabla 6 – Aplicación de la propuesta 2 con 2 modalidades para dos implicaciones de la muestra: datos (izquierda), valores de intensidad de implicación al suprimir cada individuo posible (centro), y ordenamiento según el incremento provocado en la intensidad de implicación. Las intensidades reales, sin individuos añadidos, son $\varphi(HIP, \overline{O}\,\overline{P}\,\overline{E}) = 0.3970013$ y $\varphi(\mathrm{HIP}, \overline{\mathrm{MAQ}}) = 0.5536308$, respectivamente. Nótese que, a diferencia de la propuesta 1, el individuo de menor conformidad a la regla es aquel cuya supresión provoca el mayor valor del índice.

Efectivos		
HIP/OPE	0	1
0	6	3
1	9	2

→

Intensidad al suprimir 1 individuo		
HIP/OPE	0	1
0	0.3766556	0.4573310
1	0.4080088	0.3470280

→

Ordenamiento		
HIP/OPE	0	1
0	3	1
1	2	4

Efectivos		
HIP/MAQ	0	1
0	2	7
1	2	9

→

Intensidad al suprimir 1 individuo		
HIP/MAQ	0	1
0	0.4208651	0.5821959
1	0.6775072	0.5289169

→

Ordenamiento		
HIP/MAQ	0	1
0	4	2
1	1	3

Para visualizar los ordenamientos en el caso de más modalidades, usamos los datos simulados descritos en el apartado anterior, y mostramos los resultados en la Tabla 7. Aunque los valores de intensidad de implicación son distintos, los ordenamientos producidos resultan ser los mismos que en la propuesta 1, debido al gran tamaño de la muestra.

Tabla 7 – Aplicación de la propuesta 2 con 5 modalidades para una implicación de una muestra simulada con 1000 individuos y 5 modalidades por variable: valores de intensidad de implicación al suprimir cada individuo posible (izquierda), y ordenamiento según el incremento provocado en la intensidad de implicación. La intensidad real, sin individuos añadidos, es $\varphi(V1, \overline{V}\,\overline{2}) = 0.3923378$. Los valores están redondeados por cuestión de espacio, pero no hay coincidencias (ejecutar código R de anexo para ver valores con más precisión). Nótese que, a diferencia de la propuesta 1, el individuo de menor conformidad a la regla es aquel cuya supresión provoca el mayor valor del índice.

Intensidad al suprimir un individuo tipo					
V1 /V2	0	0.25	0.5	0.75	1
0	0.384	0.388	0.392	0.396	0.401
0.25	0.388	0.390	0.392	0.394	0.396
0.5	0.392	0.392	0.392	0.392	0.392
0.75	0.397	0.394	0.392	0.390	0.388
1	0.401	0.397	0.392	0.388	0.384

→

Ordenamiento					
V1 /V2	0	0.25	0.5	0.75	1
0	24	21	14	5	2
0.25	20	18	12	8	6
0.5	9	10	11	16	17
0.75	3	7	13	19	23
1	1	4	15	22	25

En resumen, se puede concluir que, en el caso general, el ordenamiento de las parejas de modalidades no siempre es el mismo, y que dependa de los valores que tomen las variables en la muestra original, más cuanto menor sea el tamaño de la muestra.

4.3 Propuesta 3 : muestra auxiliar

Con una muestra auxiliar, que contiene un individuo, y sólo uno, de cada cruce de categorías, el ordenamiento es independiente de la muestra que cada uno esté analizando, al igual que ocurre con las propuestas de Bailleul y Gras (1994) y Ratsimba-Rajohn (1992). Esta muestra auxiliar, por simetría, representa una independencia estadística perfecta entre *a* y *b*, y por tanto una intensidad de implicación de 0.5. La supresión de cada individuo, provoca un desequilibrio en función de lo que aportaría dicho individuo a la implicación. Y el cálculo de ese incremento se puede interpretar como el apoyo o respeto a la regla. En la Tabla 8 se muestra la intensidad de implicación para la ausencia de cada individuo, y el ordenamiento inducido por esos cambios en la intensidad de implicación. Nótese cómo se escribe el ordenamiento en caso de "empate", pues es frecuente bajo esta propuesta.

Tabla 8 – Aplicación de la propuesta 3 con 2, 3 y 5 modalidades. Valores de intensidad de implicación al extraer de la muestra auxiliar cada individuo tipo, y ordenamiento según el incremento provocado. La intensidad con la muestra completa es 0.5.

Intensidad al suprimir un individuo tipo		
a / *b*	0	1
0	0.341	0.613
1	0.718	0.341

→

Ordenamiento		
a / *b*	0	1
0	3	2
1	1	3

Intensidad al suprimir un individuo tipo			
a / *b*	0	0.5	1
0	0.402	0.5	0.583
0.5	0.5	0.5	0.5
1	0.613	0.5	0.402

→

Ordenamiento			
a / *b*	0	0.5	1
0	8	3	2
0.5	3	3	3
1	1	3	8

Intensidad al suprimir un individuo tipo					
V1 /V2	0	0.25	0.5	0.75	1
0	0.442	0.472	0.5	0.527	0.554
0.25	0.471	0.485	0.5	0.513	0.527
0.5	0.5	0.5	0.5	0.5	0.5
0.75	0.529	0.514	0.5	0.485	0.472
1	0.560	0.529	0.5	0.471	0.442

→

Ordenamiento					
V1 /V2	0	0.25	0.5	0.75	1
0	24	20	9	5	2
0.25	22	18	9	8	5
0.5	9	9	9	9	9
0.75	3	7	9	18	20
1	1	3	9	22	24

En el caso de variables binarias, hay una diferencia de ordenamiento con las propuestas anteriores, pues el individuo con dato (0,0) es tan ventajoso de ser incluido en esta muestra auxiliar, como el individuo con datos (1,1). Lo es objetivamente, porque su introducción provoca el mismo incremento en la intensidad de la implicación.

5 Conclusiones

En el presente trabajo se ofrecen tres propuestas alternativas a las introducidas en Bailleul y Gras (1994) para medir el grado de adhesión de los individuos a las implicaciones entre variables modales, el cual sirve de base para obtener las tipicalidades y contribuciones de los individuos a las clases que se forman durante la construcción del árbol cohesitivo. Las propuestas vienen justificadas racionalmente, por el cambio que producen los individuos en el índice de la regla en cuestión. A partir de los ejemplos presentados se pudo evidenciar que corresponden a nuevas formas de ordenamiento de las parejas de modalidades, dependiendo de los valores que tomen las variables en los *n* individuos originales. Otro detalle importante, que no hemos entrado a valorar en esta propuesta, pero que vemos también influyente en una mejor clasificación de los individuos, es el de no usar un ordenamiento de posiciones, sino de los valores relativos. En el ordenamiento de Bailleul y Gras (1994), no hay "empates", y cada posición produce un salto fijo (de igual longitud) en el grado de adhesión a la regla. Con nuestra propuesta, el re-escalado desde las posiciones del ordenamiento al intervalo [0,1], se puede realizar como una homotecia, no desde los ordenamientos, sino desde los incrementos de intensidad de implicación calculados, de modo que dos individuos que producen el mismo incremento o muy similar, tendrían un mismo grado de adhesión, o muy similar, lo cual es más razonable que producir un salto de tamaño fijo de uno al siguiente. Una posible continuación de este trabajo sería la comparativa de grupos óptimos de individuos para la tipicalidad y contribución, entre la propuesta de Bailleul y Gras (1994) y las nuestras, usando muestras pequeñas, que se puedan interpretar de modo directo.

6 Agradecimientos

El segundo autor agradece el apoyo económico de los proyectos P1·1B2012-52 de la Universitat Jaume I de Castellón y MTM2013-43917-P del Ministerio de Economía y Competitividad de España.

Bibliografía

[1] Bailleul, M. (1994), *Analyse statistique implicative: variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique*, Thèse de doctorat, Université de Rennes 1.

[2] Bailleul, M. et R. Gras (1994), L'implication statistique entre variables modales, *Mathématiques et sciences humaines*, **128**, 41-57.

[3] Gras, R. (1979), *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques*, Thèse de doctorat, Université de Rennes 1.

[4] Gras, R., S. Ag Almouloud, M. Bailleul, A. Larher, M. Polo, H. Ratsimba-Rajohn et A. Totohasina (1996), *L'implication statistique : nouvelle méthode exploratoire de données*, La Pensée Sauvage, Grenoble.

[5] Gras, R. et A. Larher (1992), L'implication statistique, une nouvelle méthode d'analyse de données, *Mathématiques et sciences humaines*, **120**, 5-31.

[6] Lagrange, J.-B. (1998), Analyse implicative d'un ensemble de variables numériques; application au traitement d'un questionnaire à réponses modales ordonnées, *Revue de statistique appliquée*, **46**, 71-93.

[7] Lerman, I.C. (1970), Sur l'analyse des données préalable à une classification automatique (proposition d'une nouvelle mesure de similarité), *Mathématiques et sciences humaines*, **32**, 5-15.

[8] Lerman I.C., R. Gras R. et H. Rostam (1981), Elaboration et évaluation d'un indice d'implication pour des données binaires I, *Mathématiques et sciences humaines*, **74**, 5-35.

[9] Lerman I.C., R. Gras R. et H. Rostam (1981), Elaboration et évaluation d'un indice d'implication pour des données binaires II, *Mathématiques et sciences humaines*, **75**, 5-47.

[10] Pitarch, I. y Orús, P. (2002), *Estudio sobre la viabilidad y el interés didáctico del tratamiento de la información en la ESO*, Trabajo de Investigación, Tercer ciclo, Universidad Jaume I de Castellón.

[11] Ratsimba-Rajohn, H. (1992), *Contribution à l'étude de la hiérarchie implicative, application à l'analyse de la gestion didactique des phénomènes d'ostension et de contradiction*, Thèse de doctorat, Université de Rennes 1.

[12] Zamora L., J. Díaz y L. Portuondo (2015), Fundamental concepts on classification and statistical implicative analysis for modal variables, próxima publicación en *Revista Colombiana de Estadística*.

Anexo

6.1 Datos originales

La Tabla 9 contiene los datos originales sobre gustos musicales, extraídos de Pitarch y Orús (2002). Se lanza la pregunta *¿Te gusta la música…?* bajo 15 estilos musicales (hip-hop, ópera, máquina, jazz, flamenco, rumba, heavy, pop, clásica, dance, reggae, hardcore, rap, punk y ska), a 20 jóvenes de entre 15 y 20 años.

Tabla 9 – Datos de un estudio real sobre gustos musicales (Pitarch y Orús, 2002), que utilizamos para valorar nuestras propuestas sobre alternativas al grado de adhesión de los individuos a las implicaciones.

	HIP	OPE	MAQ	JAZ	FLA	RUM	HEA	POP	CLA	DAN	REG	HAR	RAP	PUN	SKA
al1	1	1	1	0	0	1	0	1	1	1	0	0	1	1	1
al2	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1
al3	1	0	0	0	0	1	0	0	0	1	0	1	1	1	1
al4	0	1	0	1	1	1	1	0	1	0	1	0	1	0	1
al5	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1
al6	0	0	1	0	1	1	0	1	0	1	0	1	1	0	0
al7	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0
al8	0	1	1	1	1	1	0	1	0	1	0	1	1	0	1
al9	1	0	0	0	1	1	0	0	0	1	1	0	1	1	0
al10	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
al11	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1
al12	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1
al13	1	0	1	0	0	1	1	1	0	1	0	1	0	1	1
al14	0	0	1	0	1	1	0	1	0	1	0	1	1	0	1
al15	0	1	1	1	1	1	0	1	0	1	0	1	1	0	1
al16	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1
al17	1	0	1	0	1	1	0	1	0	1	1	1	1	0	1
al18	0	0	1	0	0	1	0	1	0	1	0	1	0	0	0
al19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
al20	0	0	1	0	0	1	1	1	0	1	0	1	1	0	1

6.2 Códigos en lenguaje R para la computación

Código para calcular el ordenamiento de Bailleul y Gras (1994), que se muestra en la Tabla 2:

```
m = 5 # número de categorías
rg = array(0, dim=c(m,m), dimnames=list(categ, categ))
# preparamos matriz para poner rangos
for(i in 1:(m-1)) {
  for(j in 1:i) {
    a = as.character((m-j)/(m-1))
    b = as.character((i-j)/(m-1))
    rg[a,b] = i*(i-1)/2+j
  }
}
for(i in m:1) {
  for(j in 0:(i-1)) {
    a = as.character((m-i)/(m-1))
    b = as.character((m-i+j)/(m-1))
    rg[a,b] = m^2 - i*(i+1)/2 + j + 1
  }
}
```

```
rg
```

Código para calcular la intensidad de implicación de una pareja de variables modales:

```
phi = function(x) {
  # x : dos columnas de datos de variables modales escaladas
  # al intervalo [0,1]
  # devuelve phi con aproximación gaussiana
  n = dim(x)[1]
  mu.a = mean(x[,1])
  mu.nob = mean(1-x[,2])
  mu.anob = mean(x[,1]*(1-x[,2]))
  var.a = var(x[,1])*(n-1)/n
  var.b = var(x[,2])*(n-1)/n
  return( 1 - pnorm(q=mu.anob, mean=mu.a*mu.nob,
                    sd=sqrt((mu.a^2+var.a) *
                            (mu.nob^2+var.b)/n) )
        )
}
```

Código para generar los datos simulados, utilizados en las propuestas 1 y 2, cuyos resultados se muestran en las Tablas 5 y 7:

```
# semilla para reproducibilidad
set.seed(20151111) # fecha congreso ASI8
m=5 # número de modalidades
categ = (0:(m-1))/(m-1) # las modalidades 0, 1/(m-1),...,1
x = matrix(data=sample(x=categ, replace=TRUE, size=1000*2),
           ncol=2)
# x es la matriz de datos con 2 columnas y 1000 filas
```

Código para representar los ordenamientos de la Tabla 5 (propuesta 1), y que se pueden aplicar a cualquier conjunto de datos:

```
# propuesta 1: añadir individuo de cada tipo
# x contiene los datos los simulados anteriormente
# pero se puede poner cualquier muestra bivariante
phi0 = phi(x) # intensidad implicativa muestra de datos
m = 5 # número de modalidades de x (cotejar con x)
phi1 = matrix(data=0, ncol=m, nrow=m)
# phi1 es matriz que almacena la nueva intensidad implicativa
# al añadir un individuo de cada uno de los m^2 tipos
for(i in 1:m) { # para cada modalidad de V1
  for(j in 1:m) { # para cada modalidad de V2
    x2 = rbind(x, c(categ[i], categ[j])) # x con nuevo individuo
    phi1[i,j] = phi(x2)
    # calcula phi de la muestra con individuo añadido
  }
}
phi1 # matriz con nuevas intensidades de implicación
matrix(rank(phi1, ties="min"), ncol=m, dimnames=list(categ, categ))
# los ordenamientos resultantes de comparar
# (phi con individuo) – (phi sin individuo)
# phi1 – phi0
```

Código para representar los ordenamientos de la Tabla 7 (propuesta 2):

```
# propuesta 2: suprimir individuo de cada tipo
# x contiene los datos los simulados anteriormente
# pero se puede poner cualquier muestra bivariante
phi0 = phi(x) # intensidad implicativa muestra de datos
m = 5 # número de modalidades de x (cotejar con x)
phi2 = matrix(data=0, ncol=m, nrow=m)
# phi2 es matriz que almacena la nueva intensidad implicativa
# al suprimir un individuo de cada uno de los m^2 tipos
# si está presente en la muestra
n = dim(x)[1] # número de individuos
for(i in 1:m) { # para cada modalidad de V1
  for(j in 1:m) { # para cada modalidad de V2
    donde = (1:n)[categ[i]==x[,1] & categ[j]==x[,2]]
    # busca lugar(es) donde el individuo está en los datos
    existe = sum(donde) > 0
    # detecta si ese individuo existe en la muestra
    if(existe)  phi2[i,j] = phi((x[-donde[1], ]))
    else phi2[i,j] = NA
    # si existe, lo suprime y calcula phi
    # si no existe devuelve NA (no disponible)
  }
}
phi2 # matriz con nuevas intensidades de implicación
matrix(rank(-phi2, ties="min"), ncol=m,
       dimnames=list(categ, categ))
# los ordenamientos resultantes de comparar
# (phi con individuo) – (phi sin individuo)
# phi0 – phi2
```

Código para representar los ordenamientos de la Tabla 8 (propuesta 3):

```
# propuesta 3: muestra auxiliar con un individuo de cada tipo
# que se suprime y se compara el incremento en la intensidad
m=5 # número de modalidades
categ = (0:(m-1))/(m-1) # las modalidades 0, 1/(m-1),...,1
x = expand.grid(a=categ, b=categ) # crea muestra auxiliar completa
phi0 = phi(x) # phi de la muestra completa
phi3 = matrix(data=0, ncol=m, nrow=m)
# phi3 es matriz que almacena la nueva intensidad implicativa
# al suprimir un individuo de cada uno de los m^2 tipos
for(i in 1:m) {
  for(j in 1:m) {
    phi3[i,j] = phi(x[-(j+(i-1)*m),])
  }
}
phi3 # matriz con nuevas intensidades de implicación
matrix(rank(-phi3, ties="min"), ncol=m, byrow=TRUE,
       dimnames=list(categ, categ))
# los ordenamientos resultantes de comparar
# (phi con individuo) – (phi sin individuo)
# phi0 – phi3
```

VARIABILITY OF GRAS CLASSICAL IMPLICATION INDEX AND OTHER RULE QUALITY MEASURES UNDER SMALL SIZE SAMPLING FROM BIVARIATE BINARY PROCESSES

Pablo GREGORI[1] and Raphaël COUTURIER[2]

VARIABILITE DE L'INDICE D'IMPLICATION CLASSIQUE DE GRAS ET D'AUTRES MESURES DE QUALITE DE REGLES PAR ECHANTILLONAGE DE PROCESSUS BINAIRES BIVARIES DE PETITE TAILLE

ABSTRACT

Starting from a random bivariate Bernoulli process (A,B), we compute the variability in the (random) rule quality $A \rightarrow B$, measured by four common index (confidence, lift, Loevinger's H and classical Gras' φ). We also compute correlation coefficients among every couple of these measures. Computations are constrained to small size samples, and performed with R Statistical Software.

Keywords: *Statistical implicative analysis, association rules, confidence, lift, R language.*

RÉSUMÉ

A partir d'un processus aléatoire de Bernoulli bivarié (A,B), on calcule la variabilité dans la qualité de la règle (aléatoire) $A \rightarrow B$, mesuré par quatre indices habituels (confiance, lift, H de Loevinger et φ de Gras). On calcule aussi des coefficients de corrélation entre chaque couple de ces mesures. Les calculs sont restreints à des échantillons de petite taille, et ont été réalisés avec le logiciel statistique R.

Mots-clés : *Analyse statistique implicative, règles d'association, confiance, lift, langage R.*

1 Introduction

Association rules (AR) is a data mining technique initiated by Agrawal *et al.* (1993). It aims at turning data into knowledge, in the form of quasi-implications (or rules, i.e. sentences like « whenever a feature a is observed, then another feature b is usually observed »). Some of the most common statistical measures of the level of accomplishment of such rules are the *confidence*, *lift* and *support*, but there are dozens of such indicators (see, for instance, the review of Gras *et al.*, 2004, or Lalich *et al.*, 2007). It is a widely used technique, whose prototypical example and application field was the *market basket analysis*. The well known *apriori* algorithm (Agrawal *et al.*, 1993) and its subsequent improvements, together with the ease of interpretation of such measures, have been the keypoints for its spread use. The success of the algorithm (high speed in processing huge databases) is based upon the *a priori* rejection of candidate

[1] Departamento de Matemáticas. Instituto de Matemáticas y Aplicaciones de Castellón, Universitat Jaume I de Castellón, Campus Riu Sec E-12071, Spain, gregori@uji.es

[2] FEMTO-ST Institute, University of Bourgogne Franche-Comte, IUT Belfort-Montbéliard, Belfort, France, raphael.couturier@univ-fcomte.fr

rules of *low* support. That is the reason why this algorithm neglects rules involving a small fraction of the population, even when those rules might hide amazing and important results.

Statistical implicative analysis (SIA) was defined in order to help in the same kind of problems: measuring the quality of rules in the context of Didactics of Mathematics (Gras *et al.* 1996, 2013). The chosen statistical measure was the departure from independence between hypothesis and thesis, under some sampling scenarios. In this context, Loevinger's H index was already defined and also widely used in applications in Human Sciences, but for the author of SIA, it was not rich enough (see for instance Gras and Couturier, 2010).

Although SIA has been developing tools to deal with many types of data (Gras *et al.* 2008, 2013), we restrict ourselves to the study of the behavior of (the classical) Gras implication index for the simplest binary data: a bivariate Bernoulli process. We started this analysis in Gregori *et al.* (2013), providing an explicit computation of the probability distribution of the index, in terms of the parameters of the Bernoulli process. We also showed the effect of the conditional probability on the distribution of the expected value and quartiles of the implication index, fixing the other parameters and the sample size.

In this work, we relate the distribution of the Gras implication index to the distributions of other commonly used indexes, such as confidence, lift and Loevinger's H. We also stress the meaning of each index, in order to help researchers to better interpret them and choose the most suitable one for his or her particular work.

2 Definitions and notations

In SIA historical notation, at the very first definition of the implicative intensity, a set of variables is denoted by V and a couple of those variables are denoted by a and b. The set of sampled individuals is denoted by E, and each of those individuals is said to « hold property a » (or to be an example for property a) if a is « true » for the individual, and it is said « not to hold property a » (or to be a counterexample for property a) if a is « false » for the individual. Using logical operators (complementary, AND) we can also consider properties $\bar{b}$ and $a \wedge \bar{b}$ and check whether individuals hold those properties as true or false.

Afterwards, letters A and B denote, respectively, the subsets of individuals holding true respectively variable a and b, and finally, X and Y denote random sets of individuals, modelling the respective sets A and B, under some random sampling schemes. The surprisingness of the implication (or rule) $a \to b$ is then defined comparing the sample data against the statistical independence assumption.

In this paper, we do not remind the theory of SIA, and we use A and B to denote a couple of Bernoulli random variables that shall generate, at each instance, and for a given sample size $N = n$, a particular sample of n individuals for variables a and b, with the usual identification 0 (for false) and 1 (for true).

Example: Let (A,B) be a couple of Bernoulli variables, with $P(A=1)=P(B=1)=0.5$ and $P(B=1\,|\,A=1)=0.75$, and let $N=10$. Let us run a simulation of (A,B) and show it in Table 1.

Table 1 – A particular instance of (A,B) for $P(A=1)=P(B=1)=0.5$ and $P(B=1\,|\,A=1)=0.75$, and sample size $N=10$, seen as raw data (left), contingency table (middle), relative frequency table (right).

a	0	1	0	1	0	1	1	1	1	0
b	0	1	1	1	0	1	1	1	1	0

a / b	0	1
0	3	1
1	0	6

a / b	0	1
0	0.3	0.1
1	0	0.6

The notations for the contingency table and the relative frequency table are shown in Table 2.

Table 2 – Notations for contingency (left) and relative frequency tables (right).

a / b	0	1	Margin a
0	$n_{\bar{a}\bar{b}}$	$n_{\bar{a}b}$	$n_{\bar{a}}$
1	$n_{a\bar{b}}$	n_{ab}	n_a
Margin b	$n_{\bar{b}}$	n_b	n

a / b	0	1	Margin a
0	$f_{\bar{a}\bar{b}}$	$f_{\bar{a}b}$	$f_{\bar{a}}$
1	$f_{a\bar{b}}$	f_{ab}	f_a
Margin b	$f_{\bar{b}}$	f_b	1.00

Definition: For a given joint frequency table for (a,b), the quality index for rule $a \rightarrow b$ that we consider are defined as:

- Confidence : $C=\dfrac{f_{ab}}{f_a}$.
- Lift : $L=\dfrac{C}{f_b}=\dfrac{f_{ab}}{f_a\times f_b}$.
- Loevinger's H : $H=1-\dfrac{f_{a\bar{b}}}{f_a\times f_{\bar{b}}}$.
- Gras' φ : $\varphi=1-F(n_{a\bar{b}})$, where F is the cumulative distribution function of the binomial model of size n and success probability $\dfrac{n_a\times n_{\bar{b}}}{n^2}$, in one of its possible modelisations, the one we choose for this contribution.

After the realisation (a,b) of (A,B) in the example, the quality of the sampled rule $a\rightarrow b$ shows the index values :

- $C=\dfrac{0.6}{0.6}=1$, $L=\dfrac{0.6}{0.6\times 0.6}=1.4285714$, $H=1-\dfrac{0}{0.6\times 0.3}=1$.
- $\varphi=1-\mathrm{F}_{\mathrm{Bin}\left(10,\frac{6\times 3}{10^2}\right)}(0)=1-0.137448=0.862552$.

Obviously, the sampled realisation (a,b) is only one, and we want to show the « whole » picture for (A,B). For that purpose, we need intensive computation in order to

write all possible simulations, weighted by their probabilities, and compute all indexes. We shall analyse the variability in the rule quality under each index, and also the joint variability under each couple of indexes when the underlying process is known (by the given parameters).

Then, for (A,B) a bivariate Bernouilli random variable, determined by parameters $p_A := P(A=1)$, $p_B := P(B=1)$ and $p_{AB} := P(A=1,B=1)$ (or equivalently $p_{B|A} := P(B=1 \mid A=1) = p_{AB} / p_A$, whenever $p_A > 0$), we can represent it in the form of a joint probability table as in Table 3.

Table 3 – Notation for the parameters of a bivariate Bernoulli random variable (A,B).

A / B	0	1	Margin A
0	$p_{\bar{A}\bar{B}}$	$p_{\bar{A}B}$	$p_{\bar{A}}$
1	$p_{A\bar{B}}$	p_{AB}	p_A
Margin B	$p_{\bar{B}}$	p_B	1.00

We shall consider only simple random sampling (SRS) from (A,B) with « fixed » sample size n, and our goal is to show probabilistic properties of some quality (or interestingness) measures of the implication $A \rightarrow B$. To this end, and arising from the sampling process, we consider the random contingency table shown in Table 4, where all absolute frequencies (cases) are integer random variables.

Table 4 – Notation for the random contingency table arising from a SRS of size N from the bivariate Bernoulli variable (A,B).

A / B	0	1	Margin A
0	$N_{\bar{a}\bar{b}}$	$N_{\bar{a}b}$	$N_{\bar{a}}$
1	$N_{a\bar{b}}$	N_{ab}	N_a
Margin B	$N_{\bar{b}}$	N_b	N

Under our « fixed » sample size framework, N is not random, because $N = n$. Now we study the probability laws and relationships among the following *random* quality measures (seen as functions of random variables A and B):

– Confidence $C := \dfrac{N_{ab}/N}{N_a/N} = \dfrac{N_{ab}}{N_a} = 1 - \dfrac{N_{a\bar{b}}}{N_a}$.

– Lift $L := \dfrac{C}{N_b/N} = \dfrac{1 - N_{a\bar{b}}/N_a}{1 - N_{\bar{b}}/N}$.

– Loevinger's $H := 1 - \dfrac{N_{a\bar{b}}/N}{(N_a/N)(N_{\bar{b}}/N)} = 1 - \dfrac{N_{a\bar{b}}/N_a}{N_{\bar{b}}/N}$.

– Gras' $\varphi := 1 - F(N_{a\bar{b}})$, where F means the cumulative distribution function (cdf) of the binomial model $Bin\left(N, \frac{N_a \times N_{\bar{b}}}{N^2}\right)$ (for this « fixed sample size » case).

We have intentionally displayed the formulae in common terms (N, N_a, $N_{\bar{b}}$ and $N_{a\bar{b}}$) in order to better compare their analytical expressions. The implementation of these formulae in R language can be seen in the Appendix.

One can see that confidence is ill defined when $N_a = 0$, and so is lift when $N_a \times N_b = 0$, and Loevinger's H whenever $N_a \times N_{\bar{b}} = 0$. Gras' φ is well defined in any case, and then we shall exclude the problematic cases of the other indexes from computations.

As it was shown in the example before, and in usual practice, all these (random) quality measures are actually seen as statistics, computed from an observed sample, and they seize the quality of the instance rule $a \to b$ in different manners. It is clearly detailed in Vaillant *et al.* (2008) and we just stress it here for convenience:

– *Confidence* C stands for estimating the predictive quality of the rule (if property A is held by an individual of the sample, then an approximation to the probability for that individual to hold property B is C). The range of confidence lies within $[0,1]$. High values of confidence are attributed to a strong degree of association. However, for example, if A and B are any two statistically independent binary random variables, the law of large numbers suggests that C will be likely close to $p_{b|a} = p_b$. If one chooses B to be abundant (with p_b close to 1.0), then many samples of (A,B) will show large confidence values, even when $A = 1$ does not push the chances of $B = 1$.

– *Lift* L stands for measuring the improvement of predicting B once the rule is taken into account. For example, a value of 3.0 means that the probability of an individual to hold property B raises by a factor of 3.0 at the moment we know that the individual holds property A (i.e. $p_{b|a} = 3p_b$). In this case $A = 1$ would « lift » the chances of $B = 1$. Statistical independence corresponds to a lift of 1.0. The range of lift is within $[0, N]$.

– *Loevinger's* H measures a distance of the relative number of counter examples from independence. Statistical independence corresponds to a value of 0. A value of 1.0 means no counterexamples at all (hence the strongest association from A to B). The range of H lies within $[1 - N, 1]$.

– *Gras'* φ measures how extraordinarily small the number of counterexamples to the rule is, given the sampled amounts of $A = 1$ and $B = 1$, compared to what are expected under independence of A and B. The range of φ lies within $[0,1]$. Its purpose is to discover association even if only a small number of individuals is involved. In practice, users should pay special attention only to rules with indexes over 0.90 or 0.95, because they are more clearly away from independence. Rules with lower values (such as 0.80 and below) must be

analysed with care. We have found that many practitioners of SIA tend to interpret φ in the same way as C, and our intention is to show clearly that those indexes measure different aspects: not only by definition, but by computing their distributions and correlations. We stress that, *if the goal is to discover rules which cannot be explained by randomness, then confidence is not the tool, but* φ. Once the rule is targeted, even if the confidence were to be low, the association is very likely to exist. *If one seeks for predictive ability, then the confidence is the right tool*.

Finally, we shall mention that statistics C, L and H do not vary under the scaling of the contingency table. However φ does. Besides, a drawback of φ was highlighted in Bodin (1996): with very large sample sizes, it was given an extremely high value (0.9999) to a rule with a confidence as poor as 33%. Gras redefined φ (dubbed *entropic version*) in order to avoid such behavior. However, that behavior is not necessarily a sign of weakness: Gras index and confidence are measuring different effects indeed; 66% of counterexamples are unacceptable for the task of prediction, but it is a ridiculous fraction of counterexamples, in the eyes of probability, when checking for independence.

3 Probability distributions of absolute frequencies

Under the SRS with fixed sample size, N is not random, but takes the fixed value n. The 4-variate $\left(N_{\bar{a}\bar{b}}, N_{\bar{a}b}, N_{a\bar{b}}, N_{ab}\right)$ follows the multinomial distribution $Multinom\left(n; p_{\bar{A}\bar{B}}, p_{\bar{A}B}, p_{A\bar{B}}, p_{AB}\right)$. Recall that the probability function (pf) of a general $X \sim Multinom\left(n; p_1, p_2, \ldots, p_k\right)$ is

$$f_n\left(x_1, x_2, \ldots, x_k\right) = \begin{cases} \dfrac{n!}{x_1! x_2! \quad x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, & \text{if } x_1 + x_2 + \cdots + x_k = n \\ 0, & \text{otherwise} \end{cases}$$

The distributions of the two bivariate margins are $\left(N_{\bar{a}}, N_a\right) \sim \text{Multinom}\left(n; p_{\bar{A}}, p_A\right)$ and $\left(N_{\bar{b}}, N_b\right) \sim Multinom\left(n; p_{\bar{B}}, p_B\right)$.

Although we have not implemented the SRS with a random sample size for computations, let us sketch up the theoretical distribution function, as a 2-step random procedure:

1. Sampling from $N \sim Po(\mu)$, for a given parameter μ, and
2. conditionned to $N = n$, sampling from $\left(N_{\bar{a}\bar{b}}, N_{\bar{a}b}, N_{a\bar{b}}, N_{ab}\right) \sim Multinom\left(n; p_{\bar{A}\bar{B}}, p_{\bar{A}B}, p_{A\bar{B}}, p_{AB}\right)$.

In this case, N is random, and it is easy to deduce that the (unconditional) pf of the 4-variate $\left(N_{\bar{a}\bar{b}}, N_{\bar{a}b}, N_{a\bar{b}}, N_{ab}\right)$ is:

$$
\begin{aligned}
&f\left(n_{\bar{a}\bar{b}}, n_{\bar{a}b}, n_{a\bar{b}}, n_{ab}\right) = \sum_{n=0}^{\infty} P(N=n)\, f_n\left(n_{\bar{a}\bar{b}}, n_{\bar{a}b}, n_{a\bar{b}}, n_{ab}\right) \\
&= P(N = n_{\bar{a}\bar{b}} + n_{\bar{a}b} + n_{a\bar{b}} + n_{ab})\, f_n\left(n_{\bar{a}\bar{b}}, n_{\bar{a}b}, n_{a\bar{b}}, n_{ab}\right) \\
&= e^{-\mu} \frac{\mu^{n_{\bar{a}\bar{b}} + n_{\bar{a}b} + n_{a\bar{b}} + n_{ab}}}{(n_{\bar{a}\bar{b}} + n_{\bar{a}b} + n_{a\bar{b}} + n_{ab})!} \cdot \frac{(n_{\bar{a}\bar{b}} + n_{\bar{a}b} + n_{a\bar{b}} + n_{ab})!}{n_{\bar{a}\bar{b}}!\, n_{\bar{a}b}!\, n_{a\bar{b}}!\, n_{ab}!}\, p_{\bar{A}\bar{B}}{}^{n_{\bar{a}\bar{b}}}\, p_{\bar{A}B}{}^{n_{\bar{a}b}}\, p_{A\bar{B}}{}^{n_{a\bar{b}}}\, p_{AB}{}^{n_{ab}} \\
&= \left(e^{-(\mu p_{\bar{A}\bar{B}})} \frac{(\mu p_{\bar{A}\bar{B}})^{n_{\bar{a}\bar{b}}}}{n_{\bar{a}\bar{b}}!} \right) \cdot \left(e^{-(\mu p_{\bar{A}B})} \frac{(\mu p_{\bar{A}B})^{n_{\bar{a}b}}}{n_{\bar{a}b}!} \right) \cdot \left(e^{-(\mu p_{A\bar{B}})} \frac{(\mu p_{A\bar{B}})^{n_{a\bar{b}}}}{n_{a\bar{b}}!} \right) \cdot \left(e^{-(\mu p_{AB})} \frac{(\mu p_{AB})^{n_{ab}}}{n_{ab}!} \right) \\
&= f_{Po(\mu p_{\bar{A}\bar{B}})}(n_{\bar{a}\bar{b}}) \cdot f_{Po(\mu p_{\bar{A}B})}(n_{\bar{a}b}) \cdot f_{Po(\mu p_{A\bar{B}})}(n_{a\bar{b}}) \cdot f_{Po(\mu p_{AB})}(n_{ab})
\end{aligned}
$$

It can be seen that all 4 random variables are mutually independent and have marginal Poisson distributions: $N_{\bar{a}\bar{b}} \sim Po(\mu p_{\bar{A}\bar{B}})$, $N_{\bar{a}b} \sim Po(\mu p_{\bar{A}B})$, $N_{a\bar{b}} \sim Po(\mu p_{A\bar{B}})$ and $N_{ab} \sim Po(\mu p_{AB})$.

For the two bivariate margins, one can proceed similarly, or by seeing them as sums of independent Poisson distributions. Then: $N_a \sim Po(\mu p_A)$ and $N_{\bar{a}} \sim Po(\mu p_{\bar{A}})$ are mutually independent, and so $N_b \sim Po(\mu p_{\bar{B}})$ and $N_{\bar{b}} \sim Po(\mu p_{\bar{B}})$ are. Be careful that N_a and N_b are not independent since they share a common term in the sum.

The pdf of the full 5-variate $\left(N_{\bar{a}\bar{b}}, N_{\bar{a}b}, N_{a\bar{b}}, N_{ab}, N\right)$ can be simplified to:

$$
f\left(n_{\bar{a}\bar{b}}, n_{\bar{a}b}, n_{a\bar{b}}, n_{ab}, n\right) = \begin{cases} e^{2\mu} \dfrac{\mu^n}{n!} \dfrac{\mu_{\bar{a}\bar{b}}{}^{n_{\bar{a}\bar{b}}}}{n_{\bar{a}\bar{b}}!} \dfrac{\mu_{\bar{a}b}{}^{n_{\bar{a}b}}}{n_{\bar{a}b}!} \dfrac{\mu_{a\bar{b}}{}^{n_{a\bar{b}}}}{n_{a\bar{b}}!} \dfrac{\mu_{ab}^{n_{ab}}}{n_{ab}!}, & \text{if } n_{\bar{a}\bar{b}} + n_{\bar{a}b} + n_{a\bar{b}} + n_{ab} = n \\ 0, & \text{otherwise} \end{cases}
$$

where $\mu_{\bar{a}\bar{b}} = \mu p_{\bar{A}\bar{B}}$, $\mu_{\bar{a}b} = \mu p_{\bar{A}B}$, $\mu_{a\bar{b}} = \mu p_{A\bar{B}}$ and $\mu_{ab} = \mu p_{AB}$.

4 Probability distributions of the quality measures

The probability distributions of the quality measures shall be computed directly from the 4-variate joint absolute frequencies. Exact analytical expressions are not available, then we shall show particular examples and restrict our computations to moderate values of n, since partitions of n in all variations of 4 integers are computationally expensive for large n. We are using the `partitions` package of R Statistical Software (Hankin, 2006).

The laws of C, L, H and φ can be computed as usual in transformation of discrete random variables:

– $P(C=c)=\sum_{c=\frac{n_{ab}}{n_a}}\left(N_{\bar{a}\bar{b}}=n_{\bar{a}\bar{b}},N_{\bar{a}b}=n_{\bar{a}b},N_{a\bar{b}}=n_{a\bar{b}},N_{ab}=n_{ab}\right)$, where the sum spans over all partitions of n into four non negative integers $\left(n_{\bar{a}\bar{b}},n_{\bar{a}b},n_{a\bar{b}},n_{ab}\right)$, and c attains the value $\frac{n_{ab}}{n_a}$, with $n_a := n_{a\bar{b}}+n_{ab}$.

– $\mathrm{P}(\mathrm{L}=\mathrm{l})=\sum_{\mathrm{l}==\frac{\mathrm{nn}_{\mathrm{ab}}}{\mathrm{n}_{\mathrm{a}}\mathrm{n}_{\mathrm{b}}}}\left(\mathrm{N}_{\bar{\mathrm{a}}\bar{\mathrm{b}}}=\mathrm{n}_{\bar{\mathrm{a}}\bar{\mathrm{b}}},\mathrm{N}_{\bar{\mathrm{a}}\mathrm{b}}=\mathrm{n}_{\bar{\mathrm{a}}\mathrm{b}},\mathrm{N}_{\mathrm{a}\bar{\mathrm{b}}}=\mathrm{n}_{\mathrm{a}\bar{\mathrm{b}}},\mathrm{N}_{\mathrm{ab}}=\mathrm{n}_{\mathrm{ab}}\right)$, where the sum spans over all partitions of n into four non negative integers $\left(n_{\bar{a}\bar{b}},n_{\bar{a}b},n_{a\bar{b}},n_{ab}\right)$, and l attains the value $\frac{nn_{ab}}{n_a n_b}$, with $n_a := n_{a\bar{b}}+n_{ab}$ and $n_b := n_{\bar{a}b}+n_{ab}$.

– $P(H=h)=\sum_{h=1-\frac{nn_{a\bar{b}}}{n_a n_{\bar{b}}}}\left(N_{\bar{a}\bar{b}}=n_{\bar{a}\bar{b}},N_{\bar{a}b}=n_{\bar{a}b},N_{a\bar{b}}=n_{a\bar{b}},N_{ab}=n_{ab}\right)$, where the sum spans over all partitions of n into four non negative integers $\left(n_{\bar{a}\bar{b}},n_{\bar{a}b},n_{a\bar{b}},n_{ab}\right)$, and h attains the value $1-\frac{nn_{a\bar{b}}}{n_a n_{\bar{b}}}$, with $n_a := n_{a\bar{b}}+n_{ab}$ and $n_{\bar{b}} := n_{a\bar{b}}+n_{\bar{a}\bar{b}}$.

– $P(\varphi=f)=\sum_{f=1-F\left(n_{a\bar{b}}\right)}\left(N_{\bar{a}\bar{b}}=n_{\bar{a}\bar{b}},N_{\bar{a}b}=n_{\bar{a}b},N_{a\bar{b}}=n_{a\bar{b}},N_{ab}=n_{ab}\right)$, where the sum spans over all partitions of n into four non negative integers $\left(n_{\bar{a}\bar{b}},n_{\bar{a}b},n_{a\bar{b}},n_{ab}\right)$, f attains the value of the complement to one, to the cdf of the binomial model of parameters n and $n_a n_{\bar{b}}/n^2$, with $n_a := n_{a\bar{b}}+n_{ab}$ and $n_{\bar{b}} := n_{a\bar{b}}+n_{\bar{a}\bar{b}}$.

We show in Table 5 the way we have proceeded with a toy example: (1) a list of all compositions of n is created, using the R package `partitions`, (2) the probability of each composition is computed, using the R function `dmultinom()`, and (3) the computation of the four index for each composition.

In order to obtain the probability distribution of each index, seen as a random variable, one has to check the repeated values of the index under different compositions, and sum up all their respective probabilities. As one can see in the previous table, all index but Gras, have ill definition (NA value) with positive probability.

We show results in Figure 1, in the form of plots, for specific parameter values, but the scripts are provided, so that the interested researcher can modify and get computation and plots for other parameter values.

Table 5: A list of some compositions of $n = 5$, seen as the four joint frequency tables of the contingency table, together with their probability, under $p_{\bar{A}\bar{B}} = p_{\bar{A}B} = p_{A\bar{B}} = p_{AB} = 0.25$, and the computed values of the quality index. Ill definition occurs at some compositions for all index except Gras' φ.

$N_{\bar{a}\bar{b}}$	$N_{\bar{a}b}$	$N_{a\bar{b}}$	N_{ab}	Prob	C	L	H	φ
5	0	0	0	0.0009766	NA	NA	NA	0.0000000
4	1	0	0	0.0048828	NA	NA	NA	0.0000000
3	2	0	0	0.0097656	NA	NA	NA	0.0000000
2	3	0	0	0.0097656	NA	NA	NA	0.0000000
1	4	0	0	0.0048828	NA	NA	NA	0.0000000
0	5	0	0	0.0009766	NA	NA	NA	0.0000000
4	0	1	0	0.0048828	0.0	NA	0.0000000	0.2627200
3	1	1	0	0.0195312	0.0	0.0000000	-0.2500000	0.1834910
2	2	1	0	0.0292969	0.0	0.0000000	-0.6666667	0.1124509
1	3	1	0	0.0195312	0.0	0.0000000	-1.5000000	0.0543613
0	4	1	0	0.0048828	0.0	0.0000000	-4.0000000	0.0147580
3	0	2	0	0.0097656	0.0	NA	0.0000000	0.3174400
2	1	2	0	0.0292969	0.0	0.0000000	-0.2500000	0.1905263
1	2	2	0	0.0292969	0.0	0.0000000	-0.6666667	0.0932512
0	3	2	0	0.0097656	0.0	0.0000000	-1.5000000	0.0317587
2	0	3	0	0.0097656	0.0	NA	0.0000000	0.3369600
1	1	3	0	0.0195312	0.0	0.0000000	-0.2500000	0.1634992
0	2	3	0	0.0097656	0.0	0.0000000	-0.6666667	0.0597943
1	0	4	0	0.0048828	0.0	NA	0.0000000	0.3276800
0	1	4	0	0.0048828	0.0	0.0000000	-0.2500000	0.1073742
0	0	5	0	0.0009766	0.0	NA	0.0000000	0.0000000
4	0	0	1	0.0048828	1.0	5.0000000	1.0000000	0.5817881
3	1	0	1	0.0195312	1.0	2.5000000	1.0000000	0.4722681
2	2	0	1	0.0292969	1.0	1.6666667	1.0000000	0.3409185
1	3	0	1	0.0195312	1.0	1.2500000	1.0000000	0.1846273
0	4	0	1	0.0048828	1.0	1.0000000	NA	0.0000000
3	0	1	1	0.0195312	0.5	2.5000000	0.3750000	0.5125046
2	1	1	1	0.0585937	0.5	1.2500000	0.1666667	0.3461014
1	2	1	1	0.0585937	0.5	0.8333333	-0.2500000	0.1834910
0	3	1	1	0.0195312	0.5	0.6250000	-1.5000000	0.0543613

As one can see, all distributions attain many different values (all the fractions of the joint frequencies for all the possible compositions). All these probability functions are not piecewise monotone. This is an inconvenience if one wants to write sets of most likely index values. For instance, let us observe Figure 1, first row, second column (i.e., the confidence for $p_{B|A} = 0.5$). The most likely value of C is 0.5, and the second most likely value are 0.3333333 and 0.6666667. However, lots of values in the interval $[0.3333333, 0.6666667]$ are very unlikely to happen. What to say, then, about that

interval? We can say that the probability of C to lie on that interval is 0.824941, but it contains very likely values, yet very unlikely ones too.

Another point is variance. Lift and H shows the largest variances in all cases, which is a negative point. The reason might be the range of these index, which is not bound to $[0,1]$. φ has a larger variance than C for $p_{B|A} = 0.5$ but, as mentioned before, and as we shall stress afterwards, confidence is not the *alter ego* of Gras' φ.

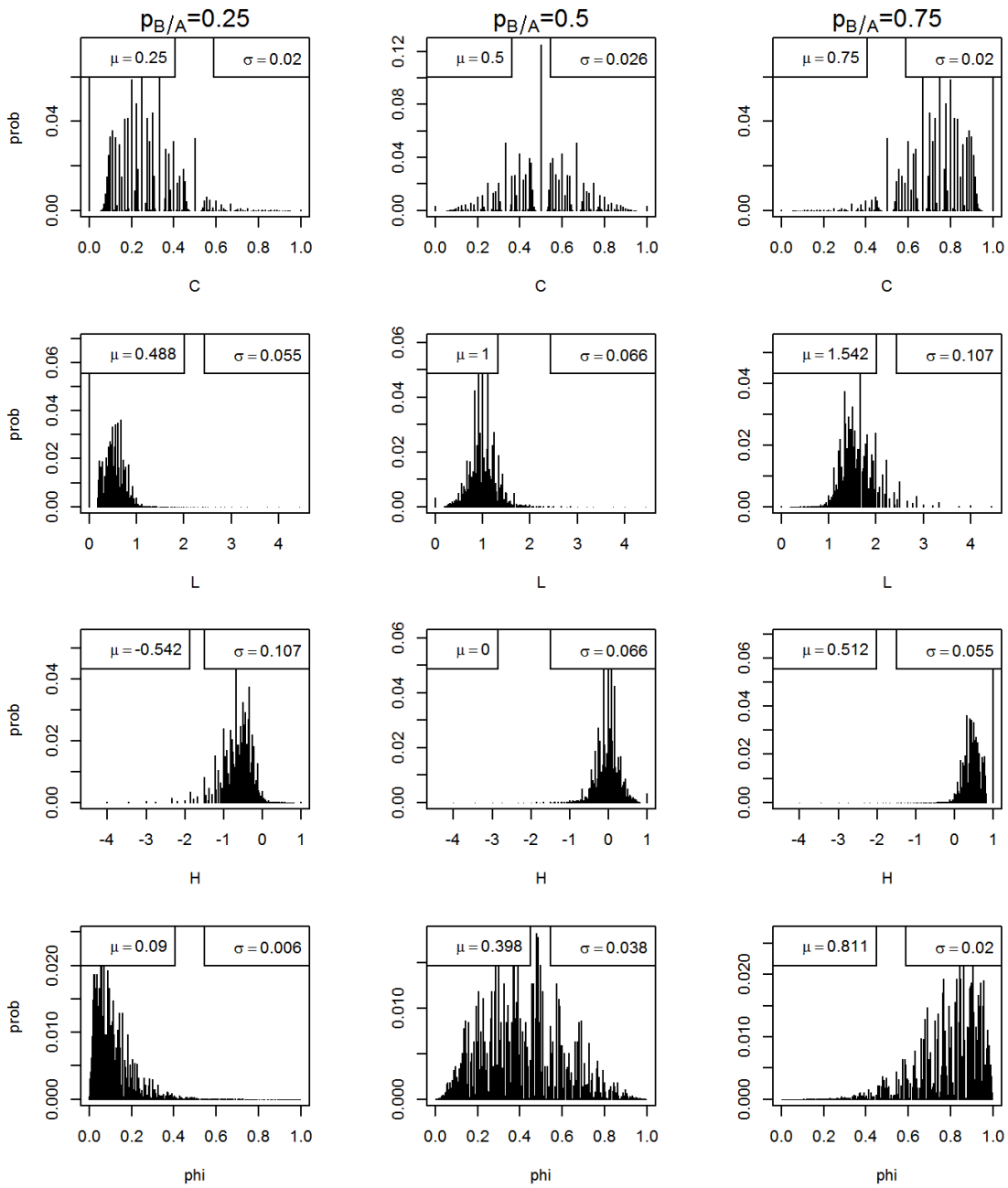


Figure 1 – Probability distribution, expectation and standard deviation of all the quality index for $n = 30$, $p_A = 0.5$, $p_B = 0.5$ and $p_{B|A} = 0.25, 0.5, 0.75$ (from left to right). Ranges of lift and Loevinger's H have been truncated for convenience.

5 Bivariate distributions : plots and correlations

A very interesting approach for the analysis of the relation among several different implication index was presented in Vaillant *et al.* (2008). In that work, a couple of large datasets served as a quarry of some thousands of rules, and several index were compared by plotting them in pairs against the quarry of rules. Some of those plots served as a motivation for the definition of new index.

In our setting, the quarry of rules is the complete bivariate random variable (A,B), with its specific parameters and the fixed sample size. And the goal is to understand the joint behavior of the considered quality measures.

To begin with, let us plot each couple's joint probability distribution, as in Figure 2. It is not complete, since the joint probabilities are not reflected in the plot, but any attempt to jump to 3D plots (with bars showing the joint probabilities) is not satisfactory neither for reflecting the joint distributions.

Another way to understand the relationship among index is to compute correlation coefficients (Pearson's r, Spearman's ρ and Kendall's τ). With these computations, we can see how similarly, or differently, the considered index are measuring the quality of the rules.

In general, such correlation coefficients are computed for two-variate samples of the type

$$\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ \hline y_1 & y_2 & \cdots & y_n \end{array}$$

by:

- Pearson's $r_{xy} := \dfrac{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{(1/n)\sum_{i=1}^{n}(y_i - \bar{y})^2}}$ (linear correlation coefficient).
- Spearman's $\rho_{xy} := r_{zw}$, where z_i and w_i are the ranks of x_i and y_i respectively (i.e., both of them have values, from 1 to n, indicating the rank of original data, when sorted increasingly). Tied ranks are averaged.
- Kendall's $\tau_{xy} := \dfrac{\#\,concordant\ pairs - \#\,discordant\ pairs}{\#\,pairs}$, where concordant pairs are couples $\{(x_i, y_i),(x_j, y_j)\}$ such that $x_i < x_j$ and $y_i < y_j$, or conversely $x_i > x_j$ and $y_i > y_j$. Discordant pairs are those couples such that $x_i < x_j$ and $y_i > y_j$, or conversely $x_i > x_j$ and $y_i < y_j$. In case of ties ($x_i = x_j$ or $y_i = x_j$), those pairs are not concordant neither discordant, and some authors prefer to use modified versions of Kendall's τ (named τ_b and τ_c).

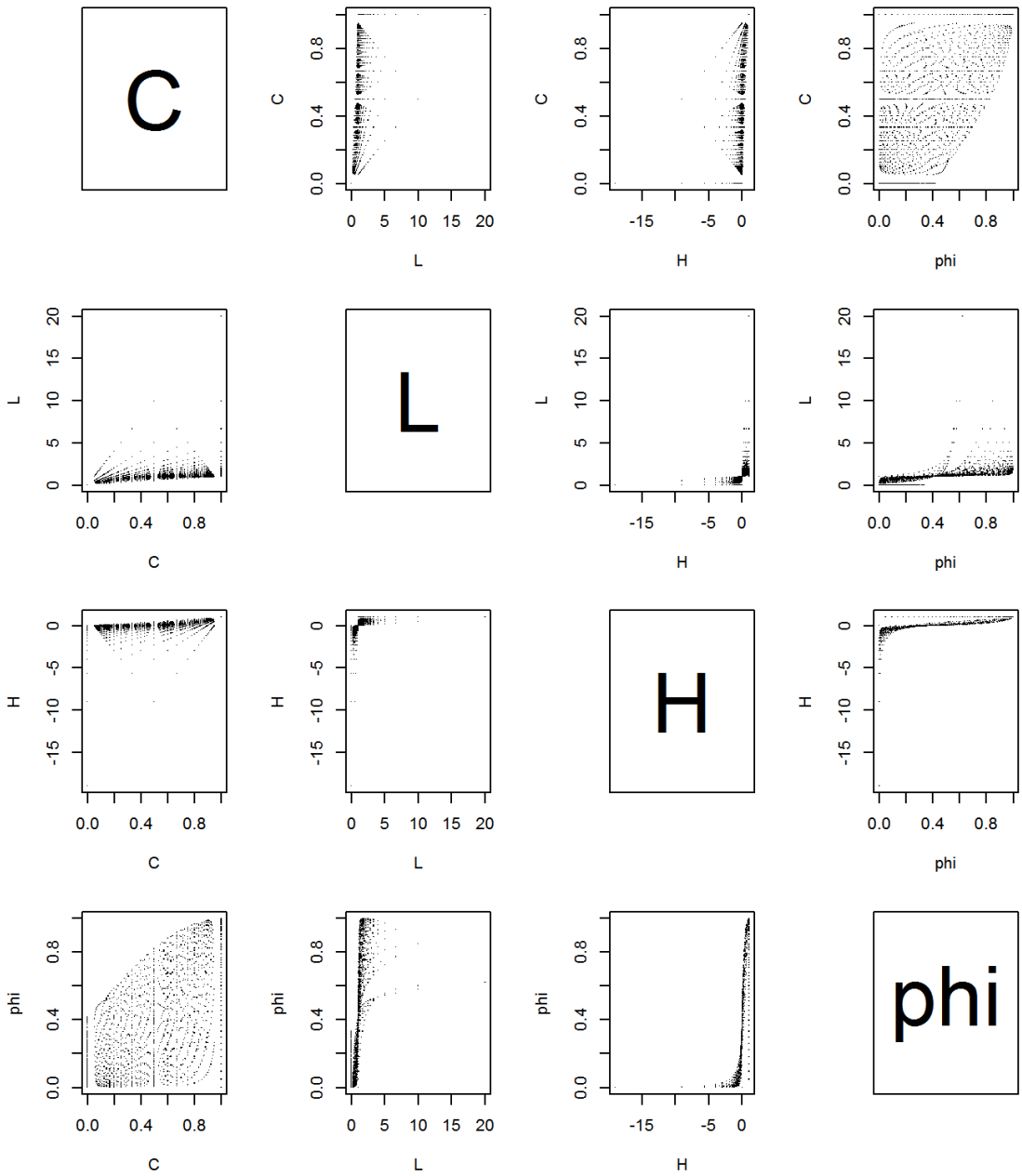


Figure 2 – Plot of joint values of all pairs of quality index, for samples of size $n = 20$, when parameters are $p_A = 0.5$, $p_B = 0.5$ and $p_{B|A} = 0.25$. Some couples of values have a very low probability and some other couples have a rather high probability, but it can hardly be reflected on the plots.

We provide the code to compute correlation coefficients among the random variables C, L, H and φ. Parameters p_A, p_B and $p_{B|A}$ of the binary process, and sample size n, need to be fixed beforehand. We display in Figures 3-5 a few cases, and invite the interested reader to use the code shared in Appendix for computations of any desired case (with moderate values of n since simulations need large memory and are time consuming).

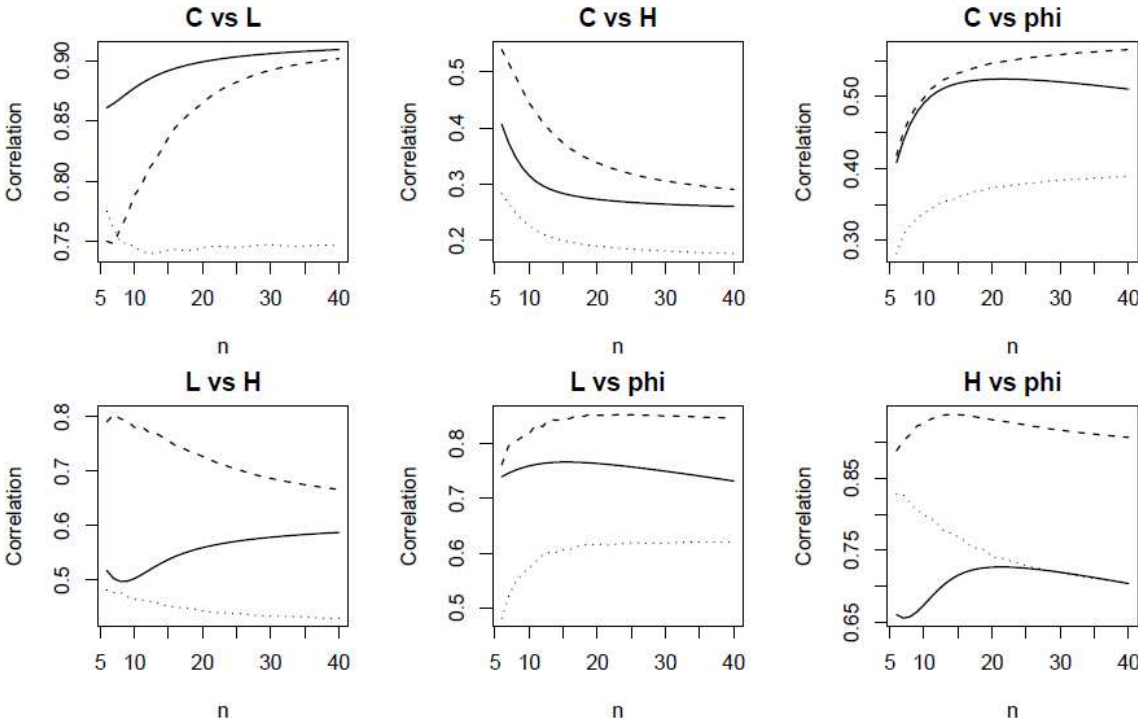


Figure 3 – Evolution, along the sample size, of the Pearson's (continuous line), Spearman's (dashed line) and Kendall's (dotted line) correlation coefficients among all pairs of quality index of the rule $A \rightarrow B$ where $p_A = 0.5$ and $p_B = 0.5$, and conditional probability $p_{B|A} = 0.25$.

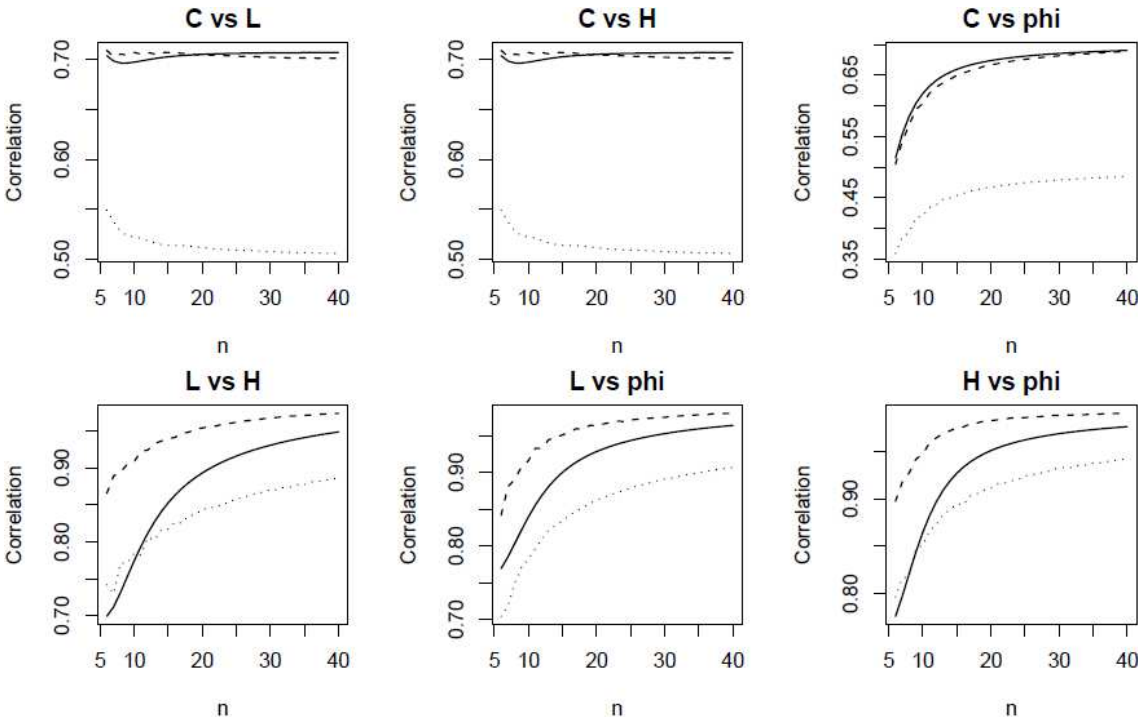


Figure 4 – Evolution, along the sample size, of the Pearson's (continuous line), Spearman's (dashed line) and Kendall's (dotted line) correlation coefficients among all pairs of quality index of the rule $A \rightarrow B$ where $p_A = 0.5$ and $p_B = 0.5$, and conditional probability $p_{B|A} = 0.5$.

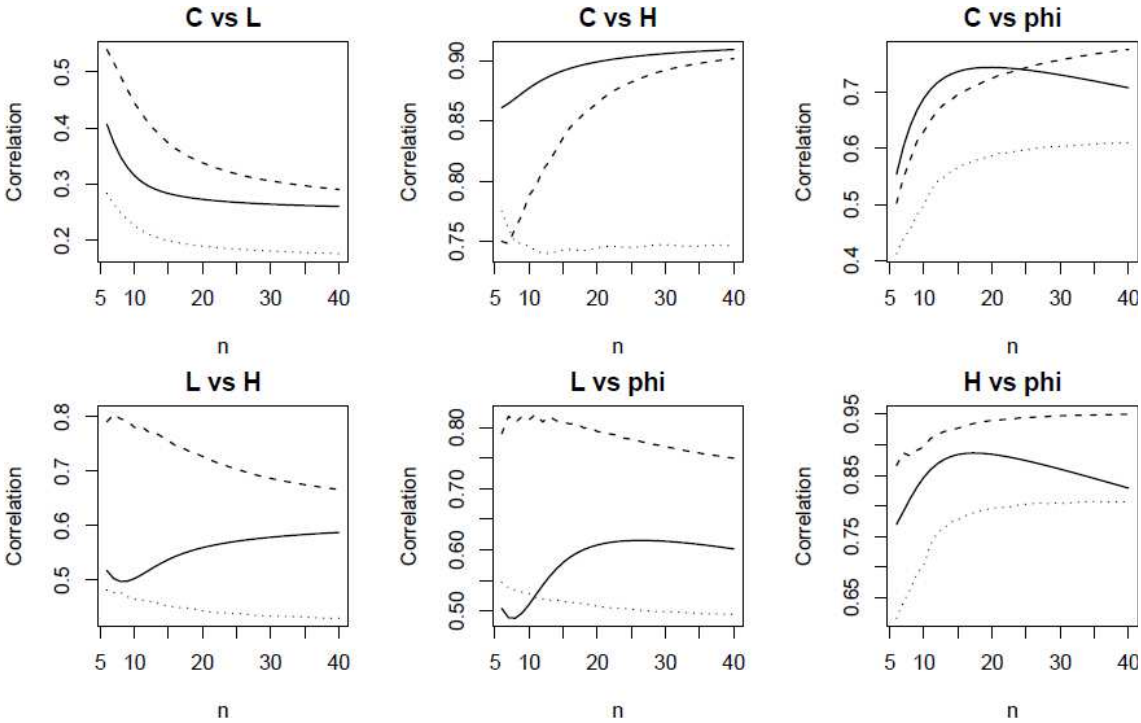


Figure 5 – Evolution, along the sample size, of the Pearson's (continuous line), Spearman's (dashed line) and Kendall's (dotted line) correlation coefficients among all pairs of quality index of the rule $A \rightarrow B$ where $p_A = 0.5$ and $p_B = 0.5$, and conditional probability $p_{B|A} = 0.75$.

Figures 3-5 show, for a range of values of the sample size, the correlation coefficients among all pairs of quality index measured on a bivariate Bernoulli process (A,B). We have used 3 different scenarios, tuned by the conditional probability $p_{B|A}$: low (0.25), average (0.5) and high (0.75). For the conclusions, one can discard the observations for low values of sample size : they are interesting in order to show the trend, but unrealistic cases for the real life applications. Besides, we display the three correlation coefficients in the plots, for the interester reader, but we shall base our discussion on the Spearman's ρ. After carefully examination we conclude that:

- When conditional probability is low ($p_{B|A} = 0.25$), confidence is unexpectedly highly correlated with lift, and (as expected) poorly correlated with Loevinger and Gras' index. The three indexes in the triplet (L, H, φ) keep rather high pairwise correlations, but not as high as one may expect. In most of cases (except for the smallest values) the larger the sample size the lower the correlation, apparently converging to a limiting value.
- When conditional probability is average ($p_{B|A} = 0.5$), the plots are more interesting. The three indexes in the triplet (L, H, φ) keep extraordinary high pairwise correlations, probably because they all measure the *distance from independence*. and this scenario is pure independence. The trend shown by the sample size suggests an increasing limiting *perfect* correlation (under Spearman's eyes). Confidence keeps a moderately good correlation with the rest of indexes.

- When conditional probability is high ($p_{B|A} = 0.75$), the triplet (L, H, φ) loses the extraordinary high pairwise correlations, although they remain rather high. In this case, confidence is surprisingly highly correlated to Loevinger's H, as well as H is with φ. The trend shown by sample size is unclear about stabilization of the values.
- In all cases, an asymptotical result ($n \to \infty$) is an interesting open question for future research.

6 Conclusions

For the simplest bivariate binary random variable (A,B), and moderate sample sizes n, we have provided scripts in R language for computing the probability functions of the rule quality measures confidence, lift, Loevinger's H and Gras' φ, seen as random variables.

We have shown, through plots for particular parameter values, the spread of those random variables with respect to their expected value, giving an idea of how close (or far away) is the value for a particular sample to other possible samples.

We have also computed correlation coefficients among the (random) quality indexes, in order to understand what these indexes are *measuring* indeed. We have found that, when (A,B) are close to statistical independence, lift, Loevinger and Gras are measuring a very similar concept (deviation from independence), and their values are strongly correlated. However, when (A,B) are far from statistical independence, their relation is direct but moderate. In contrast, confidence is in another plane, measuring a different concept. However, we have come across with a rather strong relation between confidence and lift (respectively Loevinger), when (A,B) are away from statistical independence.

With these exploratory results, we have more arguments to suggest that lift, Loevinger's H and Gras' φ do measure such a similar concept. Hence, we think that the researcher can choose freely one or the other without concerns. However we recommend that she, or he, should understand not only the interpretation of extreme values of the index, but also intermediate values:

- Lift is a very popular choice, because it has a very intuitive interpretation : a factor of improvement of prediction of b when a is present.
- However, we do not find an easy or intuitive interpretation of the intermediate values of Loevinger's H (except the trivial value $H = 0$, interpreted as pure independence).
- In our humble opinion, we prefer Gras' φ since any of its values is clearly interpreted from its definition, as a p-value of a hypothesis test against statistical independence.

7 Acknowledgements

The first author acknowledges financial support of research projects P1·1B2012-52 from *Universitat Jaume I de Castellón* and MTM2013-43917-P from *Ministerio de Economía y Competitividad* of Spain.

References

[1] Agrawal, R., T. Imielinski, and A. Swami (1993), Mining association rules between sets of items in large databases, In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207-216).

[2] Bodin, A. (1997), *Analyse implicative : modèles sous-jacents à l'analyse implicative et outils complémentaires*, Prepublication IRMAR No. 97-32, Université de Rennes.

[3] Gras, R., S. Ag Almouloud, M. Bailleul, A. Lahrer, M. Polo, H. Ratsimba-Rajohn, H. and A. Totohasina, (1996), *L'implication statistique*, La Pensée Sauvage, Grenoble.

[4] Gras, R. and R. Couturier (2010), ASpécificité de l'A.S.I. par rapport à d'autres mesures de qualité de règles d'association, *Quaderni di Ricerca in Didattica (Mathematics)*, **20**(supp.1), 175-200.

[5] Gras, R., R. Couturier, J. Blanchard, H. Briand, P. Kuntz and P. Peter (2004), Quelques critères pour une mesure de qualité de règles d'association. Un exemple: l'implication statistique, In H. Briand, M. Sebag, R. Gras and F. Guillet (eds.), *Mesures de qualité pour la fouille de données* (pp. 3-32), RNTI-E-1, Cépaduès éditions.

[6] Gras R., J.-C. Régnier, C. Marinica and F. Guillet (2013), *L'analyse statistique implicative : méthode exploratoire et confirmatoire à la recherche de causalités*, Cépaduès éditions.

[7] Gras R., S. Suzuki, F. Guillet and F. Spagnolo (2008), *Statistical Implicative Analysis, Theory and Applications*, Springer.

[8] Gregori, P., R. Couturier and R. Pazmiño-Mají (2014), On the probability distribution of the classical Gras implication index between two binary random variables, *Educação Matemática Pesquisa*, **16**(3), 969-980.

[9] Hankin, R.K.S. (2006), Additive integer partitions in R, *Journal of Statistical Software, code snippets*, **16**(1), 1-3.

[10] Lallich, S., B. Vaillant and P. Lenca (2007), A probabilistic framework towards the parameterization of association rule interestingness measures, *Methodology and Computing in Applied Probability*, **9**(3), 447-463.

[11] Loevinger, J. (1947), A systematic approach to the construction and evaluation of tests of ability, *Psychological Monographs*, **61**(4), i-49.

[12] Vaillant, B., S. Lallich, and P. Lenca (2008), On the behaviour of the generalisations of the intensity of implication: a data-driven comparative study, In

R. Gras, E. Suzuki, F. Guillet and F. Spagnolo (eds.), *Statistical Implicative Analysis: Theory and Applications* (pp. 421-448), Studies in Computational Intelligence, 127, Springer-Verlag, Berlin Heidelberg.

Appendix : code blocks

Code block for the computation of all considered index:

```
funindex = function(x, index) {
  # computes quality 'index' for a given contingency table,
  # Arguments:
  #   x: cell counts, in the form c(n00, n01, n10, n11)
  #   index: name of the desired index
  n00=x[1]; n01=x[2]; n10=x[3]; n11=x[4];
  if(index=='conf') { # confidence
    if(n11+n10 !=0 ) res = n11/(n11+n10)
    else res=NA
  }
  if(index=='lift') { # lift
    if(((n11+n10)*(n11+n01)) !=0 )
      res = sum(x)*n11/((n11+n10)*(n11+n01))
    else res=NA
  }
  if(index=='loev') { # Loevinger's H
    if(((n11+n10)*(n00+n10)) !=0 )
      res = 1 - sum(x)*n10/((n11+n10)*(n00+n10))
    else res=NA
  }
  if(index=='gras') { # Gras phi binomial model
    res = 1-pbinom(q=n10, size=sum(x),
                   prob=((n10+n11)*(n00+n10))/((sum(x))^2))
  }
  return(res)
}
```

Code block for the computation of the probability functions in fixed sample size.

```
pf.index = function(index='gras', p=rep(1,4),
                    pX=NULL, pY=NULL, pYgivenX=NULL,
                    size=4) {
  # It computes the probability function of any 'index'
  # Arguments:
  #   index: 'conf', 'lift', 'loev' or 'gras', for the confidence,
  #     lift, Loevinger's H or Gras index.
  #   p: vector of 4 probabilities c(p00, p01, p10, p11)
  #     or proportional.
  #   pX, pY, pYgivenX: alternative to 'p', marginal and
  #     conditional probability.
  #   size: sample size.
```

```
  # Value: a data frame with sample space and probability function
  #   of the random variable index, as a function of the random
  #   bivariate Bernoulli with parameters 'p' or pX, etc.
  require(partitions)
  N = size
  if( !(is.null(pX) & is.null(pY) & is.null(pYgivenX)) ) {
    p11 = pYgivenX * pX
    p10 = pX - p11
    p01 = pY - p11
    p00 = 1 - (p01+p10+p11)
    p = c(p00, p01, p10, p11)
  }
  NN = compositions(N,4) # all partition of N into 4 numbers
  # compute index value for all possible partition
  index.NN = apply(X=NN, MARGIN=2, FUN='funindex', index=index)
  # compute probability of all such partition
  pr.NN = apply(X=NN, MARGIN=2, FUN='dmultinom', size=N, prob=p)
  # REMARK: computed only for fixed sample size
  res = aggregate(x=pr.NN, by=list(index.NN), FUN='sum')
  # NA values (and their probability) are discarded
  names(res) = c(index, 'prob')
  return(res)
}
```

Code block for the computation of the bivariate probability functions in fixed sample size.

```
pf.2index = function(index=c('conf', 'conf'), p=rep(1,4),
                     pX=NULL, pY=NULL, pYgivenX=NULL,
                     size=4) {
  require(partitions)
  N = size
  if( !(is.null(pX) & is.null(pY) & is.null(pYgivenX)) ) {
    p11 = pYgivenX * pX
    p10 = pX - p11
    p01 = pY - p11
    p00 = 1 - (p01+p10+p11)
    p = c(p00, p01, p10, p11)
  }
  NN = compositions(N,4) # all partitions of N into 4 numbers
  # probability of all such partition
  pr.NN = apply(X=NN, MARGIN=2, FUN='dmultinom', size=N, prob=p)
  # REMARK: computed only for fixed sample size
  ind1 = apply(X=NN, MARGIN=2, FUN='funindex', index=index[1])
  ind2 = apply(X=NN, MARGIN=2, FUN='funindex', index=index[2])
  res = aggregate(x=pr.NN, by=list(ind1, ind2), FUN='sum')
  # NA values (and their probability) are discarded
  names(res) = c(index, 'prob')
  return(res)
}
```

Code block for the computation of the correlation coefficients.

```
cor.index = function(index = c('conf', 'conf'),
                     p=rep(1,4), pX=NULL, pY=NULL,
                     pYgivenX=NULL, size=4,
                     method='pearson') {
  if( !(is.null(pX) & is.null(pY) & is.null(pYgivenX)) ) {
    p11 = pYgivenX * pX
    p10 = pX - p11
    p01 = pY - p11
    p00 = 1 - (p01+p10+p11)
    p = c(p00, p01, p10, p11)
  }
  pf2 = pf.2index(index=index, p=p, size=size)
  ind1 = pf2[[ index[1] ]]
  ind2 = pf2[[ index[2] ]]
  pr.NN = pf2$prob/sum(pf2$prob)
  if( method=='pearson') {
    cor = ( sum(ind1*ind2*pr.NN) - sum(ind1*pr.NN)*sum(ind2*pr.NN) ) /
      ( sqrt( (sum(ind1^2*pr.NN) - sum(ind1*pr.NN)^2) *
                (sum(ind2^2*pr.NN) - sum(ind2*pr.NN)^2)) )
  }
  if( method=='kendall' ) {
    ind1.ij = outer(ind1, ind1, '-')
    ind2.ij = outer(ind2, ind2, '-')
    pr.ij = outer(pr.NN, pr.NN)
    conc.disc = sum(sign(ind1.ij * ind2.ij) * pr.ij)
    cor = sum(conc.disc)/(1-sum((pr.ij[ ind1.ij==0 & ind2.ij==0 ])))
  }
  if( method=='spearman') {
    ind1 = rank(ind1)
    ind2 = rank(ind2)
    cor = ( sum(ind1*ind2*pr.NN) - sum(ind1*pr.NN)*sum(ind2*pr.NN) ) /
      ( sqrt( (sum(ind1^2*pr.NN) - sum(ind1*pr.NN)^2) *
                (sum(ind2^2*pr.NN) - sum(ind2*pr.NN)^2)) )
  }
  return( list(cor=cor, method=method, index=index,
               size=size, p=p, cor=cor) )
}
```

POURQUOI ET COMMENT TRANSFORMER DES VARIABLES QUANTITATIVES EN CATEGORIELLES ? APPLICATION A L'INTONATION DE LA LANGUE FRANÇAISE

Martine CADOT[1] et Anne BONNEAU[2]

TRANSFORMING QUANTITATIVE VARIABLES INTO QUALITATIVE ONES: RATIONALE AND METHOD. APPLICATION TO FRENCH INTONATION.

RÉSUMÉ

L'interprétation d'une phrase en français dépend non seulement de ses mots, mais aussi de son intonation : interrogative, déclarative, dubitative, etc. Notre but est de construire un modèle statistique permettant de différencier les types intonatifs d'une phrase prononcée par un locuteur à partir des indices acoustiques issus de son enregistrement sonore. Pour étudier le lien entre la variable catégorielle « type de phrase » et les variables quantitatives d'indices acoustiques, il a fallu transformer ces dernières en variables catégorielles. Nous discutons ici des raisons statistiques qui peuvent imposer cette transformation et de la façon d'y procéder en nous plaçant d'un point de vue plus général et théorique avant d'appliquer ce formalisme aux données expérimentales recueillies.

Mots-clés : *Implication statistique, liaisons complexes entre variables, fouille robuste de données, recherche de modèle de discrimination, découpage de variables quantitatives, intonation, indices acoustiques, courbes mélodiques, phonétique.*

ABSTRACT

The interpretation of a French oral sentence depends not only on its word sequence but also on its intonation : interrogative, declarative, doubtful .... Our aim is build a statistical model able to differentiate various intonations and relying upon the acoustic cues extracted from sentences pronounced by French speakers. To study the connection between the categorical variable "type of sentence" and the quantitative variables stemming from the various acoustic cues considered in this study, it was necessary to transform the latter into qualitative variables. We discuss the statistical reasons that can impose this transformation and the way to carry it by considering a more general and theoretical point of view before applying this formalism to the collected experimental data..

Keywords : *statistical implication, complex links between variables, robust data mining, discriminant model research, quantitative versus qualitative variables, intonation, acoustic cues, melodic curves, phonetic.*

1 Introduction

Notre but est de construire un modèle statistique permettant de différencier les types intonatifs d'une phrase prononcée par un locuteur à partir des indices acoustiques de son enregistrement sonore. Le modèle que nous visons est de la forme X→Y, se lisant selon les cas : X implique Y, X cause Y, X explique Y, X produit Y, etc. (Gras *et al.*, 2013)

[1] LORIA/UdL, 54600 Villers-lès-Nancy, France, martine.cadot@loria.fr
[2] LORIA/CNRS, 54600 Villers-lès-Nancy, France, anne.bonneau@loria.fr

où X représente les indices acoustiques et Y le type intonatif de la phrase. Il est « asymétrique » dans la mesure où les variables n'ont pas toutes le même rôle[3] : la variable Y est la « variable à expliquer » (appelée également variable dépendante), et les variables acoustiques X sont les « variables explicatives » (appelées également variables indépendantes). Pour construire ce modèle nous avons utilisé une partie des données recueillies selon un « plan d'expérience » pré-établi, nous avons créé 29 variables acoustiques quantitatives X, 3 variables qualitatives (appelées également catégorielles) Z à contrôler (telles que le genre : M/F, dont il faut contrôler les effets sur le modèle pour les retirer le cas échéant[4]) et la variable catégorielle de contour mélodique Y dont nous n'avons gardé que 3 modalités.

Dans une première partie, plus théorique, nous réfléchissons à ce qu'est une liaison entre une variable catégorielle Y et un ensemble de variables quantitatives X, en nous restreignant toutefois à ce qui sera applicable à nos données et à notre problématique. Nous voyons comment cette liaison peut s'exprimer quantitativement de diverses façons, notamment après recodages des variables quantitatives, et être validée statistiquement. Pour rendre cet exposé plus lisible, nous l'avons illustré à l'aide d'un jeu de données bien connu, « les iris de Fisher » (Fisher 1936), formé des valeurs de 5 variables (X : 4 mesures de fleurs, et Y : la catégorie de la fleur parmi 3 catégories) recueillies sur 150 fleurs. À partir de cette réflexion nous donnons les raisons qui nous ont fait choisir l'ASI (Gras *et al.* 1996) et le logiciel CHIC (Version 6.0, Copyright© 2012) pour traiter nos données.

Dans la deuxième partie, plus applicative, nous exposons notre problématique, relative au rôle de l'intonation dans l'interprétation d'une phrase enregistrée en français, le plan d'expériences choisi ainsi que le prétraitement des données, puis nous montrons les résultats de l'utilisation de CHIC pour la mise au point de notre modèle sur ces données. A cette occasion, nous comparons les modèles obtenus avec CHIC en reprenant la réflexion de la première partie et notamment les divers recodages des données quantitatives.

A part CHIC pour l'ASI, et le tableur Excel pour la régression linéaire, le tableau de contingence et une partie des graphiques, tous les modèles testés l'ont été à l'aide de scripts écrits dans le langage/logiciel R (disponible gratuitement sous licence GNU http://www.r-project.org/), dont le code est mis en annexe.

2 Liaison statistique entre une variable catégorielle et des variables quantitatives

Dans cette section, nous présentons d'abord le jeu de données qui va nous permettre d'illustrer les modèles, puis les divers modèles existants, avant de choisir celui que nous utiliserons.

[3] Pour un exposé détaillé sur la planification des variables en psychologie expérimentale voir Hoc (1983), en agronomie voir Dagnelie.(2003).

[4] Pour un exposé détaillé sur la gestion des variables à contrôler en psychologie expérimentale voir Léon *et al.* (1977), en médecine voir Schwartz (1983).

2.1 Les iris de Fisher

Ce jeu de données est téléchargeable depuis UCI repository (https://archive.ics.uci.edu/ml/datasets/Iris). Il est formé d'une variable catégorielle Y (l'espèce) à 3 modalités (iris-setosa, iris-versicolor, iris-virginica), avec 50 fleurs de chaque espèce, et d'un ensemble de 4 variables numériques X, qui sont la longueur et la largeur moyennes des sépales et des pétales (Sepal_Length, Sepal_Width, Petal_L, Petal_W). Il a été utilisé par de nombreux chercheurs en analyse de données désirant éprouver une nouvelle méthode statistique, soit en classification pour retrouver les 3 groupes de fleurs, soit en discrimination pour trouver les règles d'attribution d'une fleur à un groupe. C'est dans ce deuxième cadre que nous nous plaçons. Dans la figure 1, nous avons représenté les 150 fleurs dans l'espace formé de 3 variables de X (Petal_L, Petal_W et Sepal_W), avec une couleur par espèce.

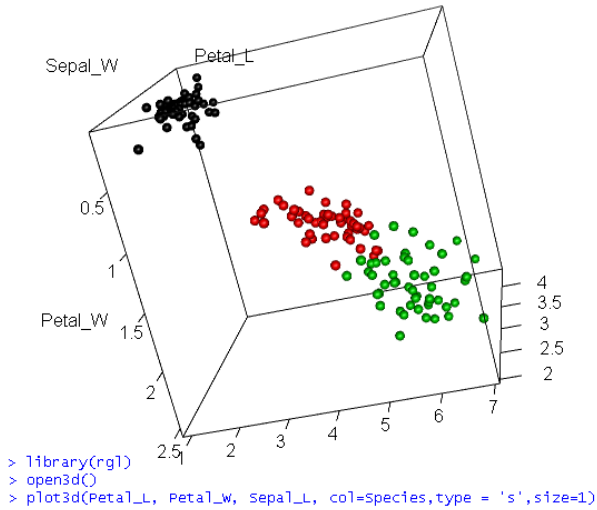


Figure 1 – Les 150 iris de Fisher dans l'espace de 3 variables, avec une couleur par espèce

On peut remarquer que la longueur et la largeur des pétales permettent d'affecter plus ou moins facilement une espèce à chaque iris : les plus petits pétales correspondent à l'espèce Setosa (en noir), les moyens à l'espèce Versicolor (en rouge) et les plus grands à l'espèce Virginica (en vert). Plus précisément, parmi les 3 nuages de points de couleurs différentes de la figure 1, seuls le nuage de points rouges et celui de points verts ont une petite partie commune d'une dizaine d'individus. Dans la mesure où un simple découpage en quelques sous-espaces de la figure 1 permet d'établir l'espèce des 150 iris avec une dizaine d'erreurs, on attend de chacune des méthodes de discrimination étudiées qu'elles atteignent des taux de reconnaissance de plus de 90%.

2.2 Les différents modèles possibles

Nous nous plaçons ici dans le cadre de la prédiction : connaissant la valeur de X et celle de Y sur un ensemble de données, nous recherchons une méthode qui permette de trouver Y_{pred} (prédiction de Y) à partir de X, en faisant le moins possible d'erreurs (Erreur=$Y-Y_{pred}$). Pour que cette méthode puisse être élevée au rang de modèle, il faut qu'elle soit suffisamment détachée des données de départ, simple et intelligible. Nous faisons un tour rapide des modèles prédictifs courants en donnant pour chacun une référence renvoyant à un manuel pédagogique dans lequel cette méthode est détaillée

parmi d'autres, une partie restreinte de ces références pouvant donc suffire pour couvrir l'ensemble des modèles décrits.

2.2.1 Les modèles statistiques

Un des modèles les plus utilisés est le ***modèle linéaire*** (Prum 1996). Sa version de base est le ***modèle de régression linéaire***, pour lequel X et Y sont deux variables quantitatives. Il se résume à la connaissance des 2 coefficients a et b de la ***droite de régression*** Y_{pred}=aX+b sur laquelle sont situés les points de coordonnées (X, Y_{pred}) dans le plan de X et Y (par exemple la première équation à droite de la figure 2 exprime la liaison entre la longueur prédite du pétale et la longueur du sépale pour les iris setosa, et elle est représentée par la droite bleue du graphique). Quand X est formé de p variables quantitatives (p>1), le modèle de régression linéaire s'écrit Y_{pred}=XA+B avec A et B des vecteurs de p valeurs fixées, et c'est l'équation d'un hyper-plan dans l'espace de dimension p+1 de X et Y.

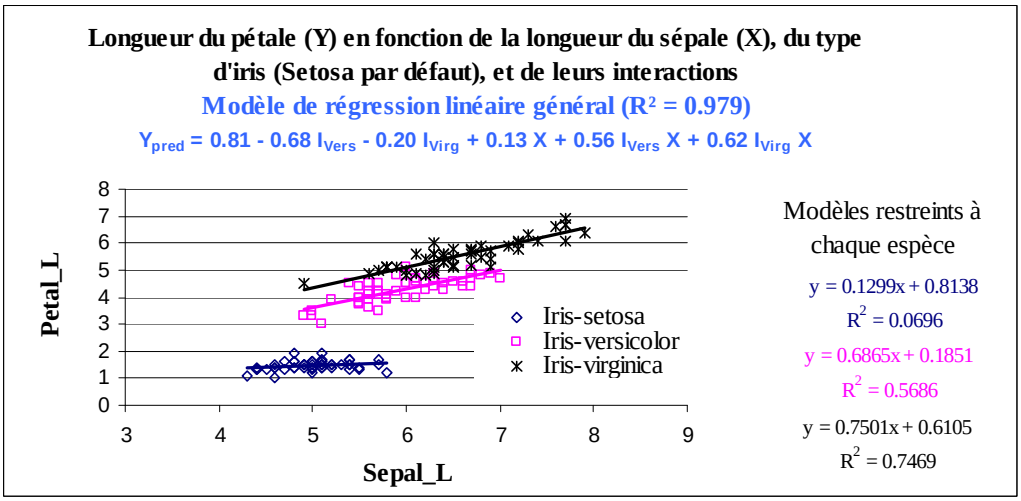


Figure 2 – Un modèle de régression de la longueur du pétale sur la longueur du sépale et l'espèce d'iris. En haut, en bleu le modèle global prenant en compte les 150 mesures, et à droite les 3 modèles spécifiques à chaque type d'iris, ne prenant en compte que 50 mesures chacun.

La régression linéaire peut s'étendre à des variables X non toutes quantitatives et on parlera plutôt de ***modèle linéaire généralisé*** (Baillargeon 2000). Par exemple en Figure 2, on a exprimé la dépendance de Y (longueur des pétales) en fonction de X (petal_L, la longueur des pétales, I_{vers} et I_{virg} les variables indicatrices de 2 types d'iris (respectivement iris-versicolor et iris-verginica, le troisième type étant la référence par défaut, obtenu quand les deux variables sont nulles simultanément). Si au lieu d'être quantitative, Y est binaire (1 : réussite, 0 : échec), on utilise le modèle ***logistique*** (Besse 2003) pour lequel l'équation devient logit(P_{est})=XA+B, où P_{est} est la probabilité que Y soit égal à 1, et logit(P)= log(P/(1-P)). Le modèle logistique peut encore s'étendre à une variable Y de comptage (Y : nombre de réussites) ou ordinale, mais plus difficilement à une variable catégorielle. Toutefois, si on « éclate » en p variables binaires la variable catégorielle Y à p catégories, on peut alors chercher p modèles logistiques, un par variable binaire. Une extension dans une autre direction est le modèle ***log-linéaire*** (Morineau, 1996), pour lequel toutes les variables sont catégorielles, X comme Y, et c'est alors chaque variable de X qu'il convient d'éclater en intervalles, ces intervalles devenant les catégories de la variable, et dans ce modèle, ce n'est plus Y qui dépend

linéairement de X mais log(Effectif$_{pred}$) qui est une fonction linéaire de X et Y. La ***discrimination linéaire*** (Nakache 2003) est un modèle un peu différent, elle consiste à trouver un changement de repère dans X pour lequel les catégories de Y sont séparées le mieux possible. En Figure 3, on a représenté la projection des 150 iris dans l'espace des deux hyperplans obtenus pour la discrimination linéaire des 3 espèces d'Iris (un seul aurait suffit, car il produit plus de 99% de la variance).

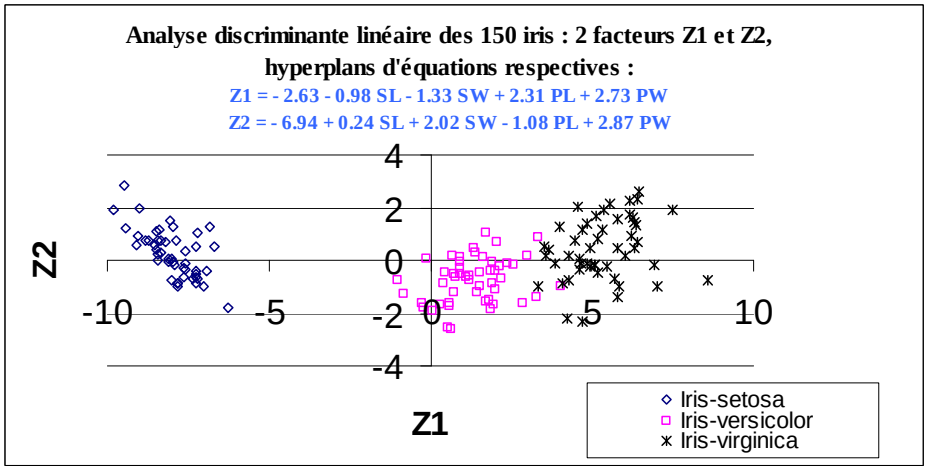


Figure 3 – Analyse discriminante des 150 iris : plus de 99% de variance selon Z1.

Le « modèle linéaire » et ses extensions, sont issus des statistiques classiques. Ils font partie des ***modèles paramétriques*** et sont assortis de conditions d'application qui permettent de garantir la qualité des estimations produites. Quand les conditions d'application ne sont pas vérifiées (par exemple la normalité des erreurs), des modèles non paramétriques (Siegel 1988) peuvent être utilisés, par exemple, si le nombre d'ex-aequo est réduit, en remplaçant les variables par leurs rangs (Droesbeke, 1996).

2.2.2 Les modèles issus de l'intelligence artificielle

A côté de ces méthodes qui, depuis plus d'un siècle, ont permis de produire des prédictions « garanties » à partir de mesures faites sur un seul échantillon choisi avec soin (Morin 1999), depuis quelques dizaines d'années de nouvelles méthodes de prédiction sont apparues avec l'avènement de l'informatique, qui se sont généralisées dernièrement avec le libre accès à des ressources sur Internet. Issues de « l'apprentissage automatique » (Mitchell 1997) ces méthodes se proposent d'évaluer la qualité des prédictions en partitionnant l'ensemble des données en plusieurs parties de diverses façons (par exemple k parties pour la ***validation croisée***, le modèle étant construit sur la réunion de k-1 parties, testé sur la partie restante, on itère le processus k fois en changeant la partie restante). Le nombre de ces méthodes est en constante augmentation (plus de 6500 packages R disponibles à ce jour (http://cran.r-project.org/), dont une grande partie développant des méthodes statistiques), dans la mesure où leur validité peut s'établir sans avoir besoin de théories statistiques ***asymptotiques*** (par exemple s'appuyant sur l'hypothèse de normalité), mais par une mise à l'épreuve sur des jeux de données caractéristiques d'un problème spécifique à résoudre.

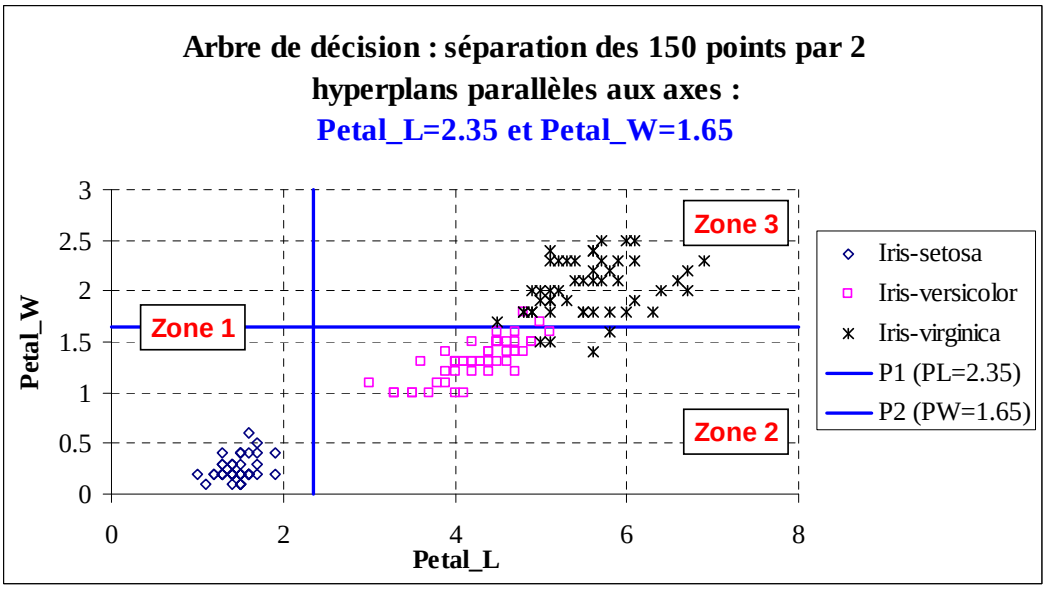


Figure 4 – Résultat sur les 150 iris d'un arbre de décision entraîné une fois sur 75 iris tirés au hasard (25 par classe) : 2 hyperplans séparateurs et 6 erreurs (2 pour la partie apprentissage et 4 pour l'autre).

Nous ne nous intéresserons ici qu'aux méthodes qui peuvent prédire une variable Y catégorielle de plus de 2 catégories à partir d'un ensemble de variables X quantitatives, et donc s'appliquer au jeu de données des iris de Fisher. Les plus pratiquées sont des extensions des réseaux neuronaux (Ripley 1996), les MSVM (Multiple Support Vector Machine, Weston 1998) et les arbres de décision (Breiman 1984). Dans la Figure 4, nous montrons le partage en régions proposé par un arbre de décision sur les données iris. Deux hyperplans parallèles aux axes ont suffi pour déterminer les 3 zones de points formées quasi-exclusivement d'une seule espèce d'Iris. La zone 1 correspond à Petal_L<2.35 et ne contient que les iris Setosa. La zone 2 correspond à Petal_L≥2.35 et Petal_W<1.65, et contient tous les iris Versicolor, sauf 1, ainsi que 5 iris Virginica. Et dernière zone correspondant à Petal_L≥2.35 et Petal_W≥1.65 contient la quasi-totalité des iris Virginica, ainsi qu'un iris Versicolor. Il y a 6 erreurs sur 150 iris, soit 4% d'erreurs. Du seul point de vue de l'interprétation graphique, on peut dire que les MSVM sont une extension des arbres de décision, avec des hyperplans frontières qui ne sont plus nécessairement parallèles aux axes, des frontières qui ne sont plus strictes, mais floues (ce sont des bandes), et dans un espace qui est une extension de X.

Signalons pour finir l'extraction ***des règles d'association***, pour lesquelles les méthodes de validation sont encore à l'état de recherches (Cadot 2006). Partant d'un jeu de données formé des valeurs de q variables catégorielles Xi sur un certain nombre de sujets, la méthode extrait automatiquement des règles de la forme (A et B et C)→D, où le membre de gauche contient un certain nombre d'affirmations (ici il y en a 3, A, B et C) du type Xi=Cik (la variable Xi a pour valeur sa catégorie Cik), celui de droite n'en contenant qu'une en général. Toute combinaison de variables et de catégories fournissant une règle, leur nombre augmente de façon exponentielle avec le nombre de variables, et le défi est de trouver le (ou les) indices de qualité permettant de ne garder que les meilleures. Si on ne garde que les règles ayant Y en partie droite, et si on a transformé les variables quantitatives en catégorielles pour les utiliser en partie gauche, cette méthode peut répondre à notre recherche en nous fournissant des règles de discrimination.

2.2.3 Où se situe l'ASI ?

Dans sa version originale, l'ASI s'applique à des variables binaires, qu'on obtient souvent par éclatement de variables catégorielles (la variable Genre avec 2 catégories F et M, donnera par éclatement 2 variables binaires F et M). Les règles fournies ressemblent donc aux règles d'association (RA) mais la méthode présente en outre 2 avantages : 1) l'interface graphique de CHIC permet de « voir » les règles et de les manipuler 2) seules les règles valides s'affichent, la validité des règles étant établie selon la théorie statistique asymptotique (comptages vérifiant la loi de poisson ou binomiale, selon un choix à préciser dans les options de CHIC).

Dans la version actuelle, on peut utiliser des variables quantitatives telles quelles, ou les faire recoder en variables catégorielles. Nous avons utilisé les données des iris pour confronter les réponses que l'ASI offre à notre problème de discrimination X→Y. Nous avons d'abord utilisé les variables sans les découper mais en les remplaçant par une valeur entre 0 et 1 selon la formule indiquée dans l'aide :

valeur_nouvelle=(valeur_ancienne-min_valeurs)/(max_valeurs-min_valeurs),

Puis nous les avons fait découper en 2, 3, ou 4 parties par le logiciel CHIC. Le graphe implicatif des variables quantitatives non recodées, seulement réajustées pour être dans l'intervalle [0 ; 1] est en figure 5, à gauche, et à droite on a celui obtenu avec les variables découpées en 3 par CHIC (les 2 autres découpages testés, qui se sont avérés moins bons, ont été mis en annexe).

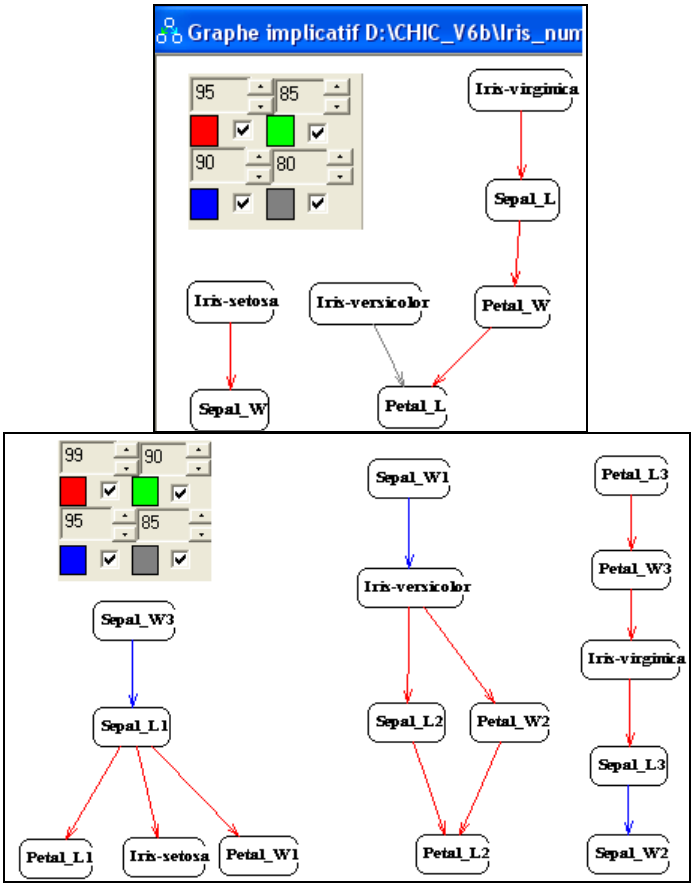


Figure 5 – Graphes implicatifs produits par CHIC sur les 150 iris. À gauche les 4 variables quantitatives à valeurs réajustées dans [0 ; 1], à droite recodées en 3 catégories chacune par CHIC.

La première différence entre les deux graphiques de la figure 5 porte sur les seuils. Comme nous voulions faire apparaître la totalité des 3 catégories, dans le graphique de gauche, nous avons dû descendre jusqu'à un seuil d'implication statistique de 0.80 pour faire apparaître Iris-versicolor, ce qui fournit donc une règle de piètre qualité pour cette espèce. La recherche de règles de bonne qualité pour chaque espèce d'iris est la première raison qui nous fait préférer le graphique de droite.

La deuxième raison qui nous fait préférer le graphique de droite est que chaque espèce apparaît dans un graphe séparé. Cette séparation exprime graphiquement ce que les autres modèles nous ont indiqué dans des tables de confusion croisant les espèces estimées par le modèle et les espèces observées : connaissant les mesures des pétales et des sépales, on peut en déduire avec peu d'erreurs l'espèce de l'iris. Comparons maintenant la figure 5 à la figure 2. La régression de Petal_L sur Sepal_L n'était significative que pour iris-virginica et iris-versicolor, pas pour iris-setosa, comme on peut le voir dans les équations à droite du graphique de la figure 2 et nous retrouvons un peu cela dans le graphique de gauche de la figure 5, dans la mesure où les 2 variables Petal_L et Sepal_L sont reliées entre elles fortement (p>0.95) et également avec les 2 espèces d'iris convenables. Le modèle de régression linéaire nous en dit plus, notamment, dans l'équation globale, les coefficients qui diffèrent significativement de 0 sont les seules interactions entre Sepal_L et chacune des 2 catégories. Mais ce qu'il dit est d'interprétation difficile, et si l'on faisait la régression pour chaque sous-ensemble de variables cela poserait des problèmes d'interprétation encore plus complexes. Si on prend maintenant le graphique de droite de la figure 5 nous pouvons dire que la liaison des variables Sepal_L et Petal_L est forte car chaque fois qu'elles sont reliées par une flèche, c'est avec la même catégorie (1, 2 ou 3), et à chaque catégorie correspond une espèce d'iris. Sauf pour l'espèce setosa, un chemin rouge (seuil=0.99) fait de 2 flèches qui se suivent joint l'espèce et les 2 catégories de Sepal_L et Petal_L, on retrouve ainsi les éléments déjà cité de la figure 2, mais en plus on a une information fine sur la relation entre les espèces et les variables Sepal_L et Petal_L : les valeurs les plus faibles des 2 variables sont pour iris-setosa, les valeurs intermédiaires pour iris-versicolor et les plus fortes pour iris-virginica. Et on voit dans ces 3 graphes que les niveaux de Petal_W sont associés aux espèces de la même façon. Par contre la variable Sepal_W se comporte différemment : le niveau le plus faible est associé à versicolor, le moyen à virgina et le plus fort à setosa, et ces associations sont plus faibles (entre 0.95 et 0.99).

Jusqu'ici, la balance penche fortement du côté du graphique de droite de la figure 5 : Il nous donne des informations qui rejoignent celles que nous avons trouvées avec les autres modèles donnés en exemples, ainsi que des informations supplémentaires que nous n'avons pas su extraire des autres modèles, pour autant qu'elles y soient.

Maintenant se pose la question de l'interprétation du sens des flèches, et notamment : pourquoi la position en hauteur de l'espèce diffère-t-elle d'un graphe à l'autre dans le graphique de droite de la figure 5. Faut-il interpréter différemment ces 3 graphes ? Ce qui nous renvoie à la sémantique de ces flèches. Dans l'ouvrage collectif sur l'ASI (Gras *et al*, 2013) nous lisons dans l'introduction (Gras et Regnier, section 4, p. 16-17) que cette flèche peut indiquer une inclusion plutôt qu'une causalité, ce que nous appellerons un « effet d'effectif ».

niveau	Sepal_L	Sepal_W	Petal_L	Petal_W
1	46	47	50	50
2	53	67	54	52
3	51	36	46	48

Tableau 1 – Effectifs selon les catégories obtenues en découpant chaque variable quantitative en 3 par CHIC (ces « informations sur le fichier » figurent sous cette dénominations dans les sorties de CHIC)

Avant donc d'essayer d'interpréter le sens des flèches pour les iris, examinons la relation entre l'ordre des effectifs et la hauteur dans le graphe. Dans le tableau 1 on a reporté les effectifs de chacune des 3 catégories des variables selon le découpage fait par CHIC. On voit que la variable Sepal_W a des catégories déséquilibrées, ce qui n'est pas le cas des 3 autres variables, dont les effectifs restent entre 46 et 54. Dans le niveau 3, si on range ces 3 variables ainsi que l'espèce par effectif croissant, on obtient (le symbole « < » se lira « précède »):

Petal_L (46) < Petal_W (48) < Iris-virginica (50) < Sepal_L (51)

C'est exactement dans cet ordre que les flèches successives joignent les items du 3ème graphe, figure 5, à droite. Pour le niveau 2, on a :

Iris-versicolor (50) < Petal_W (52) < Sepal_L (53) < Petal_L (54),

Et c'est quasiment dans cet ordre qu'on trouve le graphe 2, sauf que Petal_W et Sepal_L (53) ne sont pas l'un sous l'autre mais l'un à côté de l'autre. Pour le premier graphe, on retrouve exactement l'ordre des effectifs.

Dans notre exemple, le sens des flèches pourrait ainsi être un effet d'effectif. D'ailleurs, il paraît difficile de justifier sémantiquement un chemin de causalité différent pour chaque espèce d'iris, par exemple le fait d'avoir l'implication « sépales courtes → iris-setosa » et l'implication inverse « iris-virginica→sépales longues ». Nous décidons donc de ne pas essayer d'interpréter en termes de causalité le sens des flèches dans les graphes implicatifs.

Nous essayons maintenant d'en déduire des règles de discrimination. Pour cela, nous utilisons le fichier de données recodées après découpage en 3 des variables de X par CHIC, et nous en extrayons sous Excel un tableau de contingence (voir tableau 2) croisant la variable Y des catégories avec les 3 variables de X (Petal_L, Petal_W et Sepal_L) liées aux catégories de Y par des flèches rouges dans le graphe implicatif (Figure 5, à droite).

Nombre de Id			Species			
Petal_L	Petal_W	Sepal_L	Iris-setosa	Iris-versicolor	Iris-virginica	Total
1	1	1	40			40
		2	10			10
2	2	1		5		5
		2		29		29
		3		13		13
	3	1			1	1
		2		1	4	5
		3			1	1
3	2	2		1	2	3
		3			2	2
	3	2			6	6
		3		1	34	35
Total			50	50	50	150

Tableau 2 – Répartition des 150 iris selon leur espèce et 3 de leurs dimensions, en rouge les iris « mal classés ».

La lecture de ce tableau nous permet d'écrire le jeu de trois règles de discrimination suivant :

Si Petal_L1 alors iris_setosa (N=50, nbVrai=50, nbFaux=0)

Si Petal_L2 et Petal_W2 alors iris_versicolor (N=47, nbVrai=47, nbFaux=0)

Si (Petal_L2 et Petal_W3) ou (Petal_L3) alors iris_virginica (N=53, nbVrai=50, nbFaux=3)

Et en récupérant les intervalles de valeurs des variables figurant dans le journal de CHIC (voir tableau 3), on peut remplacer les modalités des variables quantitatives dans les règles par leur appartenance à un intervalle (par exemple Petal_L1 devient Petal_L $\in$ [1 ; 1.9]).

niveau	Sepal_L	Sepal_W	Petal_L	Petal_W
1	de 4.3 à 5.3	de 2 à 2.8	de 1 à 1.9	de 0.1 à 0.6
2	de 5.4 à 6.2	de 2.9 à 3.3	de 3 à 4.9	de 1 à 1.6
3	de 6.3 à 7.9	de 3.4 à 4.4	de 5 à 6.9	de 1.7 à 2.5

Tableau 3 – Intervalle de valeurs de chaque niveau des variables quantitatives donné par CHIC

Nous constatons que ce jeu de trois règles permet de discriminer les 150 iris avec seulement 3 erreurs. Ce qui nous permet d'affirmer que CHIC nous a fourni pour les iris un modèle tout à fait pertinent de la relation X→Y, où X est la matrice de 4 variables quantitatives et Y une variable catégorielle.

2.3 Pourquoi et comment choisit-on l'ASI

Comme on peut le constater dans la section 2.2, complétée par les annexes relatives aux sections 2.2.1 et 2.2.2, les diverses méthodes visant à modéliser X→Y arrivent presque toutes à discriminer les 3 catégories d'iris avec plus de 95% de réussite, ce qui n'est pas surprenant sachant que dans le sous-espace de X formé de la longueur et de la largeur des pétales (voir figure 1), une espèce se détache nettement des deux autres, ces dernières formant 2 blocs distincts juxtaposés avec une zone commune de moins de dix individus plus difficiles à catégoriser. C'est donc sur d'autres critères que les résultats quantitatifs que nous choisissons la méthode destinée à nous fournir un modèle.

2.3.1 Comparaison entre ASI et les autres méthodes de modélisation

Nous avons donné une liste de méthodes visant à produire un modèle simple et fiable de la relation X→Y. D'abord les méthodes statistiques linéaires pour lesquelles la fiabilité est assurée par la théorie statistique (tests d'hypothèses), et la simplicité par la linéarité, ce qui signifie que la valeur prédite de Y est combinaison linéaire des valeurs de X. C'est la formule la plus simple quand Y est quantitative, comme dans la figure 2 où Y est la longueur du pétale. Mais quand Y n'est pas quantitative, ce qui est le cas de l'espèce, il faut faire des transformations préalables des données pour conserver la linéarité, et le modèle se complexifie, même si les résultats peuvent s'exprimer par des équations et des graphiques comme en figure 3. Quand X contient peu de variables, on peut s'y retrouver, mais avec 29 variables comme dans les données d'intonation que nous voulons traiter, le choix et l'utilisation d'un tel modèle s'avère délicat. De plus les

conditions d'application exigées par ces méthodes statistiques linéaires ou dérivées sont vérifiées par les variables X des iris de Fisher, mais pas par celles des données d'intonation, qu'il faut transformer pour les adapter. Une fois ces étapes franchies, on dispose d'un modèle explicite utilisable pour la discrimination, et de la mesure de sa fiabilité, qui peut s'avérer plus ou moins forte. Quand elle est suffisamment forte, elle permet d'établir la théorie visée, d'où son intérêt.

Les méthodes de discrimination suivantes que nous avons vues (MSVM, arbres de décision) fournissent des résultats fiables sans exiger que les données vérifient autant de conditions que les méthodes statistiques précédentes, mais soit le modèle détaillé n'est pas fourni (fonctionnement des MSVM en « boîte noire »), soit il est simpliste (arbres de décision), la qualité se situant au niveau du résultat (taux élevé de prédictions réussies), pas du modèle[5]. Notons toutefois que ce n'est pas parce que ces méthodes n'exigent pas de conditions sur les distributions des données qu'elles pourront traiter correctement des données mal distribuées. Quant aux règles d'association, elles fournissent un modèle riche, mais complexe et qui reste attaché aux données traitées, par manque de méthode de validation statistique universellement reconnue. Leur avantage est de faire découvrir des relations locales inattendues, qui peuvent entrer dans la composition d'un modèle global qu'il conviendra de valider ensuite.

L'ASI est à la frontière entre les deux types de méthodes : elle fournit un modèle des données dont on peut évaluer la fiabilité grâce aux seuils indiqués dans l'interface graphique, mais sans avoir d'exigences sur la distribution des données. Son modèle sous forme de graphe implicatif permet d'exprimer des relations complexes entre les variables au sein d'un modèle global, mais comme elles ont toutes le même statut (pas de variable à expliquer ou explicative), la prédiction ne peut se faire directement dans CHIC. L'interface graphique permet de modifier les éléments du modèle (ajouter/retirer des variables, changer les seuils, déplacer des groupes de flèches, …) par simples clics.

Notre but étant d'extraire de nos données un modèle fiable, simple mais riche, nous avons choisi CHIC et ses graphes implicatifs, sachant qu'avec la facilité de manipulation procurée par l'interface graphique de CHIC, nous pouvons ajuster par clics jusqu'à obtenir le modèle qui nous convient le mieux sur les données d'intonation, sans être obligées de diagnostiquer leurs « défauts » pour les corriger au préalable (liaisons fortes entre certaines variables de X, distributions déséquilibrées, nombreux ex-aequo, valeurs extrêmes, etc.).

2.3.2 Découpage ou non des variables quantitatives

Transformer une mesure quantitative en 3 catégories fait perdre de la précision sur les valeurs (on ne dispose plus que de 3 valeurs non ordonnées), et on pourrait penser qu'il vaut mieux l'éviter pour avoir un modèle plus fiable. Nous avons vu que ce n'est pas le cas avec CHIC pour les iris, le modèle avec les variables découpées en 3 catégories ayant été préféré à celui avec les variables simplement recodées sur l'intervalle [0 ; 1]. Il semblerait donc que le recodage de variables quantitatives en variables catégorielles ne diminue par la précision du graphe implicatif. Et dans le cas des iris, il lui en a fait gagner. Cela s'explique par le fait que 3 dimensions sur 4

[5] A nos yeux, la qualité d'un modèle est fonction de son intelligibilité, de sa concision, opérationnalité, …

(Petal_L, Petal_W, Sepal_L) de la fleur d'iris sont très liées entre elles et ont tendance à varier dans le même sens, et si on ne les découpe pas, elles se retrouvent dans le même graphe et liées aux mêmes espèces d'iris (voir figure 5, à gauche), la dernière variable (Sepal_W) se trouvant dans un autre graphe avec les espèces restantes d'iris, ce qui ne fait que deux graphes pour 3 espèces, donc une mauvaise discrimination. Ce n'est qu'en découpant les variables qu'on a pu obtenir une discrimination fine de chaque catégorie, avec des graphes séparés (voir Figure 5, à droite).

Ayant établi que découper les variables quantitatives peut augmenter la précision du modèle global, il s'agit maintenant de définir la façon de procéder. La découpe d'une variable quantitative se fait automatiquement dans CHIC, une fois choisi le nombre de parties. C'est un algorithme qui détermine les seuils de découpe en optimisant le rapport entre variance inter et variance intra, indépendamment des autres variables, contrairement à l'algorithme des arbres de décision, qui optimise en fonction de la variable à discriminer. Ce choix permet à CHIC d'obtenir des niveaux de significativité non biaisés en conservant sa méthode d'estimation asymptotique, alors que les arbres de décision utilisent des méthodes de validation croisée pour éviter ce biais. L'algorithme de CHIC fonctionnant avec des variances, il faut que la distribution des valeurs soit suffisamment équilibrée, ce qui est le cas des iris. Dans le cas contraire, comme celui de certaines variables acoustiques, il est préférable de découper soi-même les variables selon d'autres critères, ne faisant pas intervenir la variable Y, pour éviter d'introduire de biais dans la discrimination, en utilisant par exemple des quantiles pour avoir des catégories d'effectifs proches.

Le découpage peut se faire selon une théorie (par exemple pour la tension artérielle, découpage selon les seuils de l'hypotension et de l'hypertension). Comme nous ne disposions pas de théorie sur les dimensions des fleurs d'iris, nous avons profité de l'algorithme de découpage automatique proposé par CHIC pour faire essayer divers découpages, en commençant par un découpage identique de toutes les variables, en deux, trois, puis quatre parties. Le découpage en trois parties, procuré par CHIC, a donné de meilleurs modèles : en demandant des seuils entre 0.80 et 0.99, nous avons obtenu 3 graphes séparés, un par espèce, et dans chaque graphe, plus d'une flèche rouge (indices>=0.99), une seule flèche bleue (0.95<=indice<0.99) et aucune flèche d'une autre couleur, donc des indices d'implication statistique tous supérieurs à 0.95. En effet, on peut voir en annexe qu'en fixant le nombre de parties à 2 ou 4 pour le découpage de toutes les variables, les espèces iris-virginica et iris_versicolor apparaissent dans le même graphe. Ensuite nous avons tenté deux découpages en un nombre différent de parties selon les variables des pétales et celles des sépales, qui sont joints en annexe de la partie 2.2.3. On peut y voir qu'un de ces deux graphes implicatifs l'emporte sur tous les autres découpages, y compris ceux avec toutes les variables découpées en 3 parties, c'est celui obtenu en découpant les pétales en trois et les sépales en deux, car il est formé de 3 graphes séparés (un par espèce), il ne contient que des flèches rouges (implication statistique>=0.99) et il est plus concis (moins de variables après découpage). Toutefois, en créant à partir de ce graphe le tableau et le jeu de règles de discrimination comme nous l'avons fait dans la section précédente pour le découpage de chaque variable en 3 parties, nous constatons que sa qualité de discrimination n'est pas meilleure (également 3 erreurs sur les 150 iris), nous privilégions donc le modèle issu du découpage de toutes les variables en trois parties, plus facile à justifier théoriquement ici (nombre de parties des variables de X égal au nombre de catégories de Y).

Ayant perdu de la qualité en passant le nombre de parties du découpage de trois à quatre, nous n'avons pas essayé au-delà de quatre parties.

3 Application aux données d'intonation

Malgré le grand nombre d'études consacrées à l'intonation du français, relativement peu d'entre elles ont porté sur une analyse fine des indices acoustiques et sur un grand nombre de locuteurs. En outre, la reconnaissance automatique des différentes intonations est actuellement faite par des systèmes de reconnaissance fondés sur des modèles cachés, qui ne dévoilent pas les paramètres utilisés pour leurs résultats, et qui ont pour l'instant des résultats assez faibles. L'interprétation objective des indices prosodiques, et tout particulièrement des courbes mélodiques est actuellement de plus en plus utilisée pour l'apprentissage des langues et, dans le domaine de la rééducation, l'apprentissage de la langue (De Bot, 1983), (Bonneau et Colotte, 2011).

Dans une première partie nous décrivons la problématique, dans la suivante les données et leur recodage, puis dans la troisième leur traitement par CHIC, et enfin l'interprétation des résultats.

3.1 La problématique

Lors d'une conversation en français entre deux personnes, il n'y a pas que les mots qui donnent du sens, il y a aussi l'intonation des phrases prononcées : par exemple si le ton du locuteur monte en fin de phrase, l'auditeur en déduit généralement qu'on lui pose une question. En effet, si la question ne comporte pas de marques grammaticales (inversion sujet-verbe « Faut-il », « As-tu » ou ajout en début de phrase de « Est-ce que », « Pourquoi », etc.) c'est par son ***contour mélodique*** qu'elle se distinguera par exemple d'une déclaration prononcée avec une intonation neutre : montant en fin de phrase pour la première et descendant pour la seconde. Delattre (1966) propose dix intonations de base du français, en distinguant quatre niveaux de hauteur de voix. Les schémas de l'article montrent clairement que pour l'auteur, les contours des intonations dites « de base » s'opposent non seulement par leur direction (montante vs descendante) et leurs niveaux de départ et d'arrivée mais également par leur forme : ainsi le contour des questions est caractérisé par un accroissement de hauteur plus fort en fin de contour (la ***vélocité*** est donc élevée en ce point), alors que pour les groupes intonatifs non terminaux (non en fin de phrase, voir Figure 8), l'accroissement est plus fort en début de contour (la vélocité est faible en fin de contour).

Plus précisément, les trois types de contour que nous étudierons dans cet article sont représentés ainsi à l'aide des 4 niveaux de Delattre :

les ***questions*** sans marques grammaticales, que nous noterons ***qis***, avec un contour montant qui part d'un niveau 2 pour aller à un niveau 4 (par exemple la question « Les agneaux ? » en réponse à l'affirmation « Je les ai vus »),
les ***phrases déclaratives***, que nous noterons ***dis***, avec un contour (final) descendant, partant du niveau 2 pour aller au niveau 1 (par exemple la réponse « Les agneaux. », à la question « Qu'as-tu vu ? »),

les ***syntagmes non terminaux***, que nous noterons ***cis***, avec un contour montant partant du niveau 1 pour aller au niveau 2 (par exemple « les agneaux » qui est au début de la phrase « Les agneaux ont vu leur mère », voir Figures 6 et 7).

Dans cette étude, notre objectif est donc de caractériser ces 3 types de contours prosodiques à l'aide d'indices acoustiques, notamment les niveaux de hauteur, la direction et l'allure générale de la pente mélodique en fin de phrase (ou de syntagme). En particulier, nous essaierons de vérifier si la hauteur atteinte en fin de question est effectivement plus élevée que celle des syntagmes non terminaux et des déclarations, et si les questions sont caractérisées par une accélération plus tardive que celle qu'on observe normalement en fin de syntagme non terminal. Enfin, nous essaierons de distinguer les syntagmes non terminaux des fins de phrases déclaratives, y compris quand celles-ci ont été prononcées avec une intonation particulière, comme le doute, qui a pour effet de modifier l'allure de la courbe qui tend à devenir légèrement montante (et non descendante, comme pour une intonation neutre).

3.2 Les données

Ce corpus est issu d'un corpus plus vaste, qui comprend notamment des phrases plus longues, et qui a été enregistré dans le cadre du projet Intonale (2009-2011) supporté par le Comite de Coordination et d'Orientation Scientifique Lorrain: Nancy University, Paul Verlaine University, CNRS,INRIA, INRA, INSERM, CHU.

Nous ne décrivons ici que les éléments nécessaires à notre étude. Celle-ci s'appuie sur 115 enregistrements sonores concernant 5 phrases simples prononcées de 3 façons différentes par 33 locuteurs différents (chaque locuteur n'a prononcé qu'une partie des phrases, comme c'est détaillé plus bas). Nous décrivons d'abord les conditions de l'expérience construite pour enregistrer la parole des locuteurs, puis comment nous avons créé les indices acoustiques et les variables à partir des enregistrements sonores.

3.2.1 Les conditions de l'expérience

Les éléments relatifs aux 115 enregistrements que nous avons analysés sont décrits en utilisant la terminologie des plans d'expérience (Hoc 1983).

Facteur P : « Phrase » :

Voici les cinq phrases (ou parties de phrases) différentes qui ont été étudiées :

P1 : Nos amis
P2 : Les élèves
P3 : Les lamas
P4 : La marelle
P5 : Les agneaux

Facteur T : « Type de phrase » :

Le contexte dans lequel les phrases ont été lues a été conçu pour que les locuteurs prononcent les phrases de manière neutre, sans émotion particulière, avec l'intonation qui respecte chacune des trois catégories qui nous intéressent ici. Certains locuteurs ont cependant parfois donné à certaines phrases une émotion particulière, notamment en fin de phrases déclaratives (expression du doute, par exemple). Nous avons conservé ces

phrases, et notre analyse a donc porté, sur les questions (***qis***), les syntagmes non terminaux (***cis***), et les déclarations (***dis***), ces dernières comportant un certain nombre de phrases non neutres.

C'est ce facteur « type de phrase »qui deviendra la variable Y de notre modèle X→Y

Facteur L : « Locuteur » :

Ce sont 6 hommes et 27 femmes, ayant le français comme langue maternelle. Le nombre de phrases que chaque locuteur a prononcées varie entre 1 et 5.

Facteur G : « Genre »

Ce facteur à 2 modalités H (Homme) et F (Femme).

Facteur S : « Rang du son » :

Pour rendre compte de la totalité du contour mélodique de chaque phrase, nous avons numéroté chaque partie de la phrase. Le numéro 1 correspond à ce qui est enregistré avant le début, le numéro 2 à la première consonne, le numéro 3 à la première voyelle et ainsi de suite. Pour la phrase interrogative « les agneaux ? » les numéros vont de 1 à 7, par contre pour les syntagmes non terminaux (de type « cis »), ils vont au-delà de 7 car pour forcer le locuteur à adopter l'intonation convenable pour « les agneaux », on lui a demandé de prononcer la phrase plus longue « les agneaux ont vu leur mère », et on a annoté le début de phrase incluant le début de la partie suivante pour être sûres d'avoir toute l'information pertinente (voir Figures 6 et 7).

Le plan d'expérience

Nous aurions aimé que ce soit un plan factoriel complet équilibré, qu'il y ait autant d'hommes que de femmes, que chacun prononce les 5 phrases dans les 3 intonations. Ce n'a pas été le cas, aucun locuteur n'a prononcé une même phrase avec des intonations différentes, et ceux qui ont prononcé 4 ou 5 phrases ont utilisé les 2 intonations « qis » (pour les phrases 1, 3 et 5, ou bien 2 et 4) et « dis » (pour les phrases restantes, la 3 étant parfois inutilisée) alors que les autres, qui ont prononcé seulement 2 ou 3 phrases, soit les phrases 1, 3 et 5, soit les phrases 2 et 4, ont utilisé uniquement l'intonation « cis ». Ce plan incomplet et déséquilibré s'explique par le poids de réalisation d'une telle expérience, autant pour ceux qui la subissent en tant que locuteur, que pour ceux qui la font passer et ceux qui transforment les enregistrements sonores en fichiers de données exploitables par des logiciels statistiques. En effet, dans le cadre du projet Intonale, chaque locuteur à prononcé une dizaine de phrases, dont nous n'avons extrait qu'une petite partie, les plus simples, pour initier notre analyse. Puis c'est de façon semi-automatique que nous avons créé le fichier destiné aux analyses statistiques, avec 115 lignes et une cinquantaine de colonnes : il nous a fallu annoter les 115 enregistrements sonores, en extraire les parties correspondant aux phrases visées, et les aligner (leur longueur dépendant de la rapidité d'élocution du locuteur), avant de créer les variables. Ce plan d'expérience n'est pas non plus un plan randomisé dans les règles de l'art, ce qui interdit d'utiliser des techniques paramétriques poussées visant à éliminer les effets des variables de contrôle telles que « locuteur », « phrase », « voyelle », etc. afin d'obtenir des valeurs épurées pour le seul facteur qui nous intéresse ici, le contour mélodique, que nous avons appelé type de phrase, ayant 3 modalités, ***cis***, ***dis*** et ***qis***.

Malgré cela, c'est une chance d'avoir pu obtenir de telles données, qui se sont avérées adaptées à notre problématique une fois choisis les indices acoustiques et la technique statistique appropriés.

Pour notre étude, tous les facteurs de ce plan d'expérience ont été pris en compte dans le modèle X→Y que nous cherchons à établir : ***Locuteur***, ***Rang du son*** dans la création des variables formant X, ***Phrase*** et ***Genre*** ont formé la matrice Z des variables à contrôler et ***Type de phrase*** est devenu la variable Y du modèle X→Y que nous cherchons à établir.

3.2.2 La construction des variables de X

Les variables de X dérivent des indices acoustiques, eux-mêmes construits à partir des enregistrements sonores.

Les indices acoustiques

Ils ont été extraits du fichier son en utilisant diverses techniques ainsi que le logiciel PRAAT (Boersma et Weenink, Version 5.4.10 du 27 June 2015, www.praat.org). Nous donnons ci-dessous quelques éléments de leur construction sans entrer dans les détails, le but étant seulement d'établir leur signification et de juger de leur précision pour pouvoir interpréter ensuite les résultats obtenus par CHIC.

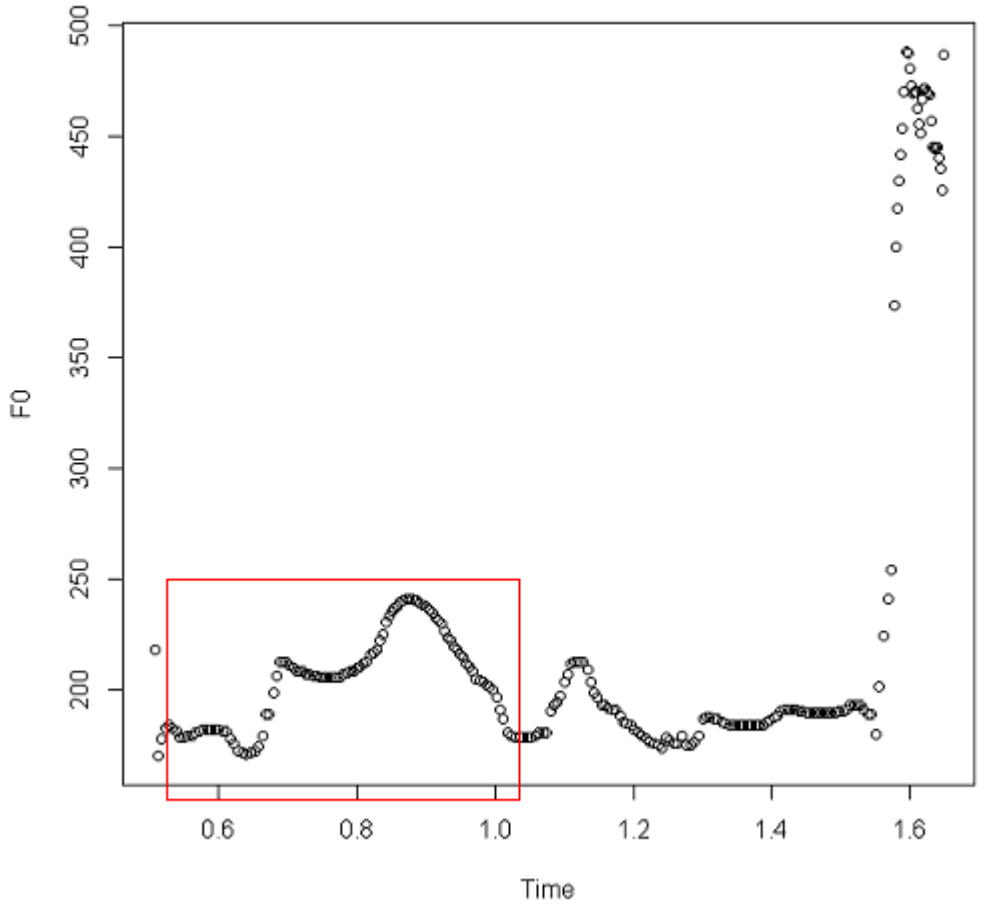


Figure 6 – Valeurs de F0 en fonction du temps pour la phrase « Les agneaux ont vu leur mère », donc de type « cis » prononcée par un locuteur. La partie de la phrase qui nous intéresse « Les agneaux », incluant la voyelle suivante « on » est encadrée de rouge.

Le premier de ces indices représente la fréquence fondamentale, F0, qui correspond à la hauteur de la voix.. Dans la figure 6, on a représenté les valeurs brutes de F0 en fonction du temps lors de la prononciation d'une phrase commençant par « les agneaux ». Puis on a repéré la partie du signal correspondant à chaque son, comme indiqué dans la figure 7. Bien sûr, les frontières verticales dessinées pour séparer les sons ne peuvent pas toujours être placées de manière précise, comme c'est le cas ici pour le passage entre les deux voyelles /o/ et /on/. Le travail de segmentation, qui demande beaucoup de temps et de rigueur, a été réalisé par l'une des auteures.

Trouver des indices qui mesurent l'évolution de F0 quand on passe d'un son à un autre (hauteur, croissance, décroissance, stabilité, vitesse, etc.) n'est pas simple, comme on peut le remarquer dans la figure 7. Les pentes des droites rouges, obtenues par régressions sur des points consécutifs, auraient pu fournir un coefficient d'accroissement de F0. D'autres indices exprimant l'accroissement ont été testés, mais finalement c'est la vélocité instantanée calculée par PRAAT, donc une valeur en chaque point, qui a été retenue à sa place et qui est notre deuxième indice.

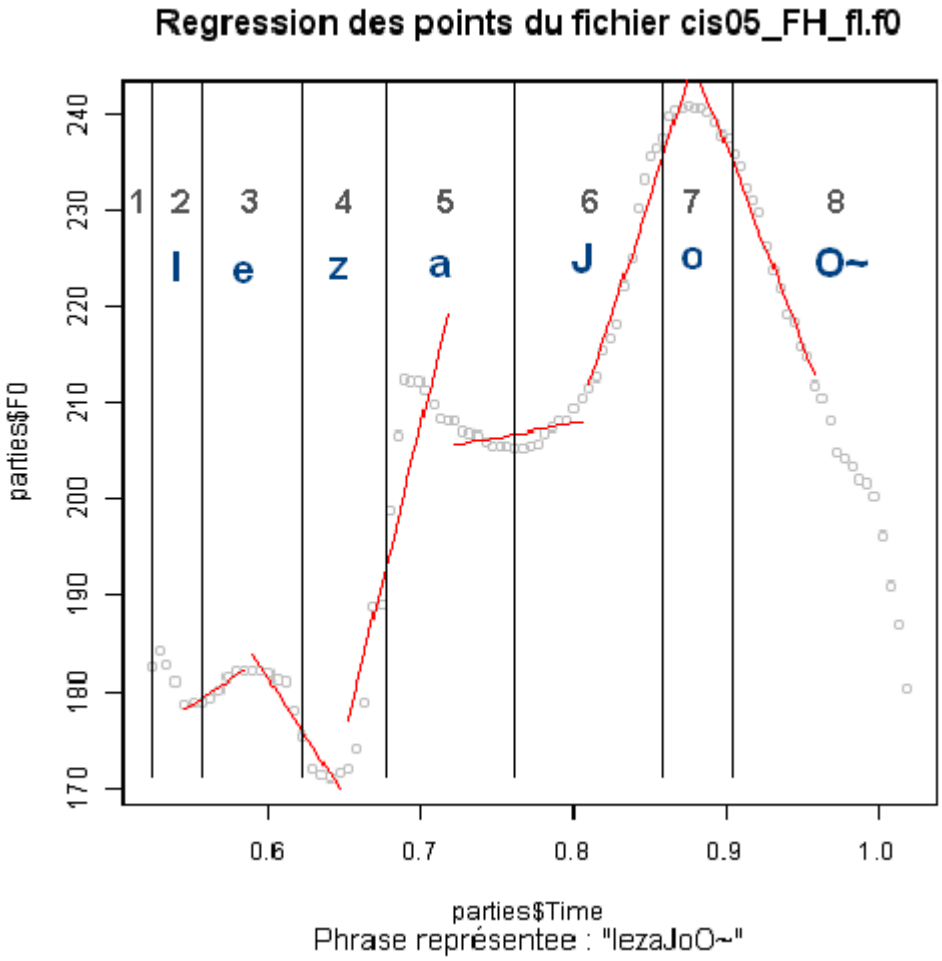


Figure 7 – La phrase a été découpée en sons par les traits verticaux, et les droites rouges ont été obtenues par des régressions sur les points allant du milieu d'une bande au milieu de la suivante.

Les 29 variables de X

Selon les enregistrements le nombre de valeurs des indices diffère pour chaque son (il y en a plus quand le locuteur parle lentement), ce qui rend difficile leur utilisation dans des modèles. Nous en avons calculé un seul par son et par locuteur en appliquant diverses fonctions sur l'ensemble des valeurs successives pour les fréquences F0 du son, max, min, moyenne (mean), final, amplitude (ex_size), durée, force/intensité (mean_int), et max ainsi que final pour la velocité. Les valeurs de ces derniers indices pour chacune des 3 positions de voyelles (numérotées 3, 5 et 7) ainsi que les différences de maxF0 entre 2 voyelles voisines (3 et 5, 5 et 7) ont donné les 29 variables suivantes :

3maxF0, 5maxF0, 7maxF0, 3minF0, 5minF0, 7minF0, 3meanF0, 5meanF0, 7meanF0, 3finalF0, 5finalF0, 3ex_size, 5ex_size, 7ex_size, 3duration, 5duration, 7duration, 3mean_int, 5mean_int, 7mean_int, 3max_vel, 5max_vel, 7max_vel, 3final_vel, 5final_vel, 7final_vel, 7finalF0, maxF0_v5_3, maxF0_v7_5.

Pour faciliter l'interprétation des fréquences F0, nous les avons transformées avant traitement statistique en « semi-tons » par la formule Semitone(F0) = 12 $\log_2$(F0/base) où nous avons choisi base = 100. Puis pour contrebalancer l'effet « locuteur », notamment les grandes différences de hauteur de voix (ici exprimée en semi-tons) femme/homme, nous avons normalisé en retirant à chaque mesure la moyenne du locuteur concerné.

3.3 Le traitement par CHIC

Nous recherchions un modèle explicite des données permettant de différencier les 3 catégories de phrases mélodiques.

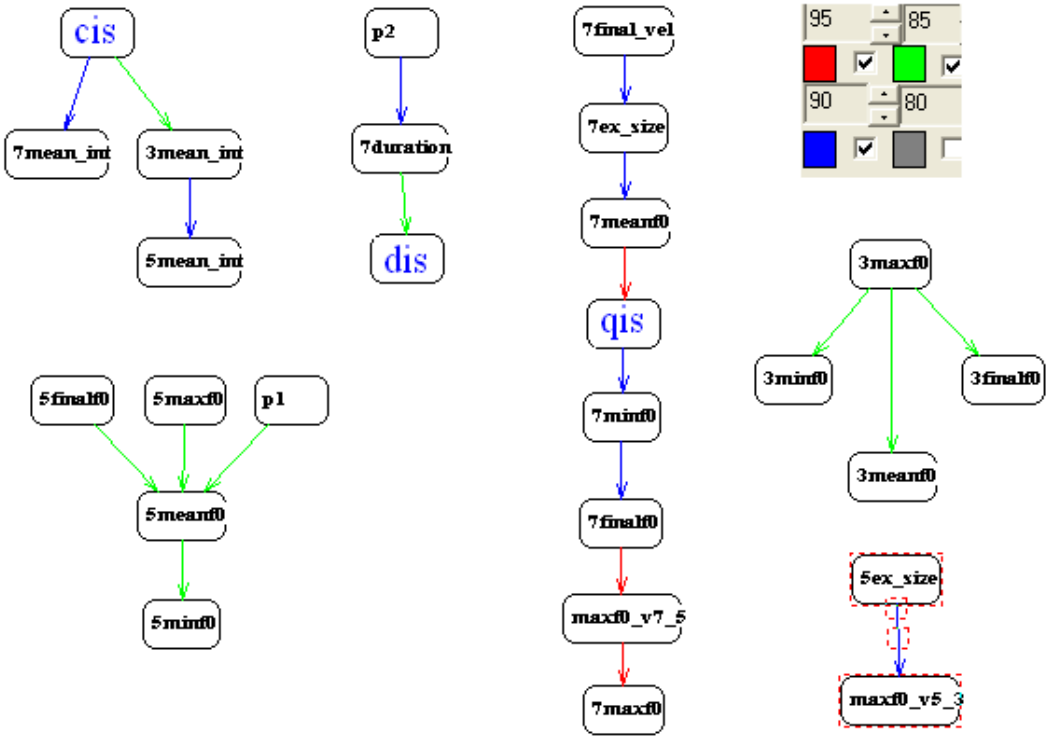


Figure 8 – Graphe implicatif des variables des données mélodiques : 29 variables acoustiques numériques recodées sur [0 ; 1], 3 catégories, 5 phrases, les couleurs correspondant aux seuils cochés en haut à droite.

Le grand nombre de variables « explicatives » que nous avons créées, et les liaisons fortes entre elles dues à leur mode de construction nous interdisaient sans correction préalable complexe des données l'utilisation de la plupart des méthodes de discrimination des statistiques classiques (décrites précédemment). Quant à l'arbre de décision que nous avons obtenu, il n'utilisait que 2 ou 3 variables parmi les 29 créées pour discriminer les 3 catégories, ce qui donnait un modèle trop pauvre en regard de la théorie de Delattre. Grâce à CHIC, nous avons pu utiliser l'ASI sur la totalité des variables afin d'extraire des modèles de contour mélodique à partir d'indices acoustiques. Nous décrivons ici ceux qui ont été obtenus à partir des graphes implicatifs en répétant la démarche suivie pour les iris de Fisher, d'abord, en Figure 8, avec la variable Y des 3 catégories (***qis, cis, dis***), les 29 variables quantitatives X recodées sur l'intervalle [0 ; 1], ainsi que 2 variables de contrôle Z (phrases P1 à P5, genre H/F), puis en Figure 9, avec chaque variable de X découpée en 3 catégories.

L'ASI sur les 29 variables recodées sur [0 ; 1]

Notre but premier étant de discriminer les 3 catégories de Y, pour le graphe implicatif de la figure 8, nous avons du « descendre » jusqu'au seuil de 0.85 pour faire apparaitre la catégorie 'dis'.

On peut voir que la catégorie ***qis*** (questions) est nettement discriminée par le son 7, qui correspond à la dernière voyelle de la phrase (par exemple le « i » de « nos amis »), qui doit avoir des valeurs élevées de F0, que ce soit le min, le max, la moyenne, la valeur finale, la différence entre le max et le min (7ex_size), la vélocité finale, ainsi que la différence entre le max sur cette voyelle et le max sur la voyelle précédente (maxF0_v7_5). Nous retrouvons donc bien pour les questions la forte vélocité en fin de contour signalée par Delattre, ainsi que la forte montée de tonalité pour la dernière voyelle. Par contre nous ne retrouvons pas pour ***dis*** et ***cis*** les hypothèses de Delattre, mais d'autres informations qui restent à expliquer : ***dis*** est discriminée par la durée importante de la dernière voyelle, et ***cis*** par l'importance de l'intensité avec laquelle chaque voyelle est prononcée. De plus, nous voyons apparaitre la spécificité de 2 phrases, indépendante de l'intonation, ***p1*** dont la deuxième voyelle (numéro 5) a des valeurs fortes pour F0, et ***p2*** dont la troisième et dernière voyelle (numéro 7) est plus longue. Pour p2, il s'agit de la variabilité inter-vocalique, le /e/ est déjà long et en plus il est devant une consonne allongeante. Pour trouver des relations plus intéressantes entre les catégories ***dis*** et ***cis*** et les 29 variables, il faut créer des variables correspondant aux valeurs faibles, donc les découper en intervalles.

L'ASI sur les 29 variables découpées en 3 catégories chacune

Nous avons scindé la distribution de chacune de ces variables en trois intervalles en regroupant dans le premier intervalle les valeurs les plus basses de la distribution, dans le deuxième les valeurs moyennes, et dans le troisième les valeurs les plus élevées.

Comme le montre la Figure 9, le graphe implicatif a spontanément séparé et regroupé entre eux les trois types de phrases analysées dans cette étude, et associé à chacun de ces types des groupes de variables (graphes numérotés de 1 à 3). L'importance de la voyelle finale, ici la troisième voyelle, bien connue dans le domaine

prosodique, apparaît de manière très claire puisque ce sont les valeurs fréquentielles de cette voyelle qui sont associées par le logiciel à chacun des trois types de phrase.

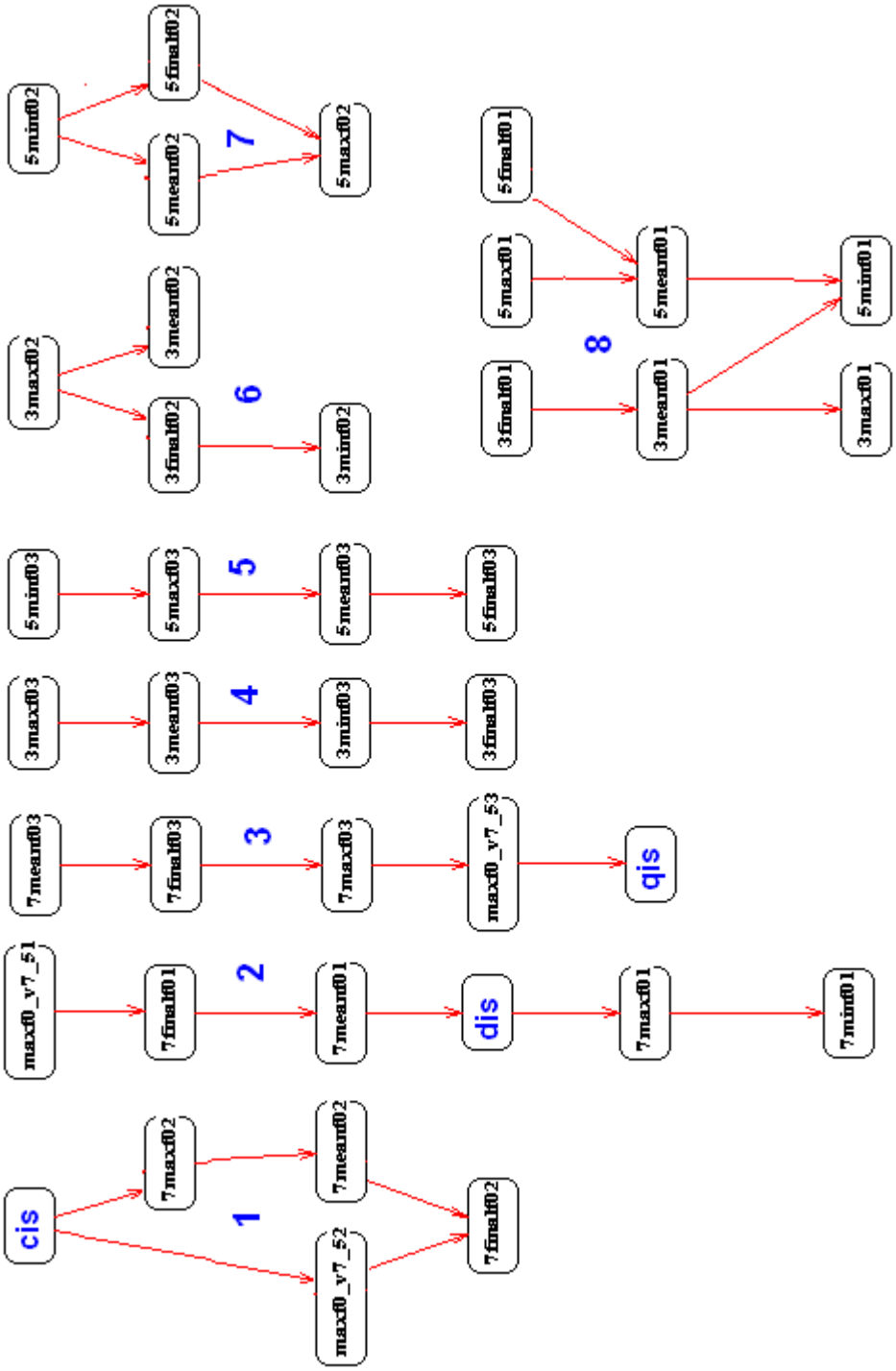


Figure 9 – Graphe implicatif comprenant toutes les valeurs de p>0.9999 (flèches rouges) pour les variables concernant F0. Chaque graphe séparé est numéroté en bleu.

Regardons maintenant quelles sont les variables précises et les régions fréquentielles (intervalles de distribution) qui sont associées par CHIC aux phrases. Les variables associées aux questions (qis) et aux déclarations (dis) sont la valeur moyenne,

maximale, finale, et minimale de la fréquence fondamentale pour la voyelle finale, ainsi que la différence entre la valeur maximale de la voyelle finale et celle de la pénultième (l'ordre des variables est légèrement différent selon les types de phrases). Pour les questions (qis), nous constatons que, quelle que soit la variable, ce sont les valeurs élevées de sa distribution qui sont significatives. Ce résultat est en adéquation avec les descriptions prosodiques des questions sans marque grammaticale (contour mélodique montant entre la pénultième et la voyelle finale, fréquences finales élevées). En ce qui concerne les déclarations (dis), ce sont les valeurs basses de la distribution qui sont significatives. Encore une fois ce résultat est en adéquation avec les descriptions prosodiques des phrases déclaratives (contour descendant, fréquences finales basses).

La caractérisation des groupes nominaux sujets (cis) est plus délicate. Les segments de ce type, situés à l'intérieur d'une phrase, possèdent en général un contour mélodique montant, mais de moindre envergure que celui des questions. Les valeurs fréquentielles caractéristiques de la dernière voyelle de ce groupe sont en général moins élevées que celles des questions et plus élevées que celles des déclarations. De manière intéressante, on voit également que ce sont les valeurs moyennes de la distribution des variables qui sont associées aux groupes nominaux sujets.

On s'aperçoit également que maxF0_V7_5, qui est la différence entre le max de la voyelle précédente et celui de la voyelle finale, est très important (troisième intervalle) pour les questions, moyen pour les groupes nominaux sujets, et faible pour les déclarations. Ce qui va encore une fois dans le sens de la description en niveaux de Delattre, qui indiquent des passages d'un niveau 2 à 4 pour les questions 1 à 2 pour les syntagmes non terminaux, et 2 à 1 pour les déclarations.

Nous continuons actuellement notre interprétation pour les graphes de 4 à 8, dans lesquelles interviennent d'autres variables que la catégorie des phrases. Ces analyses seront fournies ultérieurement sur demande.

4 Conclusion et perspectives

Comme nous l'avons souligné lors de l'analyse de nos résultats, le logiciel CHIC nous a permis 1) de confirmer les descriptions prosodiques théoriques, notamment celles de Delattre, à l'aide d'indices spécialement conçus pour vérifier ses schémas ; 2) de disposer d'éléments précis sur la manière dont ces indices codent les principales intonations du français

Le modèle que nous cherchions était un modèle de discrimination, du type X→Y, un modèle courant en statistique, qui pouvait être établi par des méthodes statistiques classiques mais également par des méthodes plus récentes de l'apprentissage automatique, et par l'ASI, à travers CHIC. Nous les avons testées et comparées sur les iris de Fisher, jeu de données ayant des caractéristiques proches du nôtre (variables de X quantitatives très liées, variable Y à 3 catégories, discrimination facile de Y à partir de X même pour un observateur inexpérimenté) et constaté que CHIC nous fournissait de façon plus ergonomique un modèle des données tout aussi fiable, mais plus simple, tout en étant plus riche, qui pouvait servir à la discrimination, moyennant toutefois une utilisation détournée de l'ASI (qui n'est pas une méthode de discrimination).

L'ASI est une méthode statistique de traitement de données catégorielles qui a été étendue aux variables quantitatives à condition de les recoder sur l'intervalle [0 ; 1], mais il est aussi possible de les transformer en variables catégorielles en les découpant en plusieurs intervalles. Nous avons comparé sur les iris de Fisher le modèle des données en recodant les variables sur [0 ; 1] puis en les découpant en plusieurs parties. Les différents modèles obtenus ne se recouvrent pas, chacun apporte des informations spécifiques, mais c'est un découpage de chaque variable de X selon 3 intervalles qui fournit le modèle de discrimination souhaité X→Y.

Le parallèle fait avec les iris nous a conduites à utiliser CHIC pour traiter nos données d'intonation. En effet, avec le plan d'expérience que nous avions, les locuteurs adaptaient leur intonation au type de phrase demandé, en produisant un contour mélodique mesuré par nos indices acoustiques. Et selon la théorie de Delattre, les indices acoustiques devaient permettre de discriminer les contours mélodiques. Nous étions donc dans un cas de discrimination « facile », comme pour les iris de Fisher, et on aurait pu choisir n'importe laquelle des diverses méthodes exposées dans la première partie pour une discrimination de bonne qualité. Aux raisons qui nous avaient fait choisir CHIC pour les iris : avoir facilement un modèle explicite et riche, s'en est ajoutée une autre : l'ensemble des variables acoustiques s'avérant plus complexes que les dimensions des fleurs d'iris, la plupart des autres méthodes s'appliquaient mal.

Nous avons étudié les 2 graphes implicatifs obtenus pour les données d'intonation, et c'est également celui avec découpage en 3 de chaque variable qui a permis de retrouver au mieux les hypothèses de Delattre. Mais l'autre graphe a apporté certaines informations supplémentaires, justes mais qui ne se sont pas avérées pertinentes car elles n'ajoutent rien à la théorie de Delattre sur les intonations de base du français.

Il nous reste à interpréter toutes les autres sorties produites par CHIC (jointes en annexe), refaire le traitement en découpant en 4 les variables issues de F0, pour voir si on peut retrouver les 4 niveaux de la théorie de Delattre, et estimer le pourcentage de schémas intonatifs qui peuvent être identifiés à l'aide des indices choisis.

Pour conclure, à la question du titre « Pourquoi et comment transformer des variables quantitatives en catégorielles ? », nous pouvons répondre ainsi :

« Pourquoi ? » : pour prédire une variable catégorielle, quand on dispose de variables explicatives quantitatives très liées entre elles, il est mieux de les transformer en variables catégorielles.

« Comment ? » : si on ne dispose pas de théorie donnant des seuils de coupes, et donc le nombre de parties pour chaque variable, il est mieux d'essayer d'abord un découpage de chaque variable quantitative en k intervalles pour prédire k catégories. Le choix de ces intervalles peut se faire avec des quantiles pour obtenir des effectifs voisins.

Mais c'est une heuristique que nous proposons ici, bien sûr, à éprouver sur d'autres données et d'autres problématiques.

Références

[1] Baillargeon, (2000), *La régression linéaire*, Edition des trois Sources, Quebec

[2] Besse, P. (2003), *Pratique de la modélisation Statistique*, Document interne du Laboratoire de Statistique et Probabilités, Université Paul Sabatier de Toulouse, http://www.math.univ-toulouse.fr/~besse/pub/modlin.pdf

[3] Bonneau, A. et Colotte, V (2011), Automatic Feedback for L2 Prosody Learning, in *Speech Technologies, Book 2*, I. IPSIC (editor), Intech, June 2011, 55-70, http://hal.inria.fr/inria-00579255/en

[4] Cadot, M ; (2006), Extraire et valider les relations complexes en sciences humaines : statistiques, motifs et règles d'associatio, Thèse de l'université de Franche-Comté.

[5] Dagnelie P., (2003), *Principes d'expérimentation : planification des expériences et analyse de leurs résultats*. Grenoble, Presses Agronomiques de Gembloux, Edition électronique : http://www.dagnelie.be

[6] Droesbeke J.-J., Fine J., éditeurs. (1996) *Inférence non paramétrique, les statistiques de rangs*. Journées d'Etude en Statistiques de l'Association pour la Statistique et ses Utilisations, Edition de l'Université de Bruxelles, Ellipses

[7] De Bot, K. (1983), Visual feedback on intonation I: Effectiveness and induced practice behaviour, *Language and Speech* , **Vol.6 , No.4**, 331–350

[8] Delattre, Pierre (1966), Les dix intonations de base du français, *The french Review*, **vol 40**, 1-14.

[9] Fisher,R.A. (1936), The use of multiple measurements in taxonomic problems *Annual Eugenics*, **7, Part II**, 179-188

[10] Gras R. et collaborateurs (1996), L'implication statistique, une nouvelle méthode exploratoire de données, La pensée sauvage, Grenoble

[11] Gras, R., Régnier, J.-C., Marinica, C., Guillet, F. (2013), L'analyse statistique implicative : méthode exploratoire et confirmatoire à la recherche de causalités, 2ème éd., Cépaduès, Toulouse

[12] Hoc J.-M. (1983), L'analyse planifiée des données en psychologie, PUF Paris

[13] Logiciel CHIC : Classification Hiérarchique Implicative et Cohésitive, Version 6.0, Copyright (c) 2012

[14] Mitchell, T. (1997). *Machine Learning*, McGraw Hill

[15] Morin H., (1999), *Théorie de l'échantillonnage*, Les presses de l'université, Laval

[16] Morineau, A., Nakache, J.-P., Krzyzanowski, C., (1996), *Le modèle log-linéaire et ses applications*, Cisia-Ceresta, Paris.

[17] Nakache, J.-P., Confais J.,(2003), Statistique explicative appliquée : analyse discriminante, modele logistique, segmentation. Editions Technip, Paris, France

[18] Prum, B., (1996), *Modèle linéaire. Comparaison de groupes et régression*, Eyrolles, Collection INSERM, Paris

[19] Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984), *Classification and Regression Trees*. Wadsworth.

[20] Ripley, B. D. (1996), Pattern Recognition and Neural Networks. Cambridge.

[21] Schwartz, D., (1991), Méthodes statistiques à l'usage des médecins et des biologistes, Flammarion, Paris

[22] Siegel S., Castellan N.J. Jr.,(1988), *Nonparametric statistics for the behavorial sciences*, McGraw-Hill, London.

[23] Weston J. and Watkins, C. (1998), *Multi-class support vector machines*. Technica l Report CSD-TR-98-04,Royal Holloway, University of London, Department of Computer Science

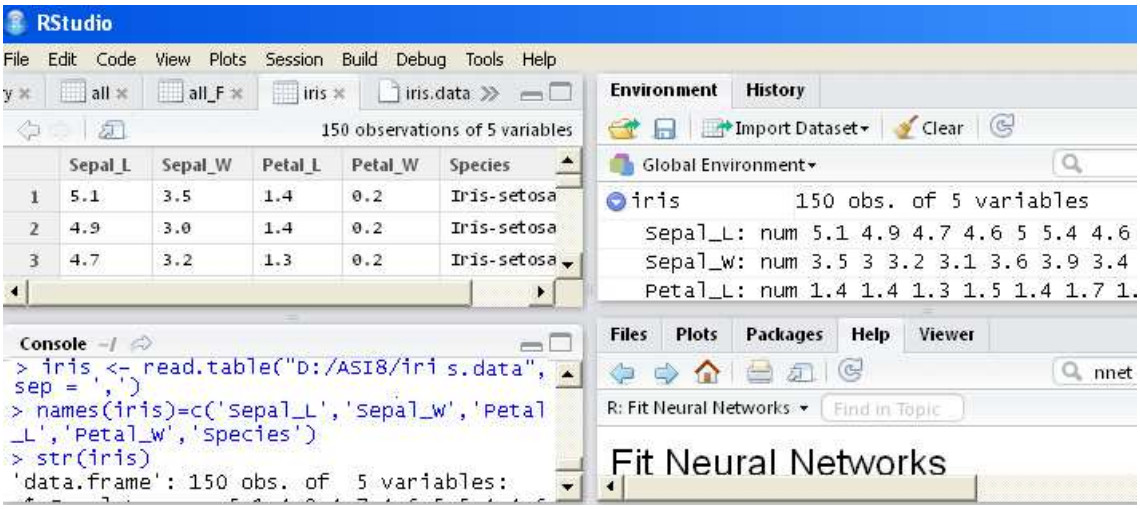
Annexe de la partie 2.2.1 : Scripts R et leur résultats

L'import des données dans l'environnement R

Les iris ont été récupérées dans UCI repository () sous forme d'un fichier iris.data, de 5 colonnes, une par variable, séparées par des virgules, et de 150 lignes, sans intitulés. On les importe dans R, puis on nomme les variables et on affiche leur structure de la façon suivante, avec en bleu les lignes de code écrites après le symbole d'invite « > » et en noir le résultat de la commande, quand il peut s'afficher.

```
> iris <- read.table("D:/ASI8/iris.data",sep = ',')
> names(iris)=c('Sepal_L','Sepal_W','Petal_L','Petal_W','Species')
> str(iris)
'data.frame':  150 obs. of  5 variables:
 $ Sepal_L: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal_W: num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal_L: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal_W: num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species: Factor w/ 3 levels "Iris-setosa",..: 1 1 1 1 1 1 1 1 1 1
...
```

On peut voir dans la copie d'écran ci-dessous de RStudio (environnement de programmation de R disponible sous Windows, http://www.rstudio.com/) en haut à gauche les données importées dans R sous forme tabulaire :



Indication pour les grands débutants en R : quand on veut utiliser une fonction d'un package, il faut le télécharger auparavant puis le charger en mémoire par la commande library(nom_package) tapée dans la console. Il n'y a plus qu'à copier-coller à la suite les commandes (en les validant une à une dans l'interface classique de R, ou en un seul bloc dans la console de RStudio), et une fois celles-ci validées (en enfonçant la touche « entrée »), observer les résultats. Attention à ne pas mettre des minuscules à la place de majuscules ou inversement, R est un langage sensible à la casse.

La régression linéaire généralisée

En R, comme dans beaucoup d'autres langages statistiques, c'est la fonction ***glm*** (Generalized Linear Model) qui permet d'obtenir les divers modèles, en faisant varier

les arguments. Mais bien sûr il y en a d'autres plus spécifiques adaptées à chaque cas, et il en apparaît régulièrement de nouvelles disponibles dans de nouveaux packages R (en un an, le nombre de packages R a plus que doublé).

Dans l'aide sur ***glm***, obtenue en tapant help(glm), on apprend que la fonction de base a été implémentée par Simon Davies, et réécrite et complétée régulièrement depuis.

```
glm(formula, family = gaussian, data, weights, subset,    na.action,
start = NULL, etastart, mustart, offset,    control = list(...), model
= TRUE, method = "glm.fit",    x = FALSE, y = TRUE, contrasts = NULL,
...)
```

Cette écriture compliquée se simplifie si on omet les arguments suivis d'un signe égale, ils sont alors remplacés par leur valeur par défaut, qui suit leur signe égale. Par exemple la droite de régression des iris-setosa, en bleu dans la figure 2, s'obtient par

```
> setosa=iris[iris$Species=='iris-Setosa',]
> model.setosa=glm(setosa$Petal_L~setosa$Sepal_L)
> summary(model.setosa)
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.81377    0.34390   2.366   0.0221 *
setosa$Sepal_L  0.12989    0.06853   1.895   0.0641 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On a bien retrouvé les coefficients de l'équation de régression des iris-setosa. L'argument family a été omis, donc sa valeur est gaussian qui est la loi normale. Il convient de contrôler que les résidus suivent bien la loi normale :

```
> shapiro.test(model.setosa$residuals)
        Shapiro-Wilk normality test
data:  model.setosa$residuals
W = 0.9671, p-value = 0.1762
```

On peut juger que c'est le cas puisque puisque p>0.05

Le modèle logistique

Dans ce cas, on doit créer une variable dichotomique (ici versicolor) qui prend la valeur 1 pour l'espèce qu'on souhaite discriminer, et 0 sinon. L'argument family change pour indiquer que la fonction est un logit.

```
> mlogistic_versicolor=glm(versicolor~Petal_L+Petal_W+Sepal_L+Sepal_W,
data=iris, family=binomial(link=logit))
> summary(mlogistic_versicolor)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1185  -0.7929  -0.3835   0.7980   2.1149
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   7.3229     2.4980   2.932 0.003373 **
Petal_L       1.2993     0.6823   1.904 0.056864 .
Petal_W      -2.7043     1.1627  -2.326 0.020021 *
Sepal_L      -0.2527     0.6495  -0.389 0.697154
Sepal_W      -2.7794     0.7859  -3.537 0.000405 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour les modèles dérivés du modèle linéaire avec une variable Y à expliquer de plus de 2 catégories, nous renvoyons aux nombreux ouvrages en ligne sur les fonctions statistiques R. Les données des iris y sont traitées de multiples façons.

Annexe de la partie 2.2.2 : scripts R et leur résultats

Pour toutes ces méthodes issues de l'apprentissage automatique, le script de l'application aux iris a été adapté des exemples figurant dans l'aide des packages. L'ensemble des 150 iris numérotés de 1 à 150, espèce après espèce, a été découpé en deux parties, la partie destinée à l'entraînement a ses numéros dans l'ensemble « samp » obtenu en prenant au hasard 25 numéros de chaque espèce, l'ensemble « test » ayant les numéros dans l'ensemble complémentaire, codé « -samp ». La qualité de prévision se mesure uniquement sur l'ensemble test. Nous avons néanmoins fourni les tables de confusion croisant classe prédite et classe observée sur les deux ensembles afin de pouvoir confronter ces méthodes aux autres. Bien sûr, si on faisait un autre tirage de « samp », on pourrait obtenir des résultats légèrement différents, et également si on modifiait les paramètres des fonctions utilisées, en les adaptant aux caractéristiques des données, ce que nous n'avons pas fait ici, cette annexe visant seulement à illustrer la partie théorique correspondante.

Les données d'iris sont conservées de la partie précédente, sous la même forme. Voici le code permettant de générer les 25 numéros d'iris par espèce tirés au hasard (on a passé d'une couleur différente les numéros correspondant à des espèces différentes) :

```
> samp <- c(sample(1:50,25), sample(51:100,25), sample(101:150,25))
> samp
 [1]  21  48  31  14  44  15   9  36  40  16   2  39   6  24  35  33
8  10
[19]  27  26  12  41  37  11  46  65  84  56  99  91  82  52  93  79
77  85
[37]  60  90  75  64  95  59 100  58  87  57  63  74  94  96 119 106
104 114
[55] 150 122 123 111 107 103 131 145 115 117 120 147 141 110 121 132
127 148
[73] 133 125 134
```

Les réseaux neuronaux

Package nnet de R mis à jour le 29 juin 2015, auteur Brian Ripley, http://cran.r-project.org/web/packages/nnet/nnet.pdf

```
model.nnet <- nnet(Species ~ ., data = iris, subset = samp, size = 2, rang = 0.1,
decay = 5e-4, maxit = 200)
# table de confusion pour les données de test
table(iris$Species[-samp], predict(model.nnet, iris[-samp,], type = "class"))
# pour info table de confusion pour les données d'entraînement
table(iris$Species[samp], predict(model.nnet, iris[samp,], type = "class"))
```

Données test : 4 erreurs sur 75				Données d'entraînement			
	Iris-setosa	Iris-versicolor	Iris-virginica		Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	25	0	0	Iris-setosa	25	0	0
Iris-versicolor	0	21	4	Iris-versicolor	0	24	1
Iris-virginica	0	0	25	Iris-virginica	0	1	24

Les « Support Vector Machine » multiples

Package e1071 de R mis à jour le 19 février 2015, auteurs David Meyer et al., http://cran.r-project.org/web/packages/e1071/e1071.pdf

```
model.svm <- svm(iris[samp,1:4], iris[samp,5])
# table de confusion pour les données de test, puis d'entrainement
table(iris$Species[-samp], predict(model.svm, iris[-samp,1:4]))
table(iris$Species[samp], predict(model.svm, iris[samp,1:4]))
```

Données test : 3 erreurs sur 75				Données d'entraînement			
	Iris-setosa	Iris-versicolor	Iris-virginica		Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	25	0	0	Iris-setosa	25	0	0
Iris-versicolor	0	22	3	Iris-versicolor	0	24	1
Iris-virginica	0	0	25	Iris-virginica	0	1	24

Les arbres de décision

Package rpart de R mis à jour le 29 juin 2015, auteurs Terry Therneau et al., http://cran.r-project.org/web/packages/rpart/rpart.pdf

```
model.arbre_dec=rpart(Species ~ Sepal_L + Sepal_W + Petal_L + Petal_W, data=iris, subset=samp)
# table de confusion pour les données de test, puis d'entrainement
table(iris$Species[-samp], predict(model.arbre_dec, iris[-samp,], type = "class"))
table(iris$Species[-samp], predict(model.arbre_dec, iris[-samp,], type = "class"))
```

Données test : 4 erreurs sur 75				Données d'entraînement			
	Iris-setosa	Iris-versicolor	Iris-virginica		Iris-setosa	Iris-versicolor	Iris-virginica
Iris-setosa	25	0	0	Iris-setosa	25	0	0
Iris-versicolor	0	23	2	Iris-versicolor	0	25	0
Iris-virginica	0	2	23	Iris-virginica	0	2	23

```
# Voici le modèle
model.arbre_dec

n= 75
node), split, n, loss, yval, (yprob)
```

```
* denotes terminal node

1) root 75 50 Iris-setosa (0.33333333 0.33333333 0.33333333)
  2) Petal_L< 2.35 25  0 Iris-setosa (1.00000000 0.00000000 0.00000000) *
  3) Petal_L>=2.35 50 25 Iris-versicolor (0.00000000 0.50000000 0.50000000)
    6) Petal_W< 1.65 27  2 Iris-versicolor (0.00000000 0.92592593 0.07407407) *
    7) Petal_W>=1.65 23  0 Iris-virginica (0.00000000 0.00000000 1.00000000) *
```

Annexe de la partie 2.2.3 : autres résultats sur les iris avec CHIC

D'abord des éléments supplémentaires pour le découpage des variables en 3 parties, puis un rapide tour des autres essais de découpage : chaque variable en 2 parties, les dimensions des pétales en 2 parties et celles des sépales en 3 parties, puis chaque variable en 4 parties.

Découpage de chacune des 4 dimensions de la fleur en 3 parties

En figure 5, de la section 2.2.3, à droite, nous avons représenté le graphe implicatif avec 3 graphes bien séparés, un par type d'iris, qui nous montrait l'association forte entre les 4 variables de dimension de la fleur et le type d'iris. Ci-dessous figure une sortie de CHIC montrant les fréquences des couples de variables :

Nous avons encadré de rouge les nombres de co-occurences de chacune des 3 valeurs de largeur de pétale avec le type d'iris. On constate que 33% des 150 fleurs sont de l'espèce « iris-setosa » et que la largeur de leur pétale est petite (Petal_W1), que 32% sont de l'espèce « iris-versicolor » et que la largeur de leur pétale est moyenne (Petal_W2), et que 31% sont de l'espèce « iris-virginica » et que la largeur de leur pétale est grande (Petal_W3), soit un taux de ***coïncidence***[6] de 96% obtenu entre l'espèce d'iris et la mesure de 1 à 3 de la largeur de ses pétales. De la même façon, la coïncidence est de 94% pour la longueur des pétales (en bleu), de 73% pour la longueur des sépales (en orange), et seulement de 55% pour la largeur des sépales (en vert).

[6] Nous utilisons ici le terme de ***coïncidence*** à la place du terme plus statistique de ***prédiction*** dans la mesure où nous raisonnons à partir de comptages faits sur la totalité des iris, sans utiliser d'échantillonnage ou de découpage en plusieurs sous-ensembles (entraînement et test, etc.).

	Sepal_L1	Sepal_L2	Sepal_L3	Sepal_W1	Sepal_W2	Sepal_W3	Petal_L1	Petal_L2	Petal_L3	Petal_W1	Petal_W2	Petal_W3	Iris-setosa	Iris-versicolor	Iris-virginica
Sepal_L1		0.00	0.00	0.05	0.13	0.13	0.27	0.04	0.00	0.27	0.03	0.01	0.27	0.03	0.01
Sepal_L2			0.00	0.17	0.10	0.08	0.07	0.23	0.06	0.07	0.21	0.07	0.07	0.21	0.08
Sepal_L3				0.09	0.22	0.03	0.00	0.09	0.25	0.00	0.10	0.24	0.00	0.09	0.25
Sepal_W1					0.00	0.00	0.01	0.20	0.11	0.01	0.20	0.11	0.01	0.18	0.13
Sepal_W2						0.00	0.13	0.15	0.17	0.13	0.14	0.18	0.13	0.15	0.17
Sepal_W3							0.20	0.01	0.03	0.20	0.01	0.03	0.20	0.01	0.03
Petal_L1								0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.00
Petal_L2									0.00	0.00	0.31	0.05	0.00	0.32	0.04
Petal_L3										0.00	0.03	0.27	0.00	0.01	0.29
Petal_W1											0.00	0.00	0.33	0.00	0.00
Petal_W2												0.00	0.00	0.32	0.03
Petal_W3													0.00	0.01	0.31
Iris-setosa														0.00	0.00
Iris-versicolor															0.00
Iris-virginica															

Ce tableau indique donc comment chaque dimension de la fleur prise séparément peut contribuer à prédire son espèce, alors que le graphe implicatif (figure 5 à droite) montre quelle combinaison de variables permet de prédire l'espèce, les détails se trouvant dans le tableau 2.

Un autre traitement de CHIC, « la réduction des variables », donne les résultats suivants :

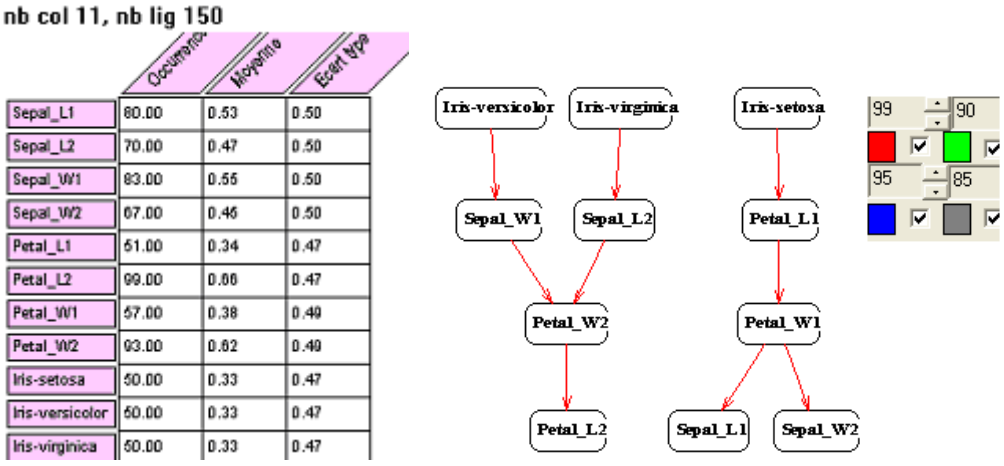
Calcul avec un seuil d'équivalence fixé à 0.9 :
Sepal_L1 Petal_W1 Sepal_W3 Iris-setosa Petal_L1 représentant de la classe : Petal_W1
Iris-versicolor Petal_W2 Petal_L2 Sepal_W1 Sepal_L2 représentant de la classe : Petal_W2
Petal_L3 Sepal_L3 Iris-virginica Petal_W3 représentant de la classe : Iris-virginica

Cette fois, toutes les variables sauf Sepal_W2 se retrouvent « équivalentes au seuil de 0.9 » à une espèce d'iris. Cette information, qui ne fait que conforter celle donnée par le graphe implicatif dans le cas des iris, peut s'avérer très utile quand on dispose d'un grand nombre de variables dont certaines très liées entre elles, pour obtenir un graphe implicatif avec un nombre réduit de « classes de variables ».

Découpage de chacune des 4 dimensions de la fleur en 2 parties

Le découpage automatique de petal_L, petal_W, sepal_L et Sepal_W en 2 se fait selon les bornes suivantes :

	Sepal_L	Sepal_W	Petal_L	Petal_W
Partie 1	de 4.3 à 5.8	de 2 à 3	de 1 à 3	de 0.1 à 1
Partie 2	de 5.9 à 7.9	de 3.1 à 4.4	de 3.3 à 6.9	de 1.1 à 2.5



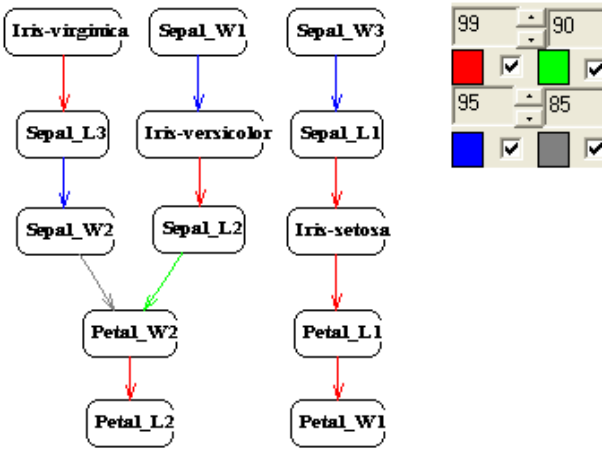
nb col 11, nb lig 150

	Occurrence	Moyenne	Ecart type
Sepal_L1	80.00	0.53	0.50
Sepal_L2	70.00	0.47	0.50
Sepal_W1	83.00	0.55	0.50
Sepal_W2	67.00	0.46	0.50
Petal_L1	51.00	0.34	0.47
Petal_L2	99.00	0.66	0.47
Petal_W1	57.00	0.38	0.49
Petal_W2	93.00	0.62	0.49
Iris-setosa	50.00	0.33	0.47
Iris-versicolor	50.00	0.33	0.47
Iris-virginica	50.00	0.33	0.47

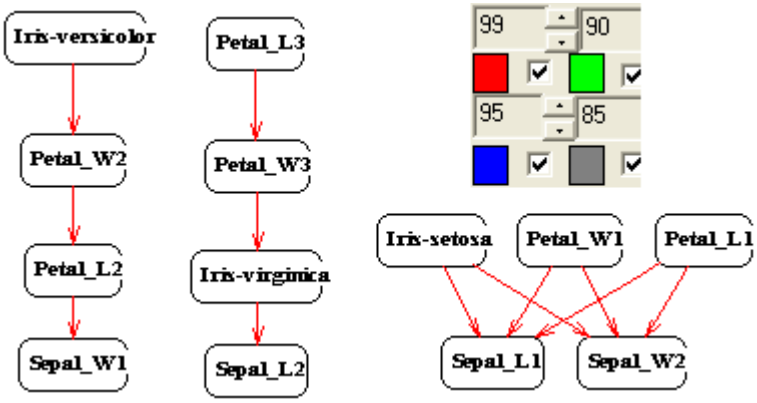
Ci-dessus, à gauche, les « informations sur le fichier » montrent que le choix de découpage fait par CHIC (selon la variance) produit pour certaines variables une répartition inégalitaire des effectifs (51 pour petal_L1 contre 99 pour pétal_L2). Ci-dessus, à droite, son graphe implicatif, qui ne permet pas de différencier Iris-versicolor de iris-verginica.

Nous avons ensuite testé trois autres découpages des variables, dont seul le deuxième a été jugé satisfaisant, car ayant fourni un graphe séparé par type d'iris :

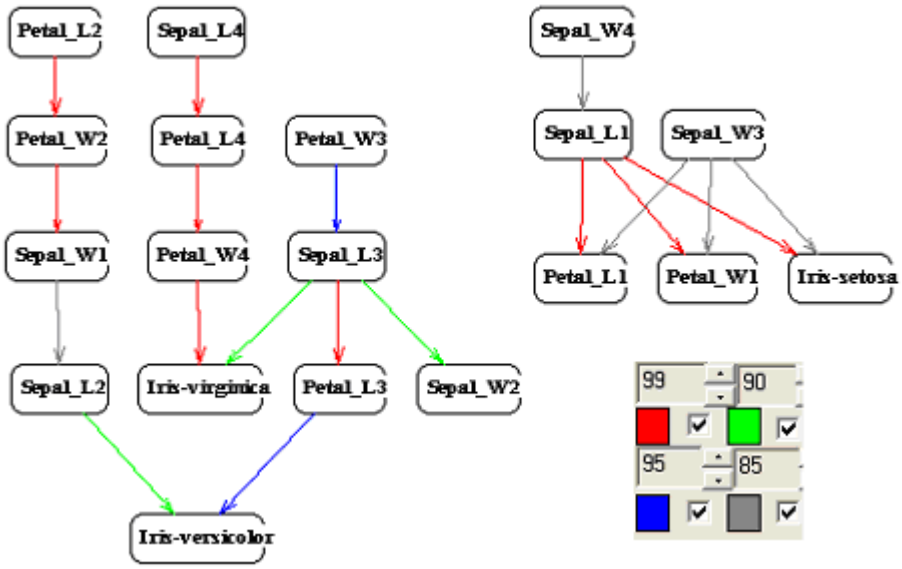
Découpage automatique de petal_L et petal_W en 2, de sepal_L et Sepal_W en 3



Découpage automatique de petal_L et petal_W en 3, de sepal_L et Sepal_W en 2



Découpage automatique de petal_L, petal_W, sepal_L et Sepal_W en 4

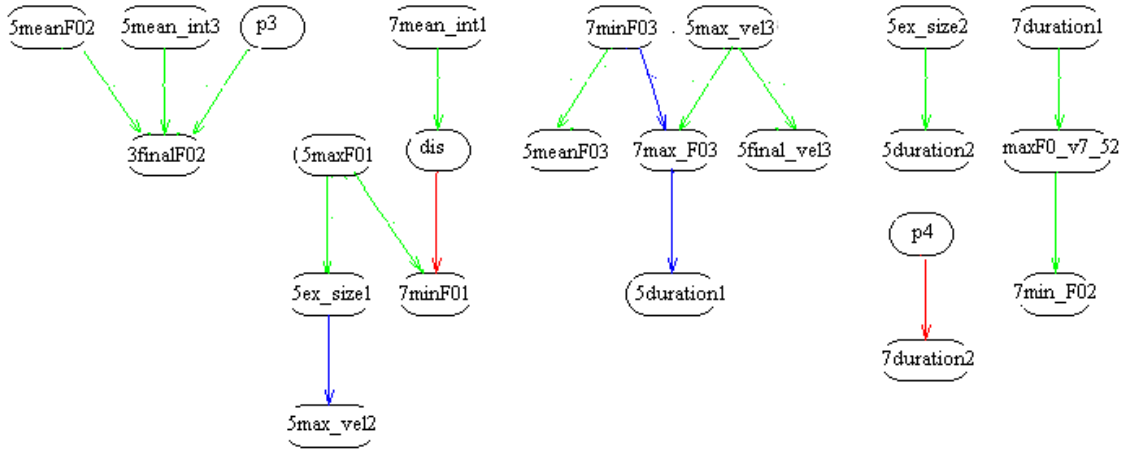


Annexe de la partie 3 : autres résultats sur l'intonation avec CHIC

Réduction des variables découpées en 3, avec recollements, au seuil de 0.9

3meanf01 3mean_int1 3finalf01 5finalf01 3maxf01 représentant de la classe : 3maxf01
maxf0_v7_51 7finalf01 7meanf01 dis 7final_vel1 7maxf01 représentant de la classe : dis
cis 3mean_int3 7mean_int3 5mean_int3 représentant de la classe : 5mean_int3
5duration2 7ex_size1 représentant de la classe : 5duration2
3minf03 5minf03 3maxf03 5maxf03 3meanf03 5meanf03 5finalf03 3finalf03 représentant de la classe : 5meanf03
qis maxf0_v7_53 7maxf03 7finalf03 7meanf03 7final_vel3 7ex_size3 représentant de la classe : 7maxf03
7meanf02 7maxf02 7final_vel2 maxf0_v7_52 7finalf02 représentant de la classe : maxf0_v7_52
3duration1 p4 représentant de la classe : p4
5maxf01 maxf0_v5_31 5meanf01 5minf01 représentant de la classe : 5maxf01
3maxf02 5finalf02 3meanf02 5meanf02 5minf02 5maxf02 représentant de la classe : 5meanf02
5final_vel3 5ex_size3 maxf0_v5_33 représentant de la classe : 5final_vel3
3finalf02 3minf02 représentant de la classe : 3finalf02
7duration3 p2 représentant de la classe : 7duration3
3final_vel3 3ex_size3 représentant de la classe : 3ex_size3
7ex_size2 7minf03 représentant de la classe : 7minf03

et le graphe implicatif avec les variables réduites et les seuils de 99.9, 99 et 95



APPORT DE LA COMBINAISON DE LA METHODE D'ANALYSE STATISTIQUE IMPLICATIVE (A.S.I.) AVEC LA THEORIE DE REPONSES AUX ITEMS (IRT).

Hayette KHALED[1], Raphael COUTURIER[2]

CONTRIBUTION OF STATISTICAL IMPLICATIVE ANALYSES (SIA) COMBINED WITH THE ITEM RESPONSE THEORY (IRT).

RÉSUMÉ

Dans ce papier, nous illustrons les avantages de la combinaison de la méthode d'Analyse Statistique Implicative (ASI) avec la Théorie de Réponses aux Items (IRT) à travers un exemple d'analyse des résultats des étudiants en Informatique de l'université A/Mira de Bejaia. L'ASI est utilisée pour découvrir et analyser les implications les plus pertinentes entre les différents modules de formation étudiés. L'IRT, quant à elle, permet l'analyse de la qualité des items et leur calibrage (difficulté) en permettant la construction d'échelles quasiment indépendantes des données d'évaluation (étudiants et modules). Une analyse des résultats est proposée ainsi qu'une comparaison avec les résultats que nous avons obtenus dans le papier Khaled *et al*. (2014) dans lequel nous avons appliqué l'ASI aux notes des étudiants.

Mots-clés : *ASI, IRT, Notes des Etudiants.*

ABSTRACT

In this paper we illustrate the benefits of combining Statistical Implicative Analysis (SIA) method with the Items Response Theory (IRT) throw a study of a sample of computer science students of Bejaia University. SIA is used to discover and analyze the most relevant implications between different studied modules. The IRT, in turn, allows the analysis of the items quality and their calibration (difficulty) by allowing the construction of scales independent from the test data (students and modules). An analysis of the results is proposed with a comparison with the results that we have obtained in the paper khaled*and al.*(2014) in which we applied the SIA on students' marks.

Keywords : *SIA, IRT, Students' Marks.*

1 Introduction

La théorie de réponses aux items (IRT) est un modèle statistique de la mesure relativement récent Pini (2012), elle permet de faire face à des problèmes auxquels la psychométrie classique n'apporte pas toujours des réponses et des solutions satisfaisantes, elle s'efforce de produire une estimation des propriétés de l'item quasiment indépendante d'un groupe particulier d'individus. En d'autres termes, elle cherche à élaborer des instruments de mesure dont les caractéristiques ne soient pas excessivement influencées par tel ou tel autre groupe de référence.

[1]U A/Mira Béjaia, Targa Ouzemour, hayette.khaled@univ-bejaia.dz ou aya_info6@yahoo.fr
[2]IUT Belfort-Montbéliard, 19 Avenue du Maréchal Juin BP 527, raphael.couturier@univ-fcompte.fr

L'IRT repose sur des modèles de type probabiliste. Ces modèles sont fondés sur le principe que la réponse d'un individu à l'item (et notamment sa probabilité de fournir une réponse correcte) est déterminée par deux sortes de facteurs d'une part, la compétence du sujet (qualifié de traits latents) ; d'autre part, les propriétés de l'item lui-même : son degré de difficulté, son pouvoir de discrimination ainsi que la "chance" qui peut jouer un rôle dans certains cas.

L'IRT a été proposée pour la première fois dans le domaine de la psychométrie, dans le but de l'évaluation de la capacité Xinming et Yiu-Fai. (2014). Elle est par la suite utilisée dans différents domaines comme la médecine Michael (2013), le tabagisme chez les adolescents Hedeker*et al*. (2006), etc. En outre elle est largement utilisée dans l'éducation pour calibrer et évaluer les items dans les tests ainsi que la capacité des sujets. Au cours des dernières décennies, l'évaluation pédagogique a utilisé de plus en plus les techniques à base d'IRT afin de développer des tests.

Par exemple, dans Johns *et al*. (2006) les auteurs ont conçu plusieurs expériences pour tester les modèles d'IRT qui ont été utilisés comme un outil pour la modélisation des élèves dans un système de tutorat intelligents. Les modèles ont été testés en utilisant des données réelles des élèves du secondaire en utilisant le système de tutorat Outpost Wayang. Un modèle de validation croisée a été développé et trois métriques de mesure de la précision de la prédiction ont été comparées. Les résultats ont montré que les modèles formés prédisent avec une précision de 72% si un étudiant répond correctement à un problème de choix multiples ainsi, nous pouvons dire que ces modèles ont un bon pouvoir de prédiction.

Dans Hamdare (2014) les auteurs proposent un système d'évaluation adaptif se basant sur l'IRT. Ce système permet de personnaliser dynamiquement les questions pour les élèves en fonction de leur réponse à la question précédente, mais en plus de cela, et ce qui est très intéressant, il permet aussi d'ajuster le degré de difficulté des questions d'évaluation en fonction de la capacité de l'élève ; ainsi cela aidera les enseignants à acquérir une mesure valide et fiable des compétences des étudiants.

Dans Rowan (2001) les chercheurs de l'Université de Michigan travaillent sur l'élaboration de mesures fondées sur des enquêtes concernant le contenu pédagogique des connaissances des enseignants, en identifiant trois dimensions à mesurer, à savoir : le contenu de la connaissance, la connaissance de la réflexion des élèves ainsi que la connaissance des stratégies pédagogiques ; afin d'analyser leurs effets sur la réussite des étudiants. Les chercheurs établissent des questionnaires pour les enseignants, concernant les trois dimensions citées auparavant puis appliquent le programme commercial informatique BILOG pour l'analyse et la notation des réponses aux questionnaires qui se basent sur l'IRT pour exclure les questions qui diminuent la fiabilité de l'ajustement.

L'équipe de recherche de l'IREM d'Aix-Marseille développe une méthode adaptive d'évaluation sur ordinateur Bodin *et al*. (2010) qu'elle teste sur le socle commun en termes de connaissances et de compétences. Elle cherche ainsi à répondre aux questions suivantes: quelles sont les connaissances et les compétences nécessaires à tous les individus, indépendamment du cursus de formation que chacun pourra avoir suivi ou non ? Comment est-il possible de positionner des individus par rapport aux demandes du socle ?

La méthode adaptive consiste à former un référentiel R qui contient l'ensemble des connaissances et des compétences à évaluer. Un ensemble de questions Q est associé au référentiel R ; et l'arbre implicatif cohésif de CHIC Couturier et Gras (2005) ou Couturier (2008) est utilisé afin de déterminer les questions qui correspondent à la même unité de compétence.

Le principe du test est de prendre au hasard la première question (q1,p) à poser à la personne p puis la question suivante sera choisie selon le graphe implicatif des questions de ASI. Le test est arrêté lorsque l'amplitude de l'intervalle de confiance du score estimé devient inférieure à un intervalle de confiance choisi.

Le test adaptatif sur R ici consiste à estimer le score d'une personne sur l'ensemble de questions Q à partir de ses réponses à un sous ensemble Q' en utilisant la fonction d'estimation probabiliste de l'IRT. Les auteurs montrent que la combinaison de l'ASI et de l'IRT permet d'envisager des tests très fiables.

Tous les travaux qui utilisent l'IRT montrent que cette dernière a des apports très intéressants, sur tous ceux qui s'inscrivent dans le domaine de l'éducation qui est notre domaine d'intérêt. Ces travaux s'intéressent en général au raffinement de questionnaires d'un système du tutorat ou d'un système web d'essai ou l'étude des connaissances des enseignants et leurs effets sur la réussite des étudiants.

Il y a peu de travaux qui traitent les notes de cursus des étudiants. Dans notre travail Khaled *et al*. (2014), nous utilisons l'analyse statistique implicative (ASI) Gras et al (2008) et Gras et al (2013) afin d'identifier les implications pertinentes entre les modules.

Notre objectif dans ce papier est de montrer les avantages et apports des deux méthodes ASI et IRT à travers un exemple qui porte sur les notes des étudiants de département informatique de l'université A/Mira de Bejaia. Nous avons cité auparavant un travail qui fait aussi la combinaison de ces deux méthodes sur des questions de tests dont les réponses représentent des données binaires. Dans notre cas nous traitons des données polytomiques, c'est-à-dire avec plusieurs modalités, obtenus en divisant les notes des étudiants en trois groupes par l'algorithme des nuées dynamique pour représenter les étudiants avec des résultats faibles, des résultats moyens et de bons résultats. Ainsi même si les notes des étudiants sont globalement élevées ou au contraire très faibles, nous aurons les 3 groupes précédemment énoncés.

Dans la suite du papier nous introduisons la théorie de réponses aux items (IRT) puis nous montrons notre expérience et ses résultats. Finalement nous présentons quelques résultats obtenus avec l'ASI présentés dans Khaled *et al*. (2014) ainsi qu'une comparaison entre l'IRT et l'ASI.

2 Théorie de Réponses aux Items (IRT) Pini (2012),Hamdare (2014) Dai-Trang (2013),Rizopoulo (2006)

Les modèles d'IRT sont des fonctions mathématiques appelées fonction caractéristique d'item qui spécifient la probabilité de répondre correctement à un item, en termes de paramètres d'items et d'individus. Ils sont représentés graphiquement par une courbe continue appelée la Courbe Caractéristique d'Item (CCI) (Figure 1), elle

décrit le lien qui existe entre la situation d'un individu par rapport au trait latent, représentée par θ et la probabilité que cet individu réussi l'item i.

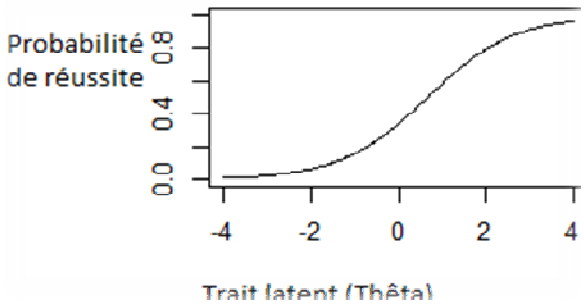


Figure 1 – La Courbe Caractéristique d'Item (CCI)

Il existe trois modèles d'IRT : le modèle à un seul paramètre (ou modèle de Rasch), le modèle à deux paramètres et le modèle à trois paramètres. Les trois modèles ont un paramètre de difficulté d'item représenté par b_i.

– l'équation du modèle à un seul paramètre est donnée comme suit :

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \tag{1}$$

– l'équation du modèle à deux paramètres est donnée comme suit :

$$P_i(\theta) = \frac{e^{\alpha_i(\theta - b_i)}}{1 + e^{\alpha_i(\theta - b_i)}} \tag{2}$$

– l'équation du modèle à trois paramètres est donnée comme suit :

$$P_i(\theta) = C_i + \frac{(1 - C_i)e^{\alpha_i(\theta - b_i)}}{1 + e^{\alpha_i(\theta - b_i)}} \tag{3}$$

En plus, le modèle à deux paramètres et celui à trois paramètres possèdent un paramètre de discrimination noté α lié à la rapidité de changement de la probabilité avec les changements en capacité. Le modèle à trois paramètres contient un troisième paramètre, appelé paramètre de pseudo-chance noté *C* qui représente la probabilité qu'un sujet de très faible capacité répondra correctement à un item *i* en choisissant aléatoirement l'une des options proposées. Il y a deux branches principales de l'IRT : l'unidimensionnelle (UIRT) et la multidimensionnelle (MIRT). La première branche opère sur l'hypothèse qu'il existe une idée ou un message à travers le contenu de tous les items et les instruments de test sont mis au point pour mesurer cette idée unique. La deuxième branche, quant à elle, suppose que les items sont regroupés dans des

domaines différents, et les instruments sont utilisés pour mesurer ces multiples constructions de domaine Dai-Trang (2013). Les modèles les plus courants d'IRT sont des modèles dichotomiques avec deux catégories de réponses pour les items et les modèles polytomiques avec des catégories multiples et ordinales de réponses pour les items. Dans notre cas nous allons traiter les notes des étudiants qui sont entre zéro et vingt qui seront divisées en trois groupes donc elles représentent des données polytomiques. Elles incluent les données d'items à choix multiples, des questions mathématiques ouvertes, les échelles de Likert, les items ordinaux, les réponses d'évaluation d'échelles et les réponses graduées de test ou de questions d'enquête. Le type de modèle IRT utilisé pour décrire l'interaction entre les sujets et les items de test dépend de la nature des données collectées. Plusieurs modèles ont été utilisés pour les données polytomiques, tels que le modèle de crédit partiel (PCM), le modèle de crédit partiel généralisé (GPCM), le modèle de la réponse nominale (NRM) et le modèle de la réponse graduée (GRM). Ce dernier modèle est proposé par Samejima (1969), c'est le mieux approprié pour le traitement des notes des étudiants, sa forme logistique est exprimée comme suit:

$$P(Y_{ijk} = k|\theta_j, b_{ik}, a_i) = \frac{1}{1+e^{-a_i(\theta_i - b_{i,k})}} - \frac{1}{1+e^{-a_i(\theta_i - b_{i,k+1})}} \quad (4)$$

3 Analyse et résultats

3.1 Description des données

La population considérée représente les étudiants de licence 2 de département Informatique de l'université A/Mira de Bejaïa pendant les années 2010, 2011 et 2012. Les étudiants de licence 2 sont évalués sur les modules programmation linéaire (PL), théorie de langages (THL), logique mathématique (LogMat), analyse numérique (ANum), système d'exploitation (SE), probabilités et statistiques (P.S), traitement de signal (TSig), génie logiciel (GL), algorithme et structure de données (ALSTRD2), architecture (ARCH), bases de données (BDD) et structure de données (STRD1).

Les notes des étudiants sont représentées par un fichier de type « .csv » (un exemple est montré dans le tableau 1). Ce fichier contient en lignes les matricules des étudiants et en colonnes les modules suivis par ces derniers ; et la variable « p » qui suit chaque intitulé de module pour dire que les notes de modules sont à partitionner en un nombre fixe d'intervalles, nous utilisons l'algorithme des nuées dynamiques Diday (1971) qui constitue automatiquement les intervalles qui ont des limites distinctes.

	ARCH p	STRD1 p
0809TMI02	11,25	11,5
09MI0034	9,38	11,38
09MI0590	10,63	9,38
…	…	…

Tableau 1 – Exemple de données de type .csv

Pour traiter ces données avec l'IRT nous avons choisi le modèle de la réponse graduée (GRM) à deux paramètres qui permet de donner la difficulté des modules et une

bonne discrimination de ces derniers. Le choix de ce modèle est dû au fait que les notes des étudiants sont divisées en trois catégories (résultats faibles, moyens et bons) et elles représentent ainsi des données polytomiques qui peuvent être traitées par le modèle GRM.

L'analyse des résultats a été effectuée en traçant les courbes caractéristiques des modules comme montré sur la figure suivante :

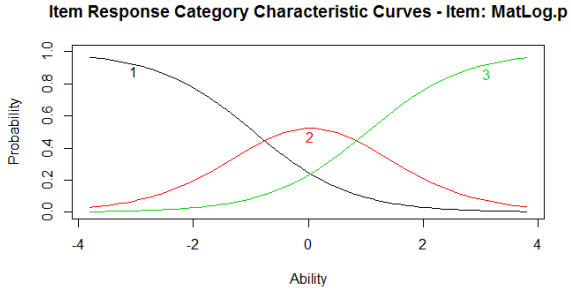


Figure 2 – Courbe Caractéristique d'Item obtenue par le modèle GRM.

La courbe 1 (en noir) représente la probabilité d'échouer le module en fonction de la capacité de l'étudiant, la courbe 2 (en rouge) représente la probabilité d'avoir un résultat moyen dans le module en fonction de la capacité de l'étudiant et la courbe 3 (en vert) représente la probabilité de réussir le module toujours en fonction de la capacité de l'étudiant. Cette figure montre que les étudiants avec de faibles résultats (capacité entre - 4 et - 2) ont une grande probabilité d'échouer le module et une probabilité presque nulle de réussir le module, contrairement aux étudiants avec de bons résultats (capacité entre 2 et 4) qui ont une grande probabilité de réussir le module et une probabilité presque nulle d'échouer au module.

3.2 Résultats du modèle IRT et analyse

Dans ce qui suit nous présentons quelques courbes caractéristiques d'items (CCI) (un item représente un module) en licence 2 durant les années scolaires 2010-2011 (Figure 3), 2011-2012 (Figure 4) et 2012-2013 (Figure 5).

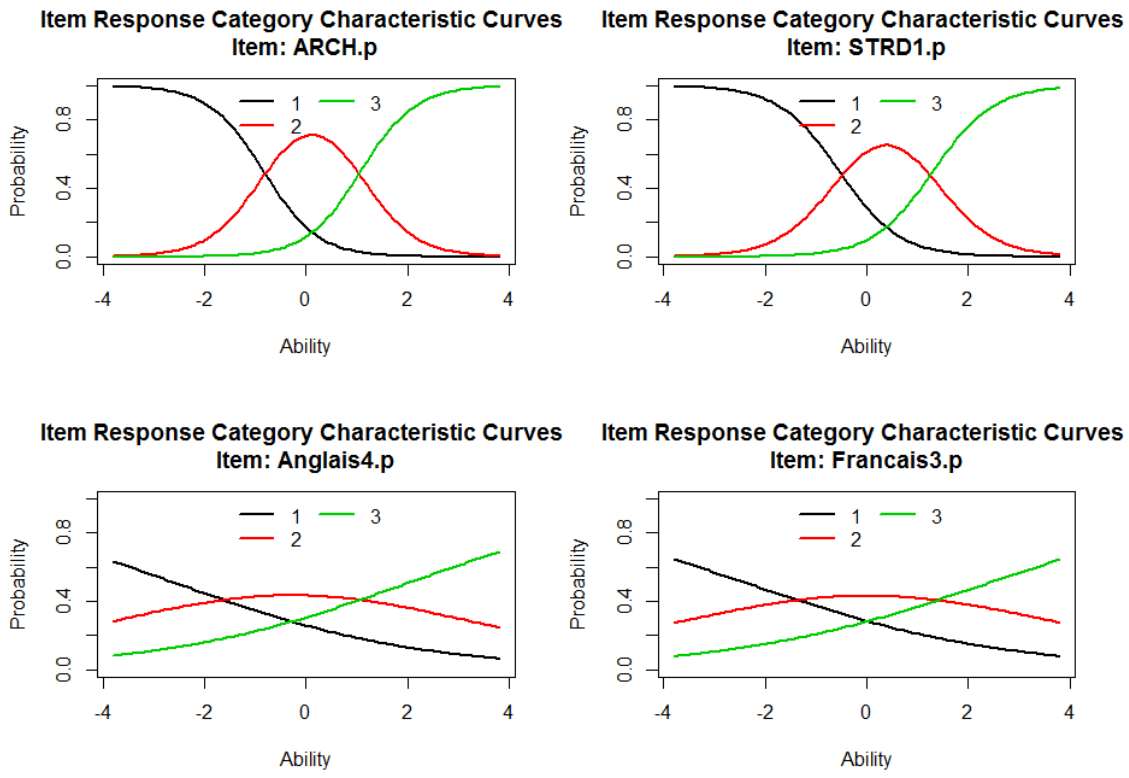


Figure 3 (suite) – CCI de licence 2 2010-2011

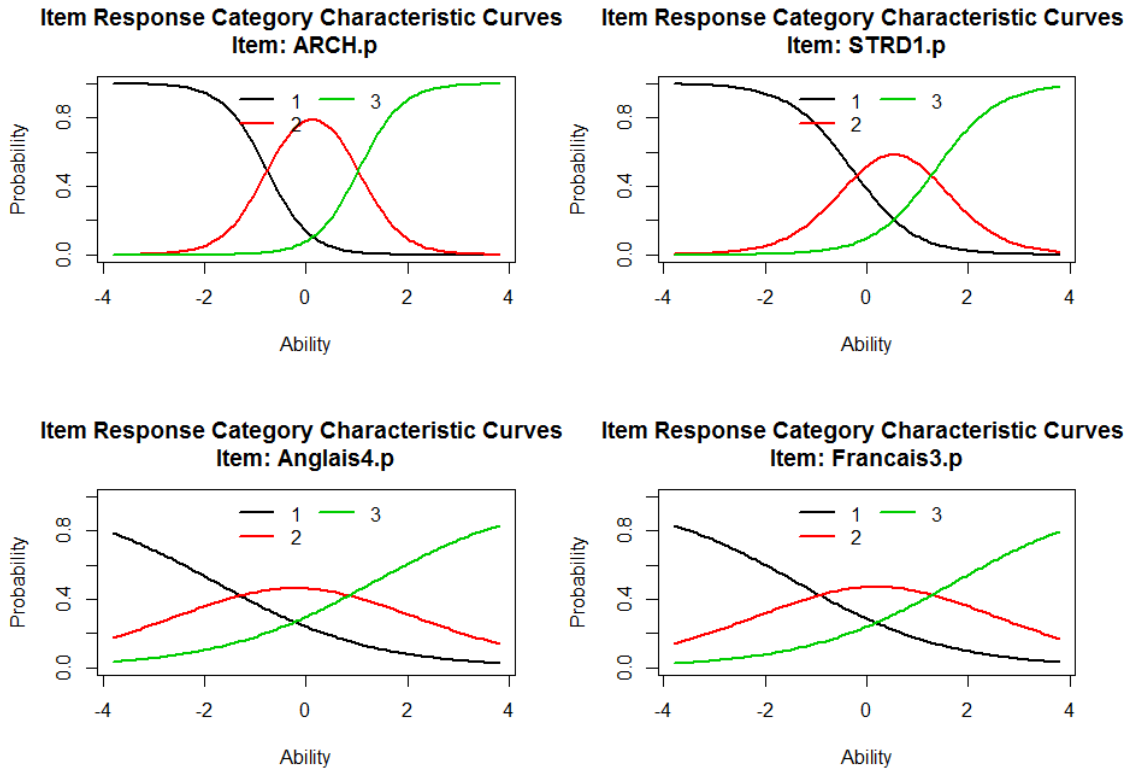


Figure 4 – CCI de licence 2 2011-2012

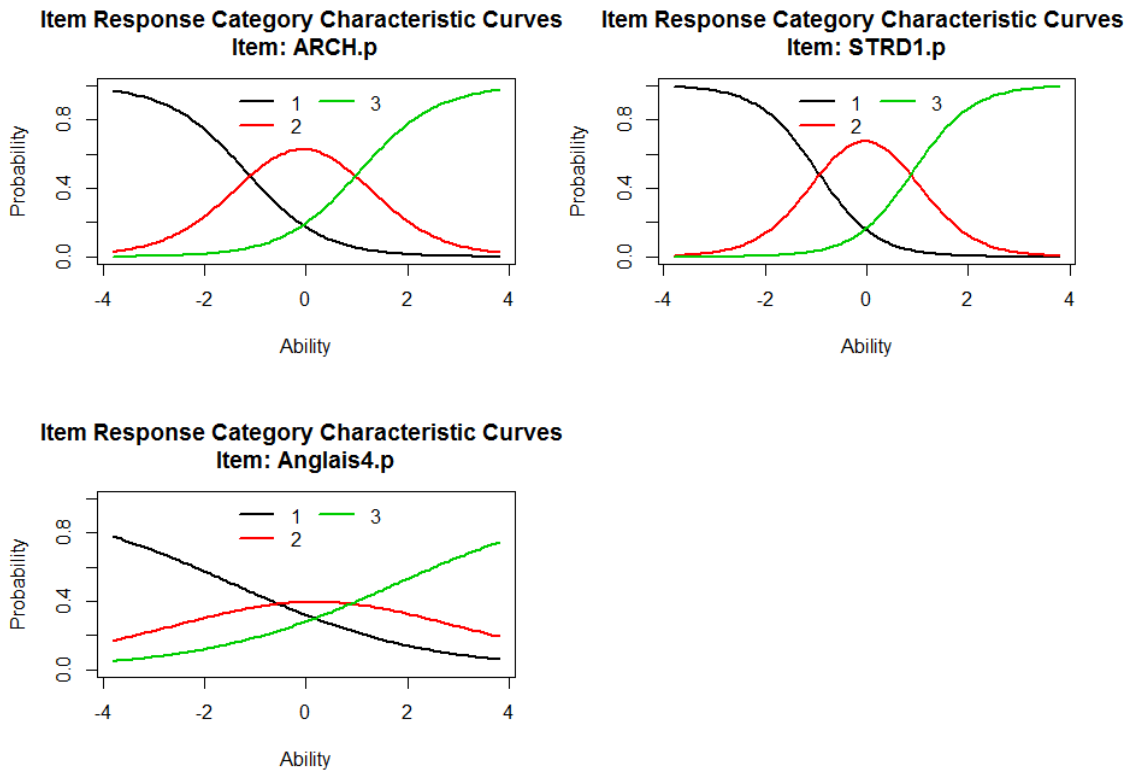


Figure 5 – CCI de licence 2 2012-2013

D'après les figures 3, 4 et 5 présentées précédemment et d'après les figures 6, 7 et 8 qui sont mises en annexe, nous observons que nous avons pratiquement les mêmes résultats pour les différentes années. Nous voyons bien que les modules de spécialité ARCH et STRD1(figures 3, 4 et 5), ANum, P.S, LogMat, TSig, ALSTRD2, BDD, SE, THL, PL et GL(figures 6, 7 et 8) sont les plus discriminants[3] (nous observons des courbes plus raides), autrement dit ces modules font bien la différence entre les étudiants avec de faibles résultats et ceux avec de bons résultats contrairement aux modules Anglais4 et Français3 (figures 3, 4 et 5), ainsi que SI, Anglais3 et Français4[4] (figures 6, 7 et 8) qui ne le sont pas.

Nous savons qu'en général les étudiants avec de bons résultats en informatique ont de bons résultats dans tous les modules de spécialité et les étudiants avec de faibles résultats en informatique ont des résultats similaires dans tous les modules de spécialité. Ceci est confirmé par les résultats présentés ci-dessus.

Les étudiants avec de bons résultats, c'est-à-dire avec une capacité supérieure à 2 par exemple ; ont une probabilité très proche de 1 de réussir les modules de spécialité et une probabilité très proche de 0 d'échouer aux modules de spécialité. Par contre les étudiants avec de faibles résultats, ceux qui ont une capacité inférieure à - 2 par

[3]Nous donnons entre parenthèses le paramètre de discrimination de chaque module :ARCH(1.867), STR1D(1.693), ANum(1.353), P.S(1.273), LogMat(1.202), TSig(1.287), ALSTRD2 (1.614), BDD (1.639), SE (1.466), THL (1.305), PL (1.272), GL (1.428)

[4]SI (0.483), Anglais3 (0.409), Anglais4 (0.421), Francais3 (0.400), Francais4 (0.062)

exemple, ont une probabilité très proche de 1 d'échouer aux modules de spécialité et une probabilité presque nulle de réussir les modules de spécialité.

Le fait d'être bon en informatique n'implique pas être bon en langues (français et anglais) et, de la même manière, le fait d'être faible en informatique n'entraîne pas d'être faible en langue. Nous observons exactement cela sur les résultats obtenus dans les figures 3, 4 et 5 et nous constatons que ces modules ne sont pas discriminants. Par rapport au module système d'information (SI)(figures 6, 7 et 8 en annexe), nous remarquons aussi qu'il n'est pas discriminant et cela s'explique par le fait que ce n'est pas un module de spécialité informatique, il porte sur l'organisation des entreprises, leur gestion, etc. Pour les étudiants qui ont des résultats moyens (dont la capacité est entre -2 et 2), les courbes numéros 2 (rouge) nous montrent qu'ils ont une probabilité moyenne (entre 0.5 et 0.7) de réussir les modules de spécialité.

3.3 Résultats de la méthode ASI et analyse

Dans ce qui suit nous présentons quelques graphes implicatifs résultants de l'application de l'analyse statistique implicative (ASI) sur les même notes d'étudiants présentée ci-dessus, ce travail a fait l'objet d'un papier Khaled *et al* (2014) publié à ISKO-Maghreb (2014).

Les intervalles des notes des étudiants ici sont également partagés en trois sous intervalles identifiant les étudiants avec des résultats faibles, moyens et bons. Par exemple, le module STRD1 a été partitionné selon les intervalles suivants :

STRD11 de 5.94 à 10.38, STRD12 de 10.5 à 13.38, STRD13 de 13.5 à 18.38

STRD11 reflète les étudiants qui sont faibles en STRD1, STRD12 les étudiants qui sont moyens et STRD13 les étudiant qui sont bons en STRD1.

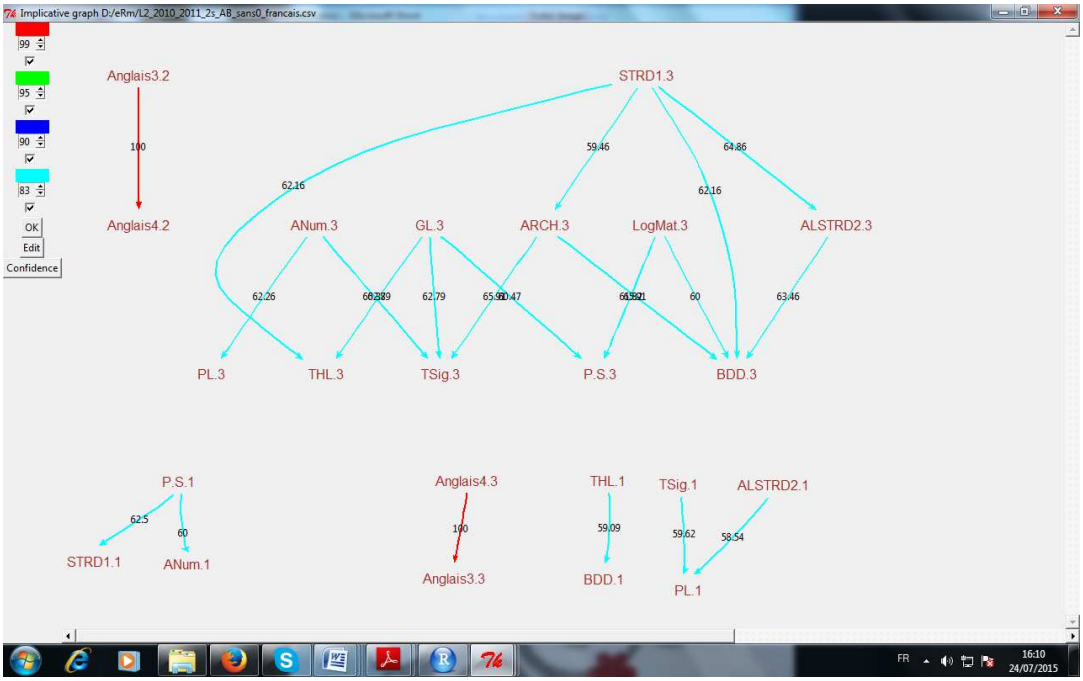


Figure 9– Graphe implicatif L2 2010-2011

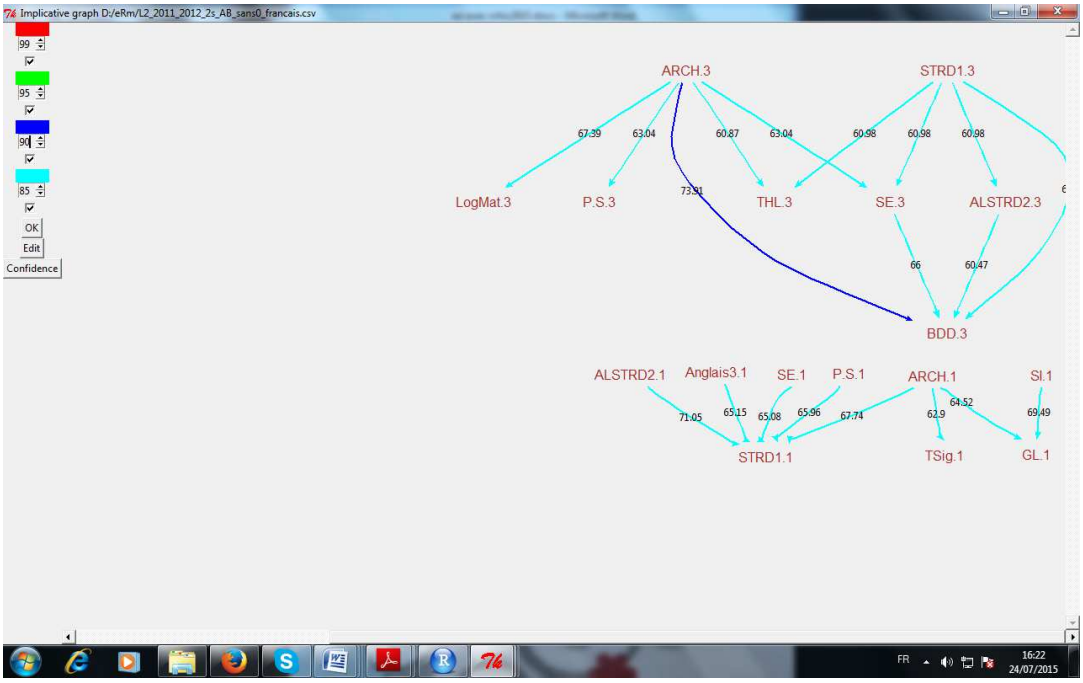


Figure 10 – Graph implicatif L2 2011-2012

Le critère utilisé pour la recherche des implications pour les graphes implicatifs est l'implifiance présenté dans GRAS *et al* (2015), ce dernier présente une combinaison ou un mariage entre les deux critères confiance et intensité d'implication.

Nous voyons sur ces figures que les arcs sont pondérés par un poids ce qui représente la « confidence en anglais », si nous prenons l'exemple de la confiance de 71,05 entre les modules ALSTRD2.1 et STRD1.1 dans la figure 10 ceci signifie que si un étudiant est faible en ALSTRD2 il a une chance de 71,05% d'être aussi faible en STRD1. Toutes les implications que nous avons obtenues sont intéressantes car elles ont presque toutes des confiances supérieures à 60.

Nous observons sur les deux figures 9 et 10 des implications qui se répètent sur deux années consécutives. Nous remarquons clairement qu'un module réussi par les bons élèves entraine d'autres modules réussis par les bons élèves (par exemple STRD1.3 -> ALSTRD2.3, STRD1.3 -> THL3, ALSTRD2.3 -> BDD3). Il en est de même pour les modules échoués par les élèves faibles qui impliquent d'autres modules échoués par des élèves faibles (par exemple PS1 -> STRD1.1). Nous ne voyons pas sur les figures les étudiants moyens, nous expliquons ceci par le fait qu'un étudiant qui est moyen dans un module peut être faible ou bon dans d'autres modules et il n'est pas forcément moyen dans tous les modules.

3.4 Comparaison des résultats de l'ASI et l'IRT

Dans le papier Khaled *et al* (2014) nous avons présenté les résultats de l'application de l'analyse statistique implicative (ASI) sur les notes des étudiants nous avons reporté quelques figures (9 et 10) dans le présent papier. Les résultats obtenus montrent que les implications entre les modules se répètent généralement sur les trois promotions (2010_2011, 2011_2012 et 2012_2013). Ces implications sont intéressantes car elles

correspondent aux réussites ou aux échecs des étudiants (les modules réussis impliquent d'autres modules réussis et les modules échoués par les étudiants impliquent d'autres modules échoués). Nous avons aussi noté des implications entre les modules qui sont liés par leur contenu ou par l'enseignant qui assure le cours.

Nous remarquons que l'ASI nous donne des implications très intéressantes entre les modules, elle nous permet de dire si un étudiant maitrise un module X alors il aura tendance à maitriser un autre module Y.

La Théorie de Réponse aux Items (IRT) quant à elle, nous montre quels sont les modules qui sont discriminants (en fonction de la capacité des étudiants) et de plus elle nous indique la difficulté des modules.

D'après les résultats obtenus dans Khaled *et al* (2014) et ceux de ce papier, nous constatons que l'Analyse Statistique Implicative et la Théorie de Réponses aux Items nous donnent des informations différentes. La première nous donne les implications entre les modules. La seconde nous informe sur la probabilité d'un étudiant de réussir ou non un module selon ses capacités et la difficulté de module. Ainsi l'ASI et l'IRT sont bien complémentaires et offrent des indications précieuses pour analyser les résultats des étudiants du département concerné. Nous encourageons les enseignants à analyser les résultats de leurs élèves ou étudiants avec l'ASI et l'IRT afin d'avoir des informations leur permettant de mieux comprendre les résultats des élèves ou étudiants.

4 Conclusion

Nous avons appliqué la méthode d'Analyse Statistique Implicative et la Théorie de Réponses aux Items sur les notes des étudiants de l'université A/Mira de Bejaia à travers les trois promotions 2010-2011, 2011-2012 et 2012-2013.

Les notes ont été partitionnées en trois intervalles en utilisant l'algorithme des nuées dynamiques pour avoir les étudiants avec de bons résultats, les étudiants avec des résultats moyens et les étudiants avec des résultats faibles. Ensuite nous avons utilisé les deux méthodes ASI et IRT pour pouvoir comparer leurs résultats.

Avec l'ASI nous avons utilisé le logiciel CHIC pour tracer les graphes implicatifs qui nous donne les implications pertinentes entre les différents modules. Pour l'IRT nous avons utilisé le modèle GRM à deux paramètres pour estimer et tracer les courbes caractéristiques d'items.

Les résultats des deux méthodes montrent que ces dernières sont complémentaires. Ainsi l'ASI nous donne les implications les plus pertinentes entre les modules ; alors que l'IRT nous donne la difficulté des modules et nous montre que les étudiants avec de bons résultats ont de bons résultats dans tous les modules de la spécialité. De même, les étudiants avec de faibles résultats sont également de faibles résultats dans tous les modules de la spécialité.

Nous conseillons aux enseignants d'utiliser ces méthodes pour analyser les notes de leurs étudiants afin de déterminer la difficulté des modules, afin d'améliorer les cours ou les sujets d'examens par exemple. En outre, ces deux méthodes peuvent être utilisées pour raffiner les modules d'une spécialité, elles permettent de déterminer les modules qui ont une relation avec la spécialité. Ainsi, il est possible d'effectuer des choix pourles modules d'une formation.

Dans un futur proche, nous comptons tester les deux méthodes sur des données plus volumineuses comme celles de Programme International pour le Suivi des Acquis des élèves (PISA). Ces données sont rassemblées dans le cadre de l'enquête internationale PISA qui est menée tous les trois ans depuis 2000 auprès des jeunes de 15 ans, dans la plupart des pays membres de l'OCDE et dans de nombreux pays partenaires. Cette enquête évalue l'acquisition de savoirs et savoir-faire essentiels à la vie quotidienne au terme de la scolarité obligatoire. Ce travail sera également l'occasion d'évaluer les performances d'autres méthodes d'apprentissage automatique à travers l'environnement statistique R.

Références

[1] Bodin, A et al (2010), Vers un test adaptatif - criterié, combinant l'utilisation de l'IRT et de l'ASI, pour l'évaluation du socle commun de connaissances et de compétences, *Quaderni di Ricerca in Didattica (Mathematics)*,n°20, 383-409.

[2] Couturier, R. (2008), Statistical Implicative Analysis. In CHIC: CohesiveHierarchical Implicative Classification, volume 127 of Studies in Computational Intelligence, Springer ; 41-52.

[3] Couturier, R. Gras, R. (2005), CHIC: Traitement de données avec l'analyse implicative, *Revue des nouvelles technologies de l'information*, n°3, 679-684.

[4] Dai-Trang, L. (2013), *Applying item responsetheorymodeling in educationalresearch*, These: Philosophie, Ames Iowa aux États-Unis : Université d'État de sciences et technologie de l'Iowa, 187 p.

[5] Diday, E. (1971), La méthode des nuées dynamiques, *Journal revue de statistique appliqué*, **n°** 19 (2), 19-34.

[6] GrasR., SuzukiS., GuilletF., Spagnolo F. (2008) F. Statistical Implicative Analysis, Theory and Applications, eds, Springer.

[7] Gras R., Régnier J.-C., Marinica C., Guillet F. (2013) L'analyse statistique implicative, Méthode exploratoire et confirmatoire à la recherche de causalités, sous la direction de Gras R., edsCépaduès Editions, 522 pages, ISBN 978.2.36493.056.8.

[8] Gras R., Couturier R., Gregori P. (2015), Un mariage arrangé entre l'implication et la confiance ?, *A.S.I.8*.

[9] Hamdare, S. (2014), An Adaptive Evaluation System to Test Student Caliber using Item Response Theory, *IJMTER*, 329-333.

[10] Hedeker, D et al. (2006), Application of Item Response Theory Models for Intensive Longitudinal Data,*in T.A. Walls & J.L. Schafer (Eds.), Models for Intensive Longitudinal Data*,Oxford University Press, New York, 84-108.

[11] Johns, J. Mahadevan, S. Woolf, B. (2006), Estimating Student Proficiency Using an Item Response Theory Model, In : Mitsuru Ikeda, Kevin D. Ashley, Tak-Wai Chan. Intelligent Tutoring Systems. Taiwan: Springer, 473-480.

[12] Khaled, H. Ghanem, S. Couturier, R. (2014), Analysis of Bejaia University Computer Science students' marks through the CHIC software and Statistical Implicative Analysis, *ISKO-Maghreb*, (n° 94).

[13] Michael, L. (2013), An application of item response theory to fMRI data: Prospects and pitfalls, *Psychiatry Research: Neuroimaging*, **n° du volume**212(n° 3), 167–174.

[14] À propos de la théorie des réponses aux items. Le cas d'items dichotomiques, http://www.irdp.ch/edumetrie/tri.htm, consulté le 28-04-2015.

[15] Rowan, B et al. (2001), Measuring Teachers' Pedagogical Content Knowledge in Surveys: An Exploratory Study, Philadelphia, *PA: Consortium for Policy Research in Education, Ministry of Education, Youth and Sports*.

[16] Rizopoulo, D. (2006), ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses, *Journal of Statistical Software*, **n° du volume** 17 (n°5), 1-25.

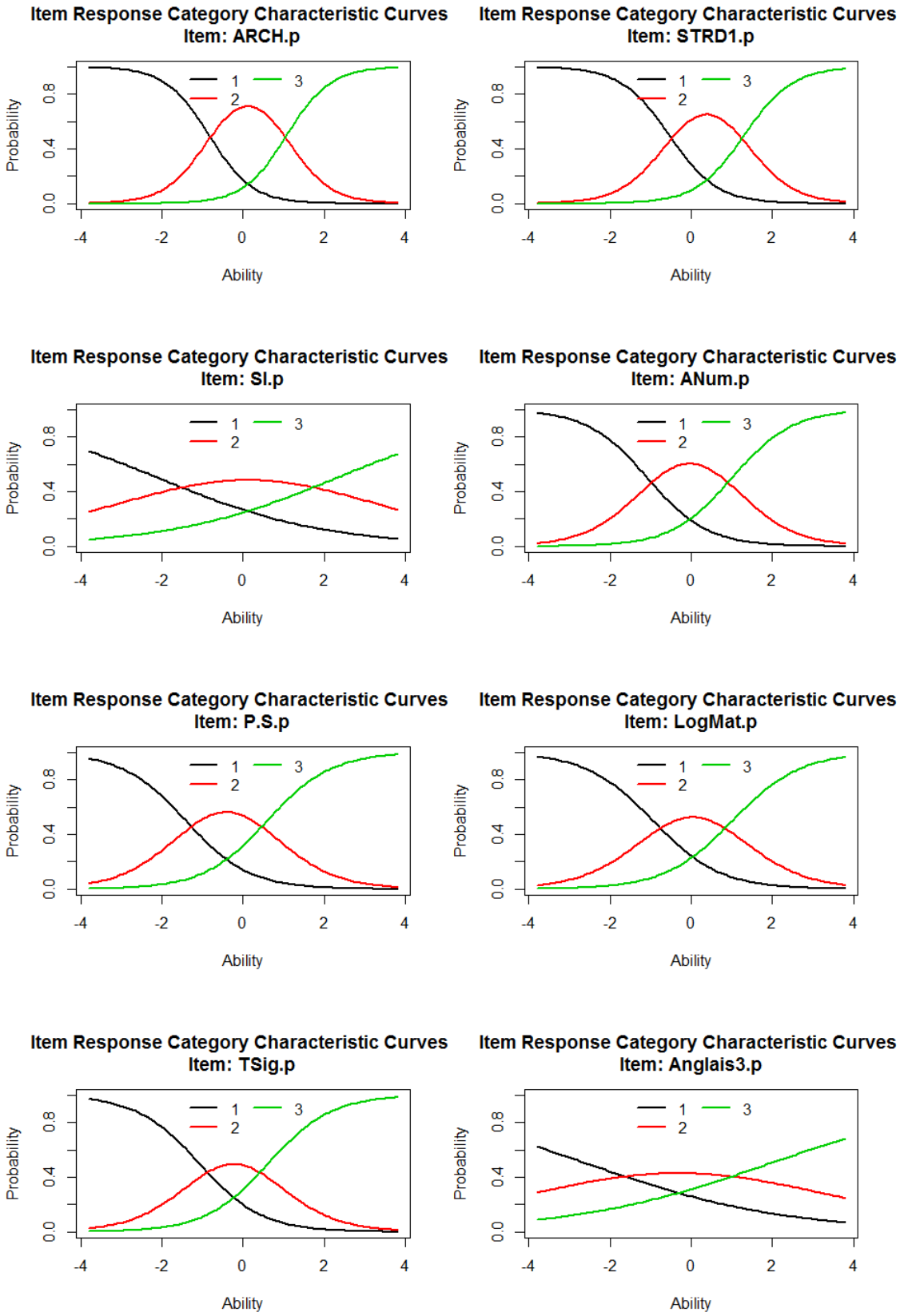
[17] Xinming, A et Yiu-Fai, Y. (2014), Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It, *SAS Institute Inc*.

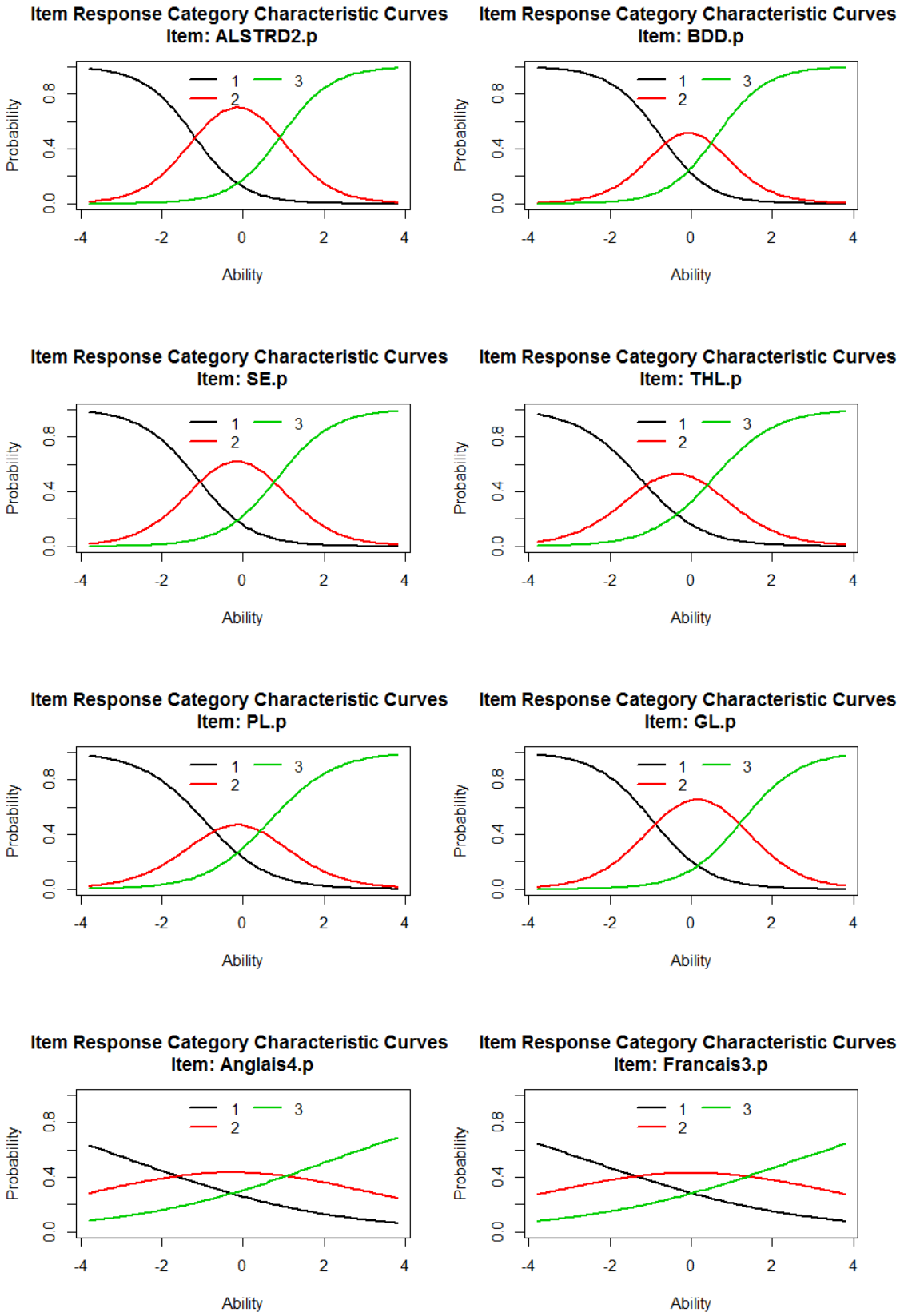
Annexe

Dans cette annexe nous présentons les courbes caractéristiques d'items de tous les modules étudiés en licence 2 des trois promotions 2010_2011, 2011_2012 et 2012_2013 :

Dans les figures 6, 7 et 8 on voit les courbes caractéristiques d'items (CCI) de tous les modules assurés en licence 2 pour les trois promotions 2010_2011, 2011_2012 et 2012_2013. Chaque CCI contient trois courbes, la noir (numéro 1) représente les étudiants qui sont faibles dans le module, la rouge (numéro 2) représente ceux qui sont moyens et la verte (numéro 3) représente ceux qui sont faibles dans le module.

Nous remarquons sur les trois figures que les courbes caractéristiques d'items sont presque les mêmes entre les trois promotions. Nous constatons aussi que les modules sont divisés en deux groupes; ceux qui sont discriminants (ARCH, STRD1, ANum, P.S, LogMat, TSig, ALSTRD2, BDD, SE, THL, PL et GL); ils représentent les modules de spécialité et ceux qui ne le sont pas (Anglais4, Français3, SI, Anglais3 et Français4).


Item Response Category Characteristic Curves
Item: ARCH.p
Item Response Category Characteristic Curves
Item: STRD1.p
Item Response Category Characteristic Curves
Item: SI.p
Item Response Category Characteristic Curves
Item: ANum.p
Item Response Category Characteristic Curves
Item: P.S.p
Item Response Category Characteristic Curves
Item: LogMat.p
Item Response Category Characteristic Curves
Item: TSig.p
Item Response Category Characteristic Curves
Item: Anglais3.p
Probability
Ability
1
2
3
0.0
0.4
0.8
-4
-2
0
2
4


Item Response Category Characteristic Curves
Item: ALSTRD2.p
Item Response Category Characteristic Curves
Item: BDD.p
Item Response Category Characteristic Curves
Item: SE.p
Item Response Category Characteristic Curves
Item: THL.p
Item Response Category Characteristic Curves
Item: PL.p
Item Response Category Characteristic Curves
Item: GL.p
Item Response Category Characteristic Curves
Item: Anglais4.p
Item Response Category Characteristic Curves
Item: Francais3.p
Probability
Ability
1
2
3
0.0
0.4
0.8
-4
-2
0
2
4

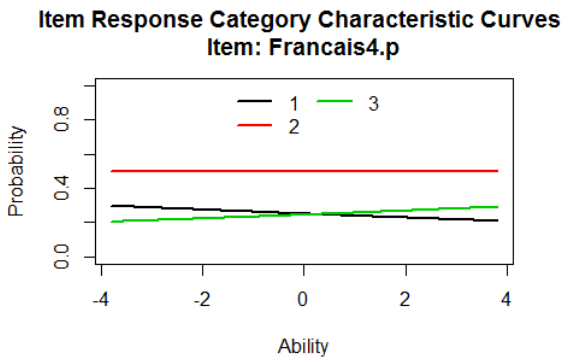
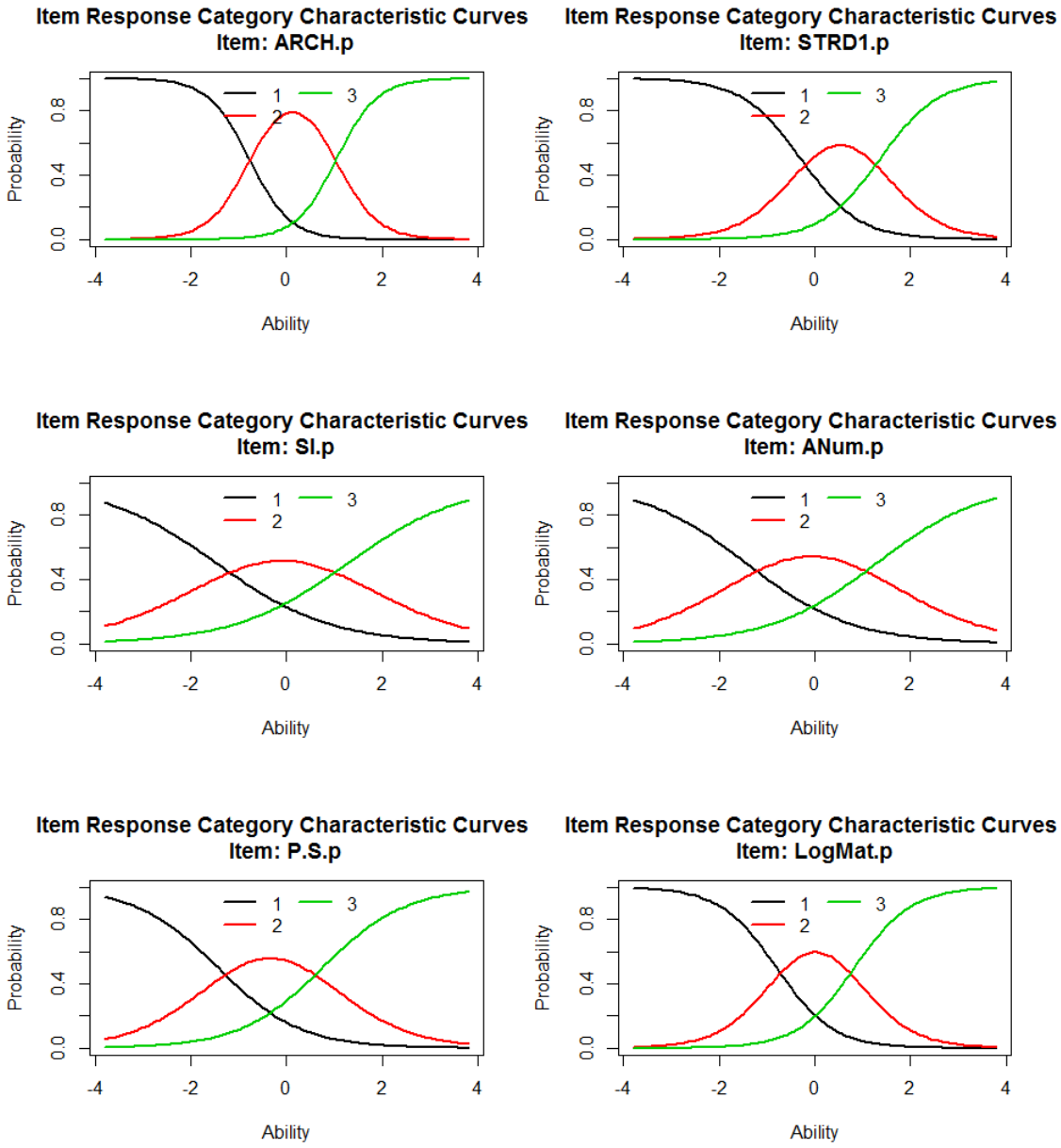
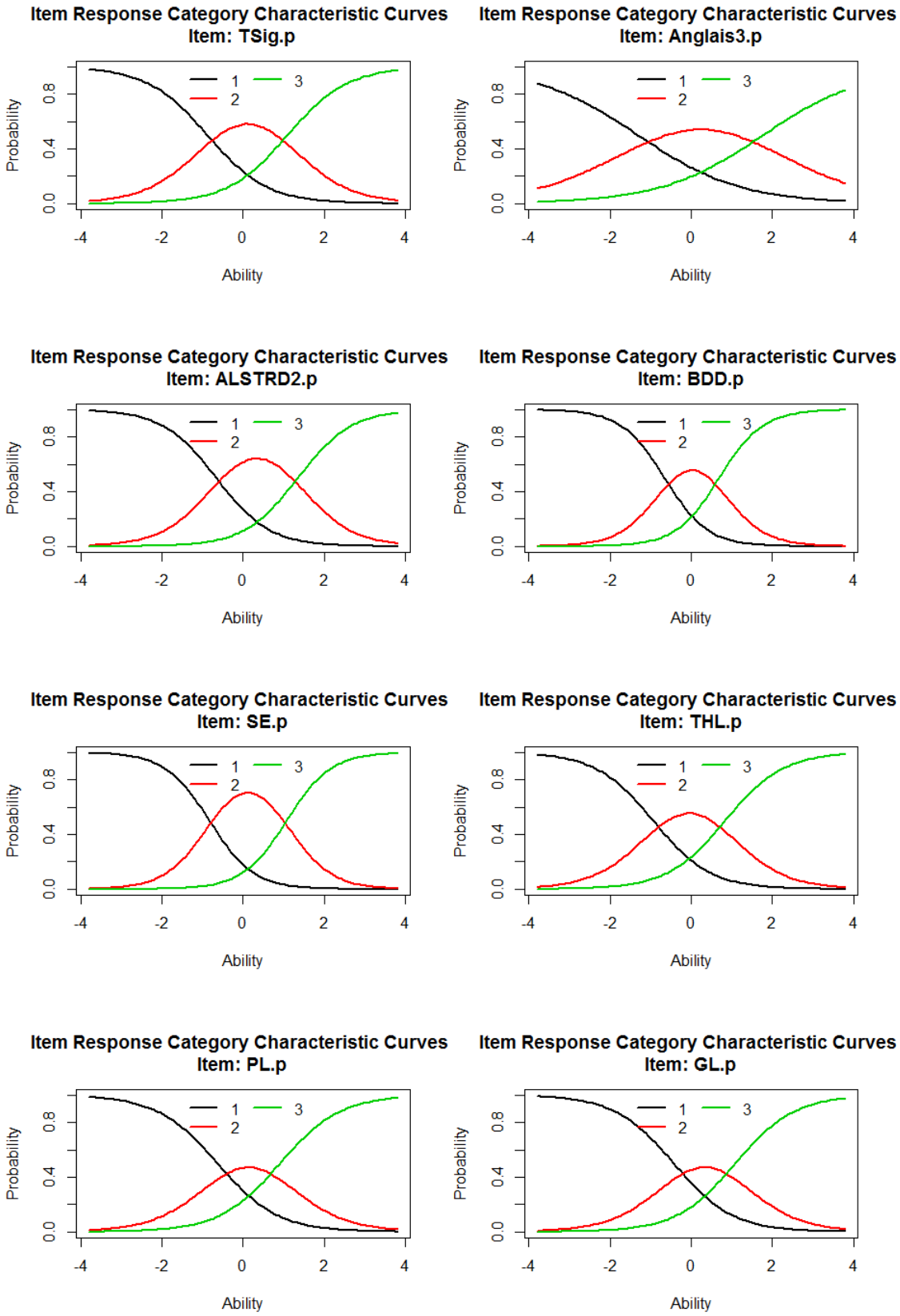


Figure 6 – CCI de licence 2 2010-2011




Item Response Category Characteristic Curves
Item: TSig.p
Item Response Category Characteristic Curves
Item: Anglais3.p
Item Response Category Characteristic Curves
Item: ALSTRD2.p
Item Response Category Characteristic Curves
Item: BDD.p
Item Response Category Characteristic Curves
Item: SE.p
Item Response Category Characteristic Curves
Item: THL.p
Item Response Category Characteristic Curves
Item: PL.p
Item Response Category Characteristic Curves
Item: GL.p
Probability
Ability
1
2
3
0.0
0.4
0.8
-4
-2
0
2
4

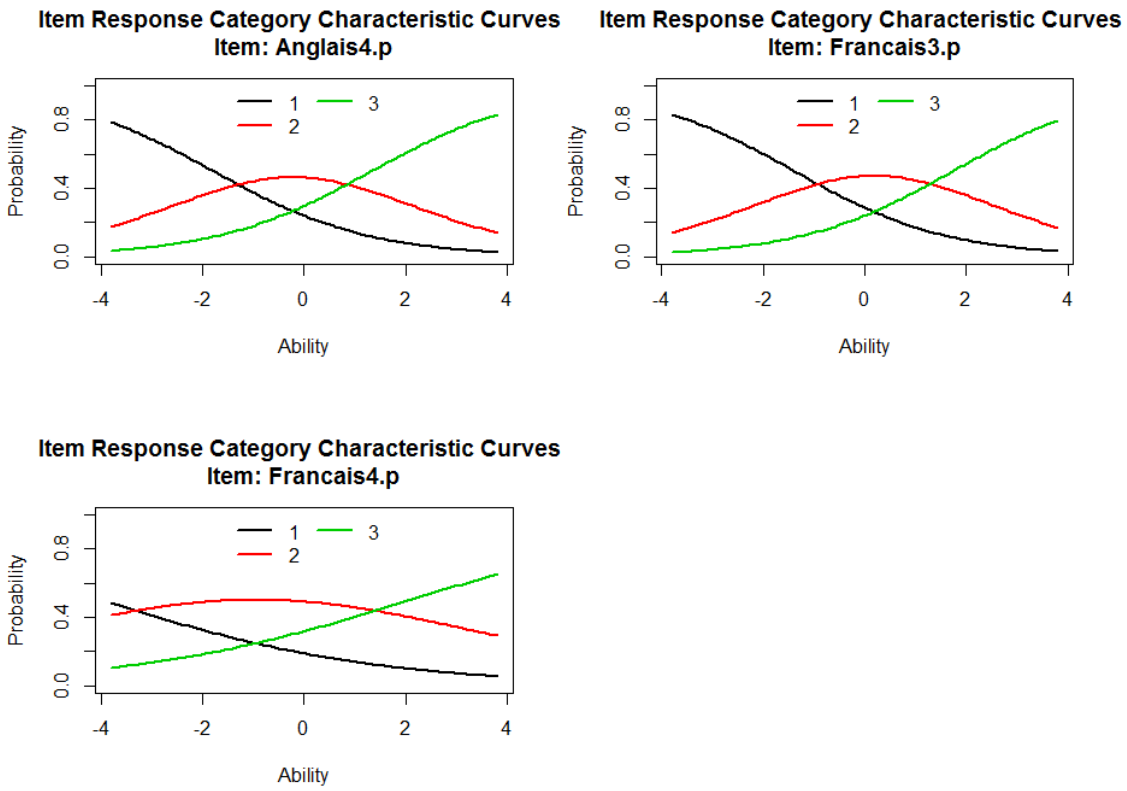
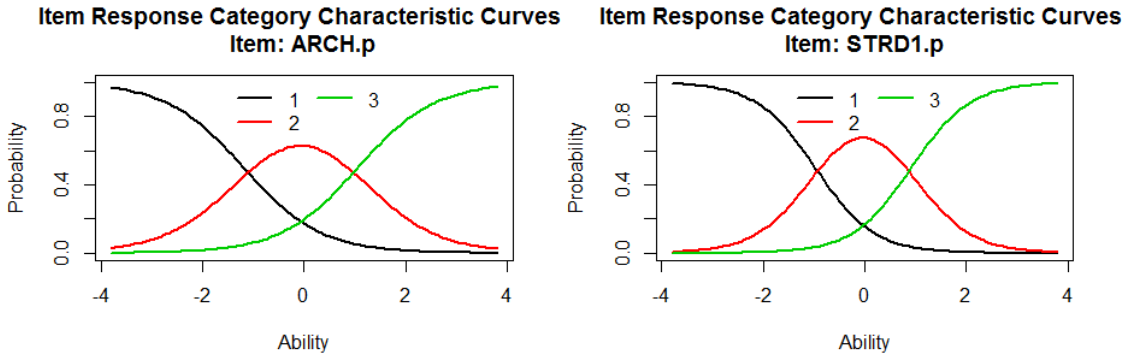
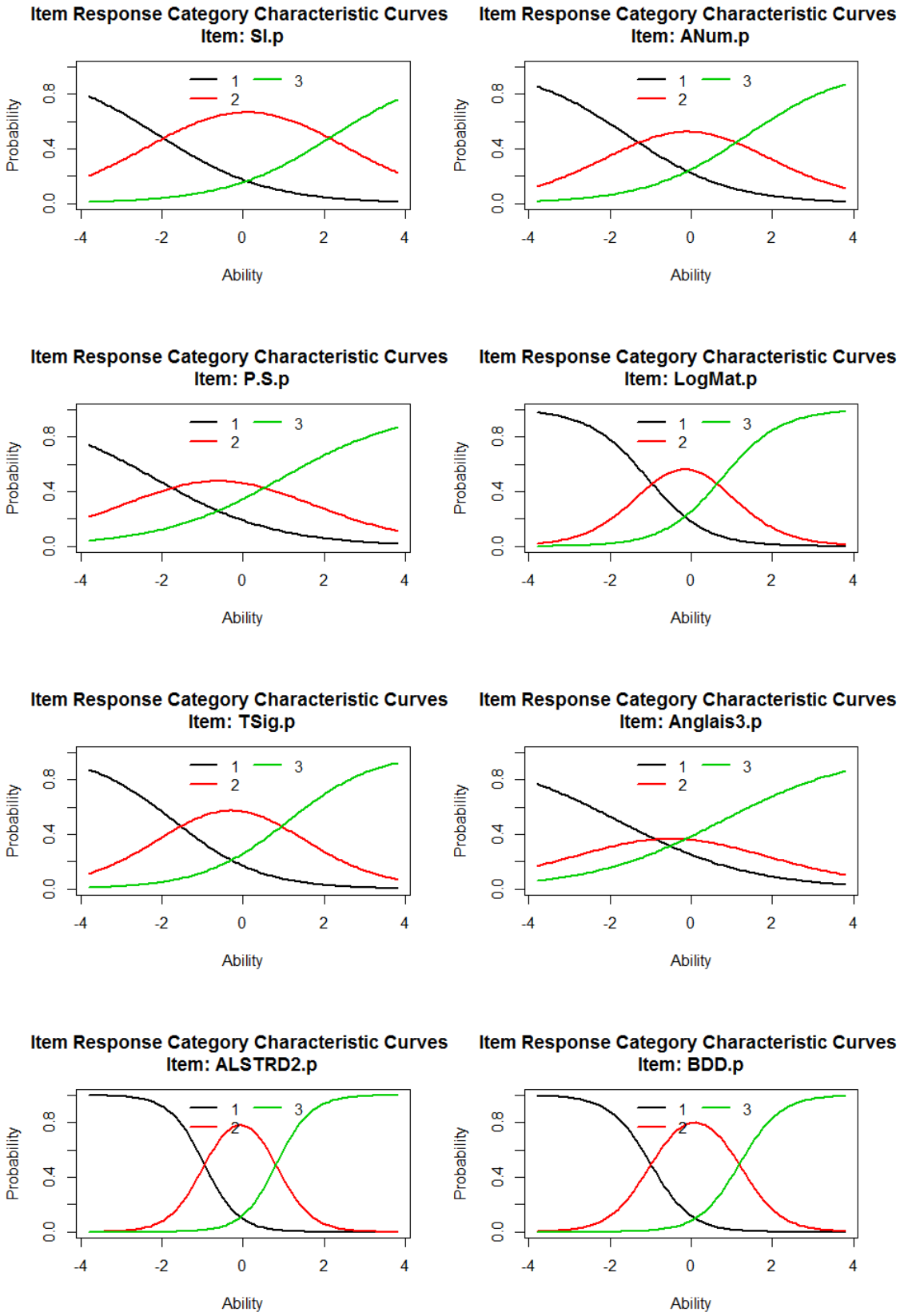


Figure 7 – CCI de licence 2 2011-2012




Item Response Category Characteristic Curves
Item: SI.p
Item Response Category Characteristic Curves
Item: ANum.p
Item Response Category Characteristic Curves
Item: P.S.p
Item Response Category Characteristic Curves
Item: LogMat.p
Item Response Category Characteristic Curves
Item: TSig.p
Item Response Category Characteristic Curves
Item: Anglais3.p
Item Response Category Characteristic Curves
Item: ALSTRD2.p
Item Response Category Characteristic Curves
Item: BDD.p
Probability
Ability
1
2
3
0.0
0.4
0.8
-4
-2
0
2
4

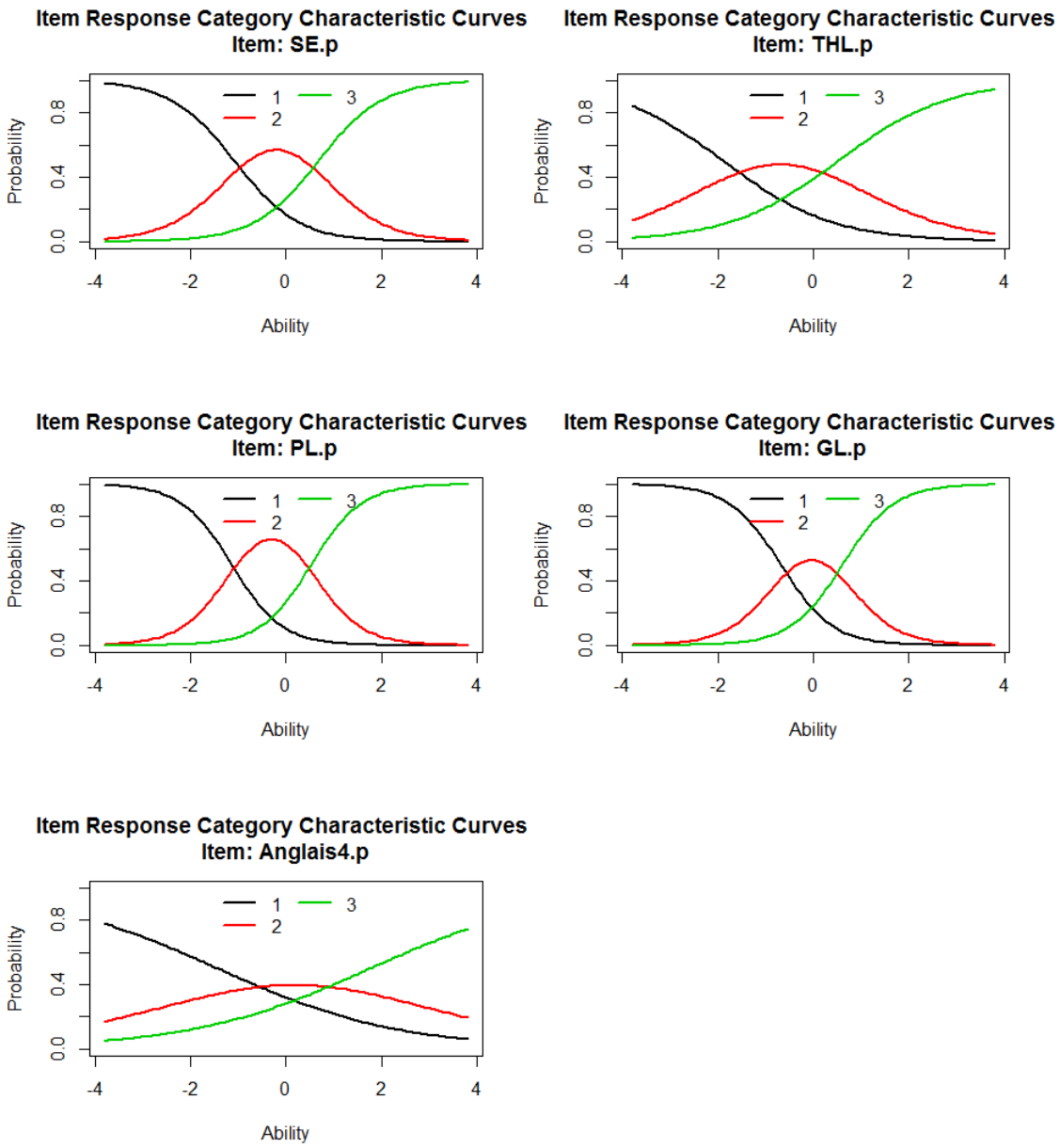


Figure 8 – CCI de licence 2 2012-2013

COORDINACIÓN DE LOS PROCESOS DE APROXIMACIÓN EN LA COMPRENSIÓN DEL LÍMITE DE UNA FUNCIÓN EN UN PUNTO. UNA APROXIMACIÓN A TRAVÉS DEL ANÁLISIS ESTADÍSTICO IMPLICATIVO

Joan PONS[1], Julia VALLS[2], Salvador LLINARES[3]

COORDINATION DES PROCESSUS D'APPROXIMATION DANS LA COMPRÉHENSION DE LA LIMITE D'UNE FONCTION EN UN POINT. UNE APPROCHE AU MOYEN DE L'ANALYSE STATISTIQUE IMPLICATIVE

COORDINATION OF THEAPPROXIMATION PROCESSES IN THE UNDERSTANDING OF THE LIMIT OF A FUNCTION AT A POINT. AN APPROCH THROUGE THE STATISTICAL IMPLICATIVE ANALISYS

RESUMEN

Esta investigación estudia la influencia de la comprensión de la coordinación de las aproximaciones en el dominio y en el rango en la comprensión de la concepción dinámica del límite en estudiantes de Bachillerato. El análisis se realizó usando el análisis implicativo (Gras et al., 2008). Los resultados indican que la comprensión de la concepción dinámica del límite se realiza mediante la coordinación de los procesos de aproximación en el dominio y en el rango, distinguiendo las aproximaciones laterales coincidentes de las no coincidentes en el rango. Además, nuestros resultados también indican que la coordinación de las aproximaciones se inicia en modo gráfico cuando las aproximaciones laterales coinciden y que los modos algebraico-numérico y numérico desempeñan un papel relevante en el desarrollo de la comprensión de la concepción dinámica de límite.

Palabras clave: *Comprensión de límite, coordinación de procesos de aproximación, modos de representación, análisis implicativo, estudiantes de bachillerato*

RÉSUMÉ

Cette recherche étudie l'influence de la compréhension de la coordination des approximations dans le domaine et dans le rang dans la construction de la conception dynamique de limite pour des élèves de la première et de la terminale. L'analyse a été effectuée en utilisant l'analyse implicative (Gras et al., 2008). Les résultats indiquent que la construction progressive de la conception dynamique de limite se fait par des processus d'approximation différenciées de la coordination des processus d'approximation dans le domaine et dans le rang et, en distinguant ceux des approximations latérales qui coïncident avec ceux qui ne coïncident pas. En outre, nos résultats indiquent également que la coordination des approximations commence en mode graphique lorsque les approximations latérales coïncident et que les modes algébriques-numériques et numériques jouent un rôle important dans le développement d'une compréhension de la notion dynamique de limite.

Mots-clés**:** *Compréhension des limites, coordination du processus d'approche, modes de représentation, analyse implicative, élèves de la première et de la terminale.*

[1] I.E.S Mutxamel, C/Mondúber, sn Mutxamel (Alicante), jpt11@alu.ua.es
[2] Universidad de Alicante, Sant Vicent del RAspeig (Alicante), julia.valls@ua.es
[3] Universidad de Alicante, Sant Vicent del RAspeig (Alicante), sllinares@ua.es

ABSTRACT

This research studies the influence of the understanding of the coordination of the approximations in the domain and range in the construction of the dynamic conception of the limit when teaching it to high school students. The analysis was carried out using the implicative analysis (Gras et al., 2008). The findings indicate that the gradual construction about the dynamic conception is made by processes that differentiate the coordination of approximation processes from those, in which the lateral approximation match of mismatched. In addition, our results indicate that the coordination of the approximations begins in graphic mode when the approximations coincide. To end, the algebraic-numeric and numeric modes play an important role in the development of the understanding of the dynamic conception of limit.

Keywords: *Understanding of limit concept, coordination of approximation processes, representation modes, implicative analysis, high school students.*

1 Introducción

La comprensión de la noción límite de una función real de variable real es un aspecto fundamental del análisis. La importancia del concepto de límite radica en que es una idea básica del análisis matemático por estar vinculada a conceptos como continuidad, derivada e integral. Las diferentes investigaciones indican que para los estudiantes el concepto de límite es una noción difícil (Sierpinska, 1985) que presenta dificultades cognitivas (Cornu, 1991). La influencia que tienen las diferentes representaciones en la comprensión del concepto de límite (Blázquez y Ortega, 2000; Lacasta y Wilhelmi, 2010; Monaghan, 2001) constituye, junto con el papel que desempeñan los obstáculos epistemológicos en el acceso al concepto de límite (Cornu, 1981; Garbin y Azcárate, 200; Moru, 2007; Sierpinska, 1985) y el papel de las concepciones espontáneas en las dificultades que presenta el propio concepto (Elia et al., 2009; Fernández-Plaza et al., 2013; Monaghan, 1991; Oehrtman, 2009; Szydlik, 2000; Williams, 1991), uno de los ámbitos en los que se han centrado las investigaciones sobre la comprensión que tienen los estudiantes del concepto de límite de una función

Una dificultad añadida a todo concepto matemático es la distinción entre la imagen del concepto y la definición del concepto (Tall y Vinner, 1981). Para estos autores la definición del concepto significa su definición formal tal como la entiende la comunidad matemática; la imagen del concepto significa toda estructura cognitiva asociada con el concepto e incluye todas las representaciones mentales con los procesos y propiedades asociadas. Pero debemos tener en cuenta que toda representación es cognitivamente parcial en relación a lo que representa, y, además, representaciones de registros diferentes no presentan los mismos aspectos de un mismo contenido conceptual (Duval, 1995). La búsqueda de soluciones, mediante enfoques didácticos, con la finalidad de superar las dificultades de los estudiantes en la comprensión del concepto de límite ha sido ampliamente investigada (Blázquez y Ortega, 2000; Engler et al., 2007; Kidron, 2010; Swinyard, 2011; Mamona-Downs, 2001; Mira et al., 2013), si bien no resulta fácil encontrar soluciones.

1.1 El límite de una función real de variable real como objeto de aprendizaje

Cauchy (1821) definió el limite como:"*cuando los valores atribuidos sucesivamente a la misma variable se aproximan indefinidamente a un valor fijo, de manera que terminen difiriendo de él tan poco como se quiera, este último valor se llaman el límite de todos los demás*" (p.4). Esta definición, al ser de carácter aritmético, hace de ella un concepto dinámico. Con posterioridad, Weierstrass, dio una definición métrica (1):

$$\lim_{x \to a} f(x) = L \Leftrightarrow \forall \varepsilon > 0, \exists \partial > 0 : \forall x, \quad 0 < |x - a| < \delta \Rightarrow |f(x) - L| < \varepsilon \qquad (1)$$

que es la que actualmente se utiliza en los primeros cursos universitarios. En contraposición a la concepción dinámica, la definición métrica de límite representa una concepción estática.

Desde una perspectiva didáctica, Blázquez y Ortega (2002) consideran la concepción dinámica como una aproximación óptima de la siguiente forma: *"Sea "f" una función y "a" un número real, el número "L" es el límite de la función "f" en el punto "a",si cuando "x" se acerca al número "a" más que cualquier aproximación, sus imágenes "f(x)" se acercan a "L" más que cualquier otra aproximación fijada"*

Desde una perspectiva cognitiva, Cottrill et al. (1996) afirman que la concepción dinámica del límite debe visualizarse y, para ello, el estudiante debe construir "*un proceso en el dominio y un proceso en el rango y usar la función para coordinarlos*" (p.174). Sin embargo, también sugieren que la concepción dinámica del límite, es más complicada de lo que se había pensado, ya que en la comprensión de la coordinación de los procesos de aproximación desempeñan un papel determinante los diferentes modos de representación.

La controversia entre la concepción dinámica versus concepción métrica es ampliamente analizada (Cottrill et al., 1996; Williams, 1991). Sin embargo, investigaciones recientes han mostrado la necesidad de inducir en los estudiantes imágenes dinámicas que sean compatibles con la definición formal de límite (Roh, 2008), y han puesto de manifiesto mediante experimentos de enseñanza que la forma en la que los estudiantes pueden desarrollar una rica y sólida comprensión de la definición de límite se apoya en la habilidad para usar de forma flexible las dos perspectivas (Swinyard, 2011).

1.2 La influencia de las diferentes representaciones en la comprensión del concepto de límite de una función en un punto

Los resultados de las investigaciones realizados sobre la influencia de las diferentes representaciones en la comprensión del concepto de límite de una función indican la necesidad de introducir la noción de límite con distintas representaciones (Blázquez, 1999). La elección de los distintos modos de representación depende de los aspectos del límite que se quieran subrayar (Engler et al., 2007). Los resultados de la investigación realizada por Engler et al. (2007) sobre los diferentes papeles que juegan los modos de representación indican que a los estudiantes les resulta más sencillo identificar la tendencia de una función en forma numérica o gráfica que en la algebraica y que cometían más errores para encontrar a qué valor se aproxima la función que para determinar a qué valores se aproxima la variable independiente. Las dificultades que

para los estudiantes representa el modo de representación algebraico, al resolver problemas de límites, se producen cuando realizan manipulaciones algebraicas (Moru, 2009), porque estas manipulaciones pueden ser la llave o la cerradura (Bergsten, 2006), puesto que las manipulaciones algebraicas pueden iniciar o bloquear el concepto. Sin embargo, también hay investigaciones que indican que la representación gráfica es más sólida que la numérica en el sentido de que los estudiantes resuelven mejor las situaciones gráficas que las numéricas (Monaghan, 2001).

Las investigaciones centradas en buscar diferentes alternativas didácticas a la definición formal también centran su atención en el papel de los modos de representación. Una de ellas utiliza el modo de representación gráfico asociándolo a la definición formal (Mamona-Downs, 2001); otra utiliza el modo de representación numérico y presenta la conceptualización del límite como aproximación óptima (Blázquez y Ortega, 2000) al considerar, que la utilización de la conceptualización de límite como aproximación óptima es más sencilla que la conceptualización formal (Blázquez et al., 2006). Sin embargo, los estudiantes que utilizan la definición formal y los que utilizan la noción de límite de un modo informal cuando trabajan con "números grandes" y con "números pequeños" cometen los mismos errores (Parameswaran, 2007).

Las investigaciones anteriores nos han proporcionado cierta información sobre las características de la comprensión del concepto de límite en los estudiantes, y en particular el papel desempeñado por las traslaciones entre los diferentes modos de representación, pero no aportan información sobre el papel de la coordinación de los procesos de aproximación en el dominio y en rango en la comprensión de la concepción dinámica del límite. Por ello nos hemos planteado las siguientes preguntas:

¿Cuál es el papel que desempeña la comprensión de la coordinación de los procesos de aproximación en el dominio y en el rango en la construcción de la concepción dinámica del límite?

¿Cómo influyen los distintos modos de representación en el acceso a la concepción dinámica de límite?

2 Método

2.1 Participantes

En la investigación han participado 129 estudiantes de bachillerato de edades comprendidas entre 16-18 años (66 de 1º y 63 de 2º curso de Bachillerato de Ciencias de la Naturaleza y de la Salud). Estos estudiantes habían sido iniciados a "*una aproximación al concepto de límite*" trabajando con: el concepto de función, la idea intuitiva de límite, los límites laterales, el cálculo de límites, límites en el infinito, límites infinitos, el cálculo de asíntotas, la noción de continuidad, sucesiones y límites de sucesiones. A los estudiantes de 1º de Bachillerato se les había introducido la noción de límite de una función en un punto dos semanas antes de contestar el cuestionario. A los de 2º de Bachillerato les habían introducido la misma noción seis meses antes de que

realizaran el cuestionario. Los estudiantes no tenían ninguna característica especial y su participación fue voluntaria.

2.2 Instrumentos y procedimiento de recogida de datos

Los datos provienen de un cuestionario y de las entrevistas realizadas a 21 estudiantes. Para la elaboración del cuestionario tuvimos en cuenta los resultados de las investigaciones sobre el concepto de límite de una función, analizamos libros de texto y materiales curriculares. Para el diseño de las diferentes cuestiones caracterizamos las nociones de aproximación dinámica y métrica de límite de una función con el objetivo de identificar los elementos matemáticos que la constituyen (Figura 1).

- *Sea f una función y x_0 un número real. El valor de la función f en $x=x_0$, $f(x_0)$ si existe o no, (E0)*
- *Idea de aproximación*
  - *x se aproxima al número a (E1)*
  - *f(x) se aproxima a L (E1)*
- *Coordinación en la concepción dinámica: cuando x se aproxima al número a, sus imágenes f(x) se aproximan a L (E2)*
- *Coordinación en la concepción métrica: se puede encontrar para cada ocasión un x suficientemente cerca de a tal que el valor de f(x) sea tan próximo a L como se desee (E4)*
- *Formalización como una manifestación de ser consciente de la existencia del límite L de la función f(x) en el punto a, escribiendo lím f(x) = L (E3)*

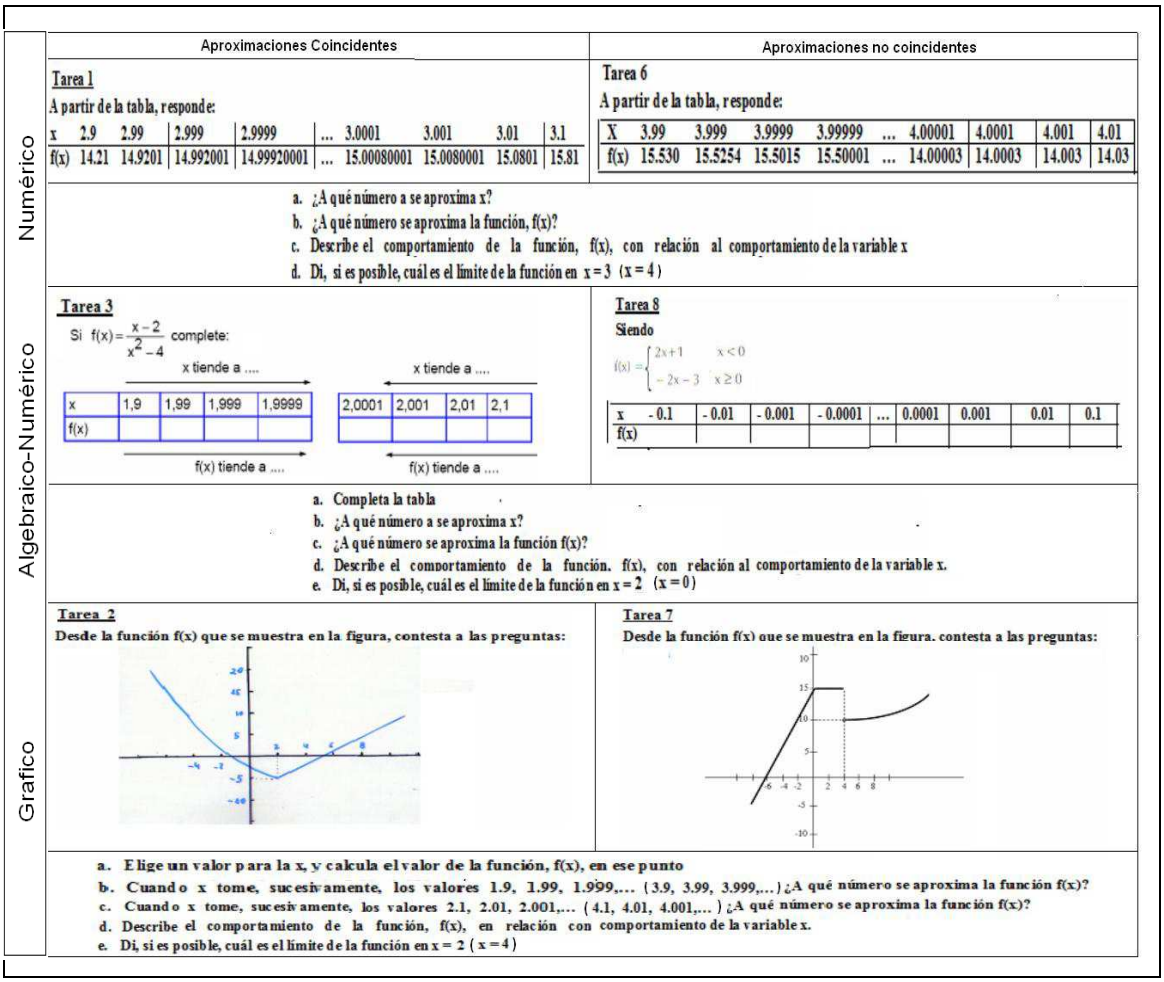
Figura 1– Elementos matemáticos considerados en el esquema límite de una función

Los ítems del cuestionario tenían como finalidad obtener información sobre la influencia que la coincidencia o no de las aproximaciones laterales, en el rango, tiene en la comprensión que los estudiantes poseen de la coordinación de las aproximaciones y el papel que desempeñan los diferentes modos de representación en esta coordinación. Se diseñaron 10 tareas con un total de 34 ítems. En esta investigación solo vamos a estudiar los datos correspondientes a las tareas 1, 2, 3, 6, 7, y 8 (Figura 2) que hacen referencia a la concepción dinámica (E0, E1 y E2) en tres modos de representación: numérico (N), gráfico (G), y algebraico-numérico (AN) y considerando la coincidencia o no de las aproximaciones laterales en el rango.

Las preguntas son:

1. "¿A qué número se aproxima la x y la f(x)?" (Elemento E1);
2. "Describe el comportamiento de la función, f(x), con relación al comportamiento de la variable x" (Elemento E2);
3. "Di, si es posible, cuál es el límite de la función en x =a" (Elemento E3).

Las aproximaciones en el dominio (la x) siempre se plantean en modo numérico y siempre son coincidentes. Las aproximaciones en el rango (la f(x)) se presentan en modos numérico y algebraico-numérico y pueden ser coincidentes o no coincidentes. En las tareas presentadas en el modo gráfico, 2 y 7, solo se pregunta a qué se aproxima la función cuando la x toma determinados valores. La tarea 3 proviene de Engler et al. (2007). Las otras las hemos elaborado nosotros.



Numérico

Aproximaciones Coincidentes

Tarea 1

A partir de la tabla, responde:

x	2.9	2.99	2.999	2.9999	...	3.0001	3.001	3.01	3.1
f(x)	14.21	14.9201	14.992001	14.99920001	...	15.00080001	15.0080001	15.0801	15.81

Aproximaciones no coincidentes

Tarea 6

A partir de la tabla, responde:

X	3.99	3.999	3.9999	3.99999	...	4.00001	4.0001	4.001	4.01
f(x)	15.530	15.5254	15.5015	15.50001	...	14.00003	14.0003	14.003	14.03

a. ¿A qué número a se aproxima x?
b. ¿A qué número se aproxima la función, f(x)?
c. Describe el comportamiento de la función, f(x), con relación al comportamiento de la variable x
d. Di, si es posible, cuál es el límite de la función en x = 3 (x = 4)

Algebraico-Numérico

Tarea 3

Si $f(x)=\frac{x-2}{x^2-4}$ complete:

x tiende a ....

x	1,9	1,99	1,999	1,9999
f(x)				

f(x) tiende a ....

x tiende a ....

2,0001	2,001	2,01	2,1

f(x) tiende a ....

Tarea 8

Siendo

$f(x)=\begin{cases} 2x+1 & x<0 \\ -2x-3 & x\geq 0 \end{cases}$

x	- 0.1	- 0.01	- 0.001	- 0.0001	...	0.0001	0.001	0.01	0.1
f(x)									

a. Completa la tabla
b. ¿A qué número a se aproxima x?
c. ¿A qué número se aproxima la función f(x)?
d. Describe el comportamiento de la función, f(x), con relación al comportamiento de la variable x.
e. Di, si es posible, cuál es el límite de la función en x = 2 (x = 0)

Grafico

Tarea 2

Desde la función f(x) que se muestra en la figura, contesta a las preguntas:

Tarea 7

Desde la función f(x) que se muestra en la figura, contesta a las preguntas:

a. Elige un valor para la x, y calcula el valor de la función, f(x), en ese punto
b. Cuando x tome, sucesivamente, los valores 1.9, 1.99, 1.999,... (3.9, 3.99, 3.999,...) ¿A qué número se aproxima la función f(x)?
c. Cuando x tome, sucesivamente, los valores 2.1, 2.01, 2.001,... (4.1, 4.01, 4.001,...) ¿A qué número se aproxima la función f(x)?
d. Describe el comportamiento de la función, f(x), en relación con comportamiento de la variable x.
e. Di, si es posible, cuál es el límite de la función en x = 2 (x = 4)

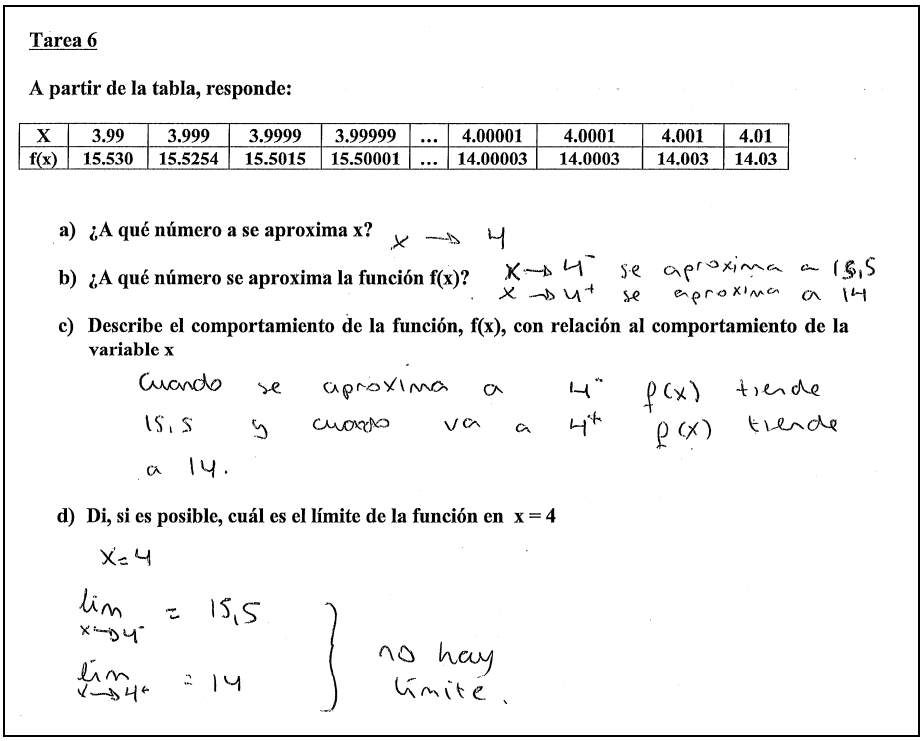
Figura 2 – Tareas que hacen referencia a los elementos de la concepción dinámica

2.3 Análisis

Cada ítem fue codificado indicando el elemento matemático que se pone de manifiesto en el ítem (Ei, i = 0, 1, 2, 3), el modo de representación usado en la presentación del problema (G, N, AN), la tarea a la que corresponde y su posición en la misma, a través de un número y una letra en minúscula. Por ejemplo, E1N6a, indica el ítem a de la tarea 6 presentado en modo Numérico y hace referencia al elemento matemático E1. Inicialmente, un grupo de tres investigadores realizaron una lectura conjunta de las respuestas a cada uno de los ítems de las tareas con el objetivo de generar criterios y unificar la puntuación dicotómica, (1 respuesta correcta, 0 incorrecta).

Ejemplificamos el análisis realizado a partir de las respuestas dadas a la tarea 6 por la estudiante EST36 (Figura 3). La respuesta a la tarea 6 fue puntuada como (1,1,1,1) dado que EST36 respondió de forma correcta ¿a qué número se aproxima la x y la f(x)? (ítems 6a y 6b) al asociar la aproximación a un número en el dominio a un número natural ($x \rightarrow 4$); la aproximación en el rango a diferentes números por la izquierda y por la derecha de x=4 (f(x) *se aproxima a 15,5 cuando* $x \rightarrow 4^-$ *y cuando* $x \rightarrow 4^+$ *se aproxima a 14*). También respondió correctamente al ítem 6c- Describe el comportamiento de la

función, f(x), con relación al comportamiento de de la variable x- al coordinar las aproximaciones en el dominio y el rango (*Cuando x se aproxima a 4^-, f(x) tiende a 15,5 y cuando se aproxima a 4^+, f(x) tiende a 14*). Finalmente, el ítem 1d se puntuó con un 1 al indicar que "*no hay límite*" después de diferenciar que los límites laterales de la función son distintos a la derecha y a la izquierda.



Tarea 6

A partir de la tabla, responde:

X	3.99	3.999	3.9999	3.99999	…	4.00001	4.0001	4.001	4.01
f(x)	15.530	15.5254	15.5015	15.50001	…	14.00003	14.0003	14.003	14.03

a) ¿A qué número a se aproxima x? $x \to 4$

b) ¿A qué número se aproxima la función f(x)? $x \to 4^-$ se aproxima a 15,5
$x \to 4^+$ se aproxima a 14

c) Describe el comportamiento de la función, f(x), con relación al comportamiento de la variable x

Cuando se aproxima a 4^- f(x) tiende 15,5 y cuando va a 4^+ f(x) tiende a 14.

d) Di, si es posible, cuál es el límite de la función en x = 4

$x = 4$

$\lim_{x \to 4^-} = 15{,}5$

$\lim_{x \to 4^+} = 14$

no hay límite.

Figura 3 – Respuesta de la EST36 a la Tarea 6

De esta manera, cada estudiante venía definido por una 28-tuplas (variables) relativas a la valoración de las respuestas a las tareas centradas en la concepción dinámica (1, 2, 3, 4, 6, 7, 8).

A continuación realizamos un análisis estadístico implicativo utilizando el software CHIC (Classification Hiérarchique Implicative et Cohésitive) (Gras et al., 2008). En la estadística implicativa, Trigueros y Escandon (2008) señalan que dada una población E, los estudiantes, y un conjunto de variables, los ítems del cuestionario, "*se busca dar sentido estadístico a una implicación no estricta $a \Rightarrow b$*" (pág. 66). En esta metodología, "*la implicación $a \Rightarrow b$ será admisible en una experiencia si el número de individuos de E que la contradicen es muy pequeño, en términos probabilísticos, en relación con el número de individuos esperado bajo la hipótesis de ausencia de relación. Si esto ocurre, se puede decir que A, conjunto de observaciones que satisfacen la característica a, está "casi" contenido en B, conjunto de observaciones que satisfacen la característica b"* (pág. 67).

El análisis implicativo genera un gráfico jerárquico y no transitivo, que permite identificar variables que se unen a través de implicaciones binarias ($a \Rightarrow b$, "a" entonces "b") o mediante implicaciones no binarias ($a \Rightarrow (b \Rightarrow c)$, si "a" y "b" entonces "c") haciendo emerger una primera estructura conceptual entre diferentes elementos matemáticos que conforman la concepción dinámica de límite. Este análisis traduce gráficamente el conjunto de las relaciones cuasi-implicativas entre las variables en distintos niveles de significación.

En nuestro estudio la población es la muestra de estudiantes de educación postobligatoria que respondieron al cuestionario y las variables son los ítems de las diferentes tareas considerando el elemento matemático y el modo de representación en que se presenta la tarea. Las relaciones implicativas entre las diferentes variables nos indican en qué medida el uso de un elemento matemático en una tarea (planteada en un modo de representación concreto y desde aproximaciones coincidentes o no en el rango) está relacionado con el uso de ese mismo u otro elemento matemático en otra tarea (en ese u otro modo de representación y coincidiendo o no las aproximaciones laterales en el rango) con una determinada probabilidad. La medida de esta relación nos permite inferir cómo se está construyendo el acceso al concepto de límite de una función en un punto.

En el análisis estadístico implicativo realizado en este trabajo (al 99% de significación), hemos usado una de las posibilidades del programa CHIC para *"suprimir o centrarse"* solamente en determinadas variables (Couturier, 2009; Gras y Kuntz, 2009) con la finalidad de resaltar determinadas relaciones entre elementos matemáticos. En los gráficos implicativos generados, los ítems que inician las ramas indican que los estudiantes que responden adecuadamente a ellos también lo hacen a la mayoría de los ítems que aparecen en la parte inferior del gráfico (Bodi et al., 2009; Zamora et al., 2009)

3 Resultados

Los resultados están organizados en tres secciones. En la primera, describimos la influencia de los modos de representación en la manera en la que los estudiantes usaban la idea de aproximación a un número (E1). En la segunda, describimos cómo la no coincidencia de las aproximaciones laterales en el rango, determina un punto de referencia en la comprensión dinámica de límite. Finalmente, en la tercera, describiremos los resultados relativos a la influencia de los modos de representación en la comprensión de la coordinación dinámica de límite (E2).

3.1 La influencia de los modos de representación en el uso de la idea de aproximación a un número

Para poner de manifiesto la influencia de los modos de representación en la comprensión de la aproximación a un número (E1) seleccionamos las variables que hacen referencia a este elemento en los modos de representación numérico y algebraico-numérico. La selección realizada dio lugar al gráfico implicativo que se muestra en la figura 4.

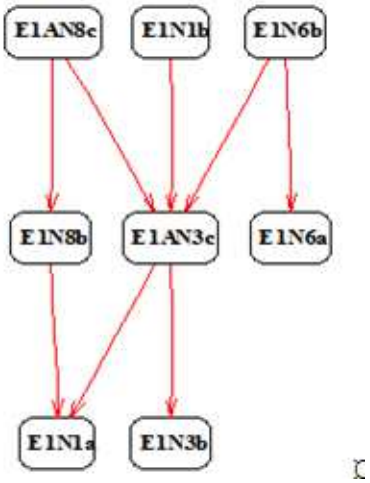


Figura 4 – Relaciones implicativas al 99% de las variables relativas al elemento E1

Desde este gráfico podemos inferir cuatro ideas relevantes relativas al papel de los modos de representación en relación a la comprensión de la idea de aproximación a un número.

La primera idea la inferimos a partir de las relaciones implicativas de la parte superior derecha de la figura 4:

- E1N1b $\Rightarrow$ E1AN3c ;
- E1N6b $\Rightarrow$ E1AN3c.

Estas dos relaciones implicativas sugieren que si los estudiantes comprenden la aproximación a un número en el rango, en modo numérico, tanto cuando las aproximaciones laterales son coincidentes (E1N1b) como cuando no son coincidentes (E1N6b), también comprenden las aproximaciones laterales coincidentes en modo algebraico-numérico (E1AN3c).

La segunda idea la inferimos a partir de las relaciones implicativas de la parte superior izquierda y central de la figura 4:

- E1AN3c $\Rightarrow$ E1N1a ;
- E1AN3c $\Rightarrow$ E1N3b ;
- E1AN8c $\Rightarrow$ E1N8b.

Estás tres relaciones implicativas indican que si los estudiantes comprenden la aproximación a un número en modo algebraico-numérico, tanto cuando las aproximaciones laterales son coincidentes (E1AN3c) como cuando no son coincidentes (E1AN8c), también comprenden las aproximaciones laterales coincidentes en el modo numérico (E1N1a, E1N3b, E1N8b).

Estas dos ideas sugieren que **los estudiantes utilizan indistintamente los modos de representación numérico y algebraico-numérico** para acceder a la idea de aproximación a un número.

La tercera idea la inferimos a partir de las relaciones implicativas que muestra la figura 5 (extraída de la figura 4) en las que las cuatro variables relacionadas hacen referencia a la idea de aproximación en el rango. En esta subestructura la idea de aproximación en modo algebraico-numérico cuando las aproximaciones laterales

coinciden (E1AN3c) desempeña un papel relevante al agrupar al resto de las variables que hacen referencia a la idea de aproximación en el rango.

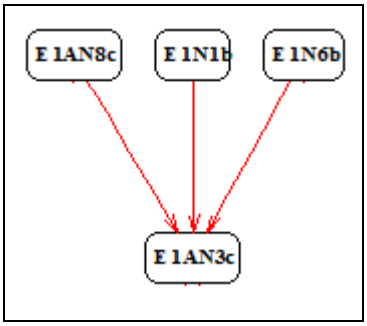


Figura5 – Relaciones al 99% entre las variables "idea de aproximación numérica", en el rango (E1)

Las relaciones implicativas:

- E1AN8c ⇒ E1AN3c ;
- E1N6b ⇒ E1AN3c ;

ponen de manifiesto que la comprensión de la aproximación a un número en el rango cuando las aproximaciones laterales no coinciden (E1AN8c y E1N6b) se apoya en la comprensión de la aproximación a un número en el rango cuando las aproximaciones laterales coinciden (E1AN3c). Es decir, **la comprensión de la aproximación a un número en el rango se inicia en modo algebraico-numérico con aproximaciones laterales coincidentes y se consolida cuando las aproximaciones laterales no coinciden (representadas en el modo numérico y numérico-algebraico).**

Globalmente consideradas, las implicaciones generadas parecen indicar que la comprensión de la aproximación en el rango (E1N1b, E1AN3c, E1N6b y E1AN8c) está vinculada a la comprensión previa de la aproximación en el dominio (E1N8b, E1N1a, E1N3b y E1N6a) pero la relación inversa no se establece. Es decir, la comprensión de la aproximación en el dominio no implica la comprensión de la aproximación en el rango. Este hecho parece establecer **una diferencia cognitiva entre la comprensión de los procesos de aproximación en el dominio y en el rango**.

3.2 La idea de aproximación a un número y la comprensión dinámica de límite. El papel de la no coincidencia de las aproximaciones laterales

Con la finalidad de analizar cómo la no coincidencia de las aproximaciones laterales determinan la comprensión dinámica del límite entendida como la coordinación de las aproximaciones en el dominio y en el rango" (E2), seleccionamos las variables que hacen referencia a la idea de aproximación y a la idea de coordinación de las aproximaciones Esta selección dio lugar al gráfico implicativo generado al 99% de significación que muestra la figura 6.

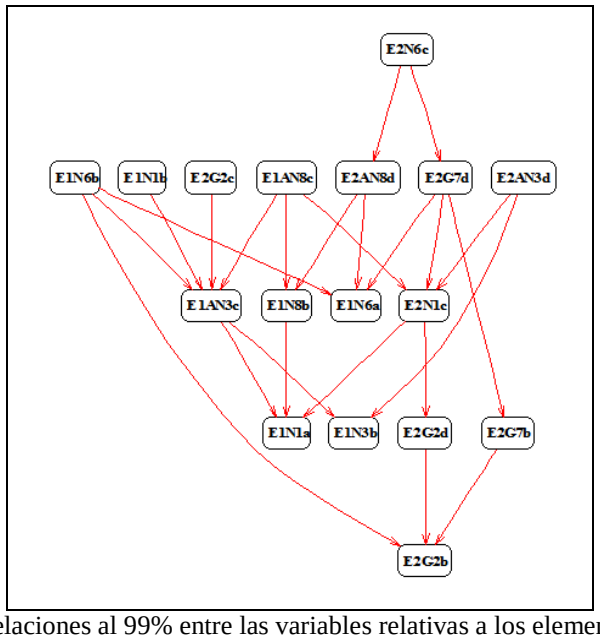


Figura 6 – Relaciones al 99% entre las variables relativas a los elementos E1 y E2

En este gráfico las relaciones en las que toma parte la idea de aproximación a un número en el rango cuando las aproximaciones laterales no coinciden permiten identificar dos ideas del papel que desempeña la idea de aproximación a un número en la comprensión dinámica de límite

- E1AN8c $\Rightarrow$ E2N1c ;
- E1N6b $\Rightarrow$ E2G2b.

En primer lugar, podemos inferir que si los estudiantes comprenden la aproximación a un número en el rango en modo algebraico-numérico cuando las aproximaciones laterales no coinciden (E1AN8c), serán capaces de identificar la coordinación de los procesos de aproximación en el dominio y en el rango en modo numérico cuando las aproximaciones laterales coinciden (E2N1c). En segundo lugar, si los estudiantes comprenden la aproximación a un número en el rango cuando las aproximaciones laterales no coinciden en modo numérico (E1N6b), coordinarán por la izquierda las aproximaciones coincidentes en el dominio y en el rango en modo gráfico (E2G2b).

Estos datos indican que comprender la aproximación no coincidente implica ser capaz de coordinar las aproximaciones coincidentes en el dominio y en el rango. Es decir, **la no coincidencia de las aproximaciones laterales a un número en el rango parece que determina la posibilidad de coordinar las aproximaciones entre el dominio y el rango que subyace a la concepción dinámica de limite**.

3.3 El papel de los modos de representación en la coordinación de las aproximaciones.

Para poner de manifiesto la influencia de los modos de representación en la comprensión de la coordinación de las aproximaciones (E2) seleccionamos las variables que hacen referencia a la coordinación en los distintos modos de representación, numérico, gráfico y algebraico-numérico. Del gráfico implicativo generado al 99% de significación que se muestra en la Figura 7, podemos extraer tres ideas relativas al papel

de los modos de representación comprensión de la coordinación de las aproximaciones en el dominio y en el rango.

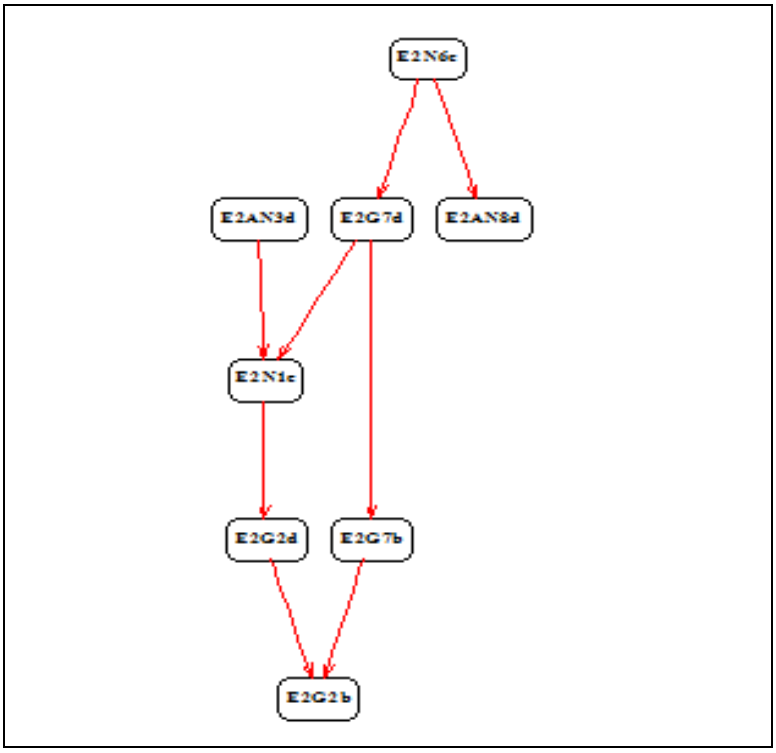


Figura 7 – Relaciones al 99% entre variables relativas al elemento E2

La primera idea la inferimos a partir de las relaciones implicativas (cuando las aproximaciones laterales coinciden) que se encuentran en la parte izquierda de la figura 7:

- E2AN3d $\Rightarrow$ E2N1c ;
- E2N1c $\Rightarrow$ E2G2d.

Estas dos relaciones implicativas indican que la coordinación de las aproximaciones cuando estas coinciden se inicia en modo gráfico (E2G2d), progresa en modo numérico (E2N1c) y se consolida en modo algebraico-numérico (E2AN3d). Es decir, **la comprensión de la coordinación de las aproximaciones coincidentes en el dominio y en el rango se inicia en modo gráfico, progresa en modo numérico y se consolida en modo algebraico-numérico**. Las relaciones implicativas entre estas variables indican que **si la comprensión que tienen los estudiantes de la coordinación de los procesos de aproximación en modo algebraico-numérico va acompañada de la comprensión de dicha coordinación en modo numérico, entonces también la comprenderán en modo gráfico**.

La segunda idea la inferimos a partir de las relaciones implicativas de la parte superior de la figura 7:

- E2N6c $\Rightarrow$ E2AN8d ;
- E2N6c $\Rightarrow$ E2G7d.

Estas dos relaciones implicativas indican que la comprensión de la coordinación de las aproximaciones no coincidentes en el rango se inicia indistintamente en modos gráfico (E2G7d) y algebraico-numérico (E2AN8d), pero se consolida en modo numérico (E2N6c).

La tercera idea la inferimos a partir de la relación implicativa que se encuentran en la parte central de la figura 7:

– E2G7d $\Rightarrow$ E2N1c.

Esta relación implicativa muestra que **el nexo de unión entre la coordinación de los procesos de aproximaciones no coincidentes en el rango y la coordinación de los procesos de aproximaciones coincidentes en el rango se realiza con los modos gráfico (G) y numérico (N).**

4 Conclusión y discusión

El objetivo de esta investigación ha sido aportar información sobre cómo los estudiantes de primero y segundo de bachillerato construyen el significado del concepto de límite de una función en un punto a partir de la coordinación de los procesos de aproximación (coincidentes o no en el rango) y qué papel desempeñan los modos de representación. Los resultados indican que la comprensión de la coincidencia o no coincidencia de las aproximaciones laterales en el rango es un elemento importante en el proceso de construcción del significado de límite de una función en un punto.

En un primer momento, la idea de aproximación en el rango se apoya en la comprensión de la idea de aproximación en el dominio indistintamente en los modos de representación numérico y algebraico-numérico. En este sentido, nuestros resultados apoyan los presentados por Engler et al. (2007) cuando indican que los estudiantes tienen más dificultades en encontrar a qué valor se aproxima la función que en determinar a qué valor se aproxima la variable independiente. Por otra parte, la idea de aproximación en el rango se inicia en modo algebraico-numérico cuando las aproximaciones laterales coinciden y se consolida en los modos numérico y algebraico-numérico cuando dichas aproximaciones no coinciden. Igualmente, la idea de aproximaciones no coincidentes a un número en el rango, en los modos numérico y algebraico-numérico, se apoya en la comprensión de la coordinación de las aproximaciones coincidentes en el dominio y en el rango. Es decir, la construcción paulatina de la concepción dinámica del límite se realiza mediante la comprensión de las aproximaciones, diferenciando las que se realizan en el dominio de las que se realizan en el rango, y dentro de estos últimos aquellos en los que las aproximaciones laterales coinciden de las que no coinciden. Sin embargo, un hecho importante puesto de manifiesto por nuestros resultados es que la comprensión de la idea de aproximaciones no coincidentes en el rango implica la comprensión de la coordinación de las aproximaciones coincidentes en el dominio y en el rango, pero no implica necesariamente el comprender la coordinación cuando las aproximaciones laterales no coinciden. Este hecho podría señalar la diferencia cognitiva que para el estudiante resulta tener el establecimiento de las aproximaciones a un número en el rango, en relación a que coincidan o no las aproximaciones laterales (Pons et al., 2011; Valls et al., 2011)

En segundo lugar, y en relación a la comprensión de la concepción dinámica de límite, nuestros resultados indican que la comprensión de la coordinación de las aproximaciones se inicia en modo gráfico cuando las aproximaciones laterales coinciden, progresa en modo numérico y se consolida en modo algebraico-numérico. Estos resultados apoyan los presentados por Monaghan (2001), Blázquez y Ortega (2001) y Engler et al. (2007) en el sentido de que los límites presentados en modo gráfico son resueltos por más estudiantes que los presentados en modo numérico. Sin embargo, un resultado relevante de nuestra investigación es que la coordinación de los procesos de aproximación, cuando las aproximaciones laterales no coinciden en el rango, se inicia indistintamente en modos gráfico o algebraico-numérico y se consolida en modo numérico. Además, del hecho de que la coordinación de los procesos de aproximaciones no coincidentes en modo gráfico se apoye en la coordinación de procesos de aproximaciones coincidentes en modo numérico, nos permite inferir que son los modos gráfico y numérico el nexo de unión entre la comprensión de las coordinaciones de las aproximaciones laterales no coincidentes y las coincidentes. Por otra parte, la comprensión de la coordinación de las aproximaciones coincidentes en el dominio y en el rango no implica necesariamente el ser capaz de coordinar los procesos de aproximaciones laterales no coinciden. La importancia de este hecho radica en que podría señalar la diferencia cognitiva que para el estudiante podría tener el establecimiento de la coordinación de las aproximaciones en el dominio y en el rango, en relación a que coincidan o no las aproximaciones laterales en el rango.

En relación a la comprensión del concepto de límite de una función nuestros resultados estarían en sintonía con los presentados por Blázquez (1999) y Blázquez y Ortega (2001) cuando afirman que la utilización de distintos registros de representación (algebraico, numérico, gráfico, verbal) mejora esta comprensión. Para estos autores un estudiante que utilice indistintamente diferentes modos de representación tendrá una sólida comprensión de la noción de límite de una función.

Las estructuras implicativas identificadas describen ciertas relaciones entre la comprensión que los estudiantes tienen de la coordinación de los procesos de aproximación en los diferentes modos de representación (numérico, gráfico y algebraico-numérico) y la coincidencia o no de las aproximaciones laterales. La información aportada puede ser útil para la organización del contenido de enseñanza y la planificación de secuencias de actividades dirigidas a desarrollar la comprensión en los estudiantes de la concepción dinámica del límite. Una implicación de este hecho es que al diseñar un proceso de enseñanza–aprendizaje de la noción dinámica de límite de una función debemos ser conscientes del papel que tienen los diferentes modos de representación para apoyar la comprensión de la coordinación de los procesos de aproximación, coincidan o no las aproximaciones laterales en el rango.

Referencias

[1] Bergsten, C. (2006). Trying to Reach the Limit. The Role of Algebra in Mathematical Reasoning. *Proceedings of the 30nd Conference of the International Group for the Psychology of Mathematics Education*, **2**(2), 153–160.

[2] Blázquez, S. (1999). *Noción de límite en Matemáticas Aplicadas a las Ciencias Sociales.* Tesis doctoral. Universidad de Valladolid. Valladolid. España.

[3] Blázquez, S. y Ortega, T. (2000). El concepto de límite en la Educación Secundaria. En *El futuro del cálculo infinitesimal.* México: Grupo Editorial Iberoamérica. S.A. de C.V.

[4] Blázquez, S. y Ortega, T. (2001). Los sistemas de representación en la enseñanza del límite. *Revista Latinoamericana de Investigación en Matemática Educativa,* **4**(3), 219- 236.

[5] Blázquez, S. y Ortega, T. (2002). Nueva definición de límite funcional. *UNO,* **30,** 67–83.

[6] Blázquez, S., Ortega, T., Gatica, S. y Benegas, J. (2006). Una conceptualización de límite para el aprendizaje inicial de análisis matemático en la universidad. *Revista Latinoamericana de Investigación en Matemática Educativa,* **9**(2), 189–209.

[7] Bodi, S.D., Valls, J. y Llinares, S. (2009). La comprensión de la divisibilidad en N. Un análisis implicativo. En P. Orús, L. Zamora, P. Gregori (ed), *Teoría y aplicaciones del Análisis Estadístico Implicativo* (215–233). Castellón, España: Innovació Digital Castelló.

[8] Cauchy, A. (1821). *Cours d'Analyse de l'École Royale Polytechnique, (Primer Partie. Analyse Algébrique).* Sevilla: Sociedad Andaluza de Educación Matemática "THALES". Edición facsímil.

[9] Couturier, R. (2009). CHIC: utilización y funcionalidades. En P. Orús, L. Zamora, P. Gregori (ed), *Teoría y aplicaciones del Análisis Estadístico Implicativo* (51–63). Castellón, España: Innovacio Digital Castelló.

[10] Cornu, B. (1991). Limits. En D. Tall (Ed), *Advanced Mathematical Thinking* (153-166). Dordrecht: Kluver.

[11] Cottrill, J., Dubinsky, E., Nichols, D., Schwingendorf, K., Thomas, K. y Vidakovic, D. (1996). Understanding the Limit Concept: Beginning whit a Coordinated Process Scheme. *Journal of Mathematical Behavior,* **15,** 167–192.

[12] Duval, R. (1995). Sémiosis et pensée humaine. Registres sémiotiques et apprentissages intellectuels. Nuchatel: Peter Lang.

[13] Elia, A., Gagatssi, A., Panaoura, A., Zachariades, T. y Zoulinaki, F. (2009). Geometric and algebraic approaches in the concept of "Limit" and the impact of the "Didactic Contract". *International Journal of Science and Mathematics Education*. Published online: 20 de February 2009.

[14] Engler, A., Vrancken, S., Hecklein, D., Müller, D. y Gregorini, M.I. (2007). Análisis de una propuesta didáctica para la enseñanza de límite finito de variable finita. *UNIÓN,* **11,** 113–132.

[15] Fernández-Plaza, J.A., Rico, L. y Ruiz-Hidalgo, J.F. (2013). Variación de las concepciones individuales sobre límite finito de una función en un punto. En A: Berciano, G. Gutiérrez, A. Estepa y N. Climent (Eds.) *Investigación en Educación Matemática XVII* (253–261). Bilbao: SEIEM.

[16] Garbin, S. y Azcarate, C, (2002). Infinito actual e inconsistencias: acerca de las incoherencias en los esquemas conceptuales de alumnos de 16–17 años. *Enseñanza de las Ciencias,* **20**(1), 87–113.

[17] Gras, R., Suzuki, E., Guillet, F. y Spagnolo, F. (Eds.) (2008). *Statistical Implicative analysis.* Theory and Applications. London: Springer.

[18] Gras, R. y Kuntz, P. (2009). El análisis estadístico implicativo (ASI) en respuesta a problemas que le dieron origen. En P. Orús, L. Zamora, P. Gregori (ed), *Teoría y aplicaciones del Análisis Estadístico Implicativo* (3–49). Castellón, España: Innovació Digital Castelló.

[19] Kidron, I. (2010). Constructing knowledge about the notion of limit in the definition of the horizontal asymptote. *International Journal of Science and Mathematics Education*, **9**(6), 695–717

[20] Lacasta, E. y Wilhelmi, M.R. (2010). Deslizamiento metadidáctico en profesores de secundaria. El caso del límite de funciones. En M.M. Moreno, A. Estrada, J. Carrillo y T.A. Sierra (Eds.), *Investigación en Educación Matemática XIV* (379–394). Lleida. SEIEM.

[21] Mamona–Downs, J. (2001). Letting the intuitive bear on the formal: A didactical approach for the understanding of the limit of a sequence. *Educational Studies in Mathematics,* **48**, 259–288.

[22] Mira, M., Valls, J. y Llinares, S. (2013). Un experimento de enseñanza sobre el límite de una función. Factores determinantes en una trayectoria de aprendizaje. *UNIÓN*, **36**, 89-107.

[23] Monaghan, J. (1991). Problems whit the language of limits. *For the learning of Mathematics,* **11**(3), 20–21.

[24] Monaghan, J. (2001). Young peoples' ideas of infinity. *Educational Studies in Mathematics,* **48**, 239–257.

[25] Moru, E. K. (2007). Talking with the literature on epistemological obstacles. *For the learning of Mathematics,* **27**(3), 34–37.

[26] Moru, E.K. (2009). Epistemological obstacles in coming to understand the limit of a function at undergraduate level: A case from the National University of Lesotho. *International Journal of Science and Mathematics Education,* **7**, 431–454.

[27] Oehrtman, M. (2009). Collapsing Dimensions, Physical Limitation, and Other Students Metaphors for Limit Concepts. *Journal for Research in Mathematics Education,* **40** (4), 396–426.

[28] Parameswaran, R. (2007). On understanding the notion of limits and infinitesimal quantities. *International Journal of Science and Mathematics Education,* **5**, 193–216.

[29] Pons, J., Valls, J., y Llinares, S. (2011). Coordination of Approximations in Secondary School Students' Understanding of Limit Concept. *Proceedings of the 35nd Conference of the International Group for the Psychology of Mathematics Education,* **4**(3), 393-400.

[30] Roh, K. H. (2008). Students' Images and their Understanding of Definitions of the Limit of a Sequence. *Educational Studies in Mathematics*, **69**, 217–233.

[31] Sierpinska, A. (1985). Obstacles épistémologiques relatives à la notion de limite. *Recherches en Didactique des Mathématiques,* **6** (1), 5–67.

[32] Szydlik, J.E. (2000). Mathematical beliefs and conceptual understanding of the limit of a function. *Journal for Research in Mathematics Education,* **31**(3), 258–276.

[33] Swinyard,C. (2011). Reinventing the formal definition of limit: The case of Amy and Mike. *Journal of Mathematical Behavior* (2011), doi:10.1016/j.jmathb.2011.01.001

[34] Tall, D. y Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limit and continuity. *Educational Studies in Mathematics,* **12**, 151–169.

[35] Trigueros, M. y Escandon, C. (2008). Los conceptos relevantes en el aprendizaje de la graficación. *Revista Mexicana de Investigación Educativa,* **13**(36), 59–85.

[36] Valls, J., Pons, J., y Llinares, S. (2011). Coordinación de los procesos de aproximación en la comprensión del límite de una función. *Enseñanza de las Ciencias,* **29**(3), 325-338.

[37] Williams, S.R. (1991). Models of limit held by college calculus students. *Journal for Research in Mathematics Education,* 22(3), 219–236.

[38] Zamora, L., Gregori, P. y Orús, P. (2009). Conceptos fundamentales del Análisis Estadístico Implicativo (ASI) y su soporte computacional CHIC. En P. Orús, L. Zamora, P. Gregori (ed), *Teoría y aplicaciones del Análisis Estadístico Implicativo* (65–101). Castellón, España: Innovació Digital Castelló.

LES SUPPORTS D'ENSEIGNEMENT DANS LA REPRESENTATION DU METIER CHEZ DES PROFESSEURS DES ECOLES DEBUTANTS

Marc BAILLEUL[1] et Laurence LEROYER[2]

SUPPORTS OF TEACHING IN THE REPRESENTATION OF TEACHING AMONG BEGINNING TEACHERS

RÉSUMÉ

Les supports d'enseignement sont au cœur de l'activité des enseignants, tant dans la phase de conception des séquences didactiques que dans celle de leurs pratiques de classe. Nous allons étudier dans ce texte la place que donnent des professeurs des écoles débutants à ces supports dans leurs représentations du métier. Nous allons pour cela mobiliser des résultats extraits de deux recherches sur la professionnalisation des enseignants, recherches menées respectivement en 2007 et 2012. En comparant ces résultats, nous mettrons en évidence des effets du changement de logique de formation voulu par l'institution entre ces deux dates.

Mots-clés : *supports d'enseignement, professionnalisation des enseignants, analyse statistique implicative.*

ABSTRACT

The teaching materials are at the heart of the activity of teachers, both in the design phase of didactic sequences than in their classroom practices. We will study in this paper the place that give novice teachers to these materials in the representations of their occupation. For this, we mobilize extracted results of two studies on the professionalization of teachers, research conducted respectively in 2007 and 2012. By comparing these results, we will highlight the effects of a change of training logic, desired by the institution, between these two dates.

Keywords : *supports of teaching, teachers professionalization, statistical implicative analysis.*

1 Une recherche sur la professionnalisation des enseignants

1.1 L'objet de recherche : la construction de la professionnalité chez les enseignants débutants

Notre intention consiste à mieux comprendre la façon dont les enseignants élaborent leur professionnalité en cours de formation (en IUFM[3], devenu ESPE4 en septembre 2011) en prenant en compte les différents contextes d'exercice (écoles, collèges, lycées) et les disciplines enseignées. Ce travail est réalisé par une équipe de chercheurs représentant plusieurs disciplines (sociologie, sciences de l'éducation, analyse du travail,…) et quatre institutions universitaires normandes, deux d'entre elles à vocation

[1] CERSE, Université de Caen, marc.bailleul@unicaen.fr
[2] ESPE, Université de Caen, 186 rue de la Délivrande, laurence.leroyer01@unicaen.fr
[3] Institut Universitaire de Formation des Maîtres
[4] École Supérieure du Professorat et de l'Éducation

formatrice (les ESPE de Basse et de Haute-Normandie), les deux autres centrées sur la recherche en sciences de l'éducation (les laboratoires CIVIIC[5] de l'Université de Rouen et CERSE[6] de Université de Caen Basse-Normandie).

Cette recherche a pour originalité de se démarquer sensiblement des recherches conduites jusqu'à ce jour sur la thématique de la professionnalisation des enseignants sous un double aspect :

- d'une part, nous n'abordons pas seulement la professionnalisation sous l'angle de la formation formelle (en IUFM puis ESPE) mais sous l'angle des transitions formation-travail et sous l'angle de la perception des apprentissages en cours d'activité professionnelle. L'enjeu consiste ici à mieux comprendre comment les enseignants parlent de et vivent leurs apprentissages professionnels par et dans la conduite des activités professionnelles ;
- d'autre part, nous croisons des grilles conceptuelles différentes pour comprendre de façon globale et non compartimentée, une réalité complexe (les conditions du développement professionnel).

1.2 Le cadre conceptuel de ces recherches

Nos travaux s'inscrivent dans le champ des réflexions sur la professionnalisation des enseignants (Bourdoncle, 1993). Nous entrons dans cette thématique en mettant en tension l'offre de professionnalisation (côté institution) et la dynamique de la construction initiale de la professionnalité (côté sujets), première étape de ce que nous appellerons le développement professionnel.

Tardif et Lessard (2000, p. 401), parlant des enseignants, insistent sur l'idée que « les connaissances issues de la pratique (connaissances ouvragées) semblent les fondements de la pratique du métier et de la compétence professionnelle […] Les routines (des enseignants) sont des moyens de gérer la complexité des situations d'interactions et de diminuer l'investissement cognitif de l'enseignant dans le contrôle des événements […] La force et la stabilité de ce contrôle ne peuvent pas dépendre de décisions volontaires de choix, de projets, mais bien de l'intériorisation des règles implicites d'action acquises avec et dans l'expérience de l'action… » Pour ces auteurs, l'activité enseignante présente ainsi un caractère syncrétique, « le syncrétisme signifie le fait que l'enseignement exige du travailleur une capacité à utiliser dans l'action quotidienne un large spectre de connaissances et de savoir-faire composites » (opus cité, p. 369). Au final, pour Tardif et Lessard, ce qui caractérise l'enseignant, c'est la production de savoirs d'expérience qui ont pour propriété d'être « ouvragés », liés aux tâches, façonnés, pratiques, « interactifs », « syncrétiques » et « pluriels », hétérogènes, complexes et « non analytiques », ouverts, « poreux », « perméables », « existentiels et faiblement formalisés », « temporels » et « évolutifs ».

Avec l'objectif de mieux comprendre la construction expérientielle des sujets dans l'activité (de travail et/ou de formation), nous proposons de différencier ce qui relève d'une logique de professionnalisation de ce qui relève d'une logique de développement professionnel.

---

[5] Centre de recherches Interdisciplinaires sur les Valeurs, les Idées, les Identités et les Compétences

[6] Centre d'Études et de Recherche en Sciences de l'Éducation

Cette notion de développement professionnel n'est pas, à proprement parler, construite. Elle surgit depuis quelques décennies dans des travaux qui s'inspirent des théories du développement (Piaget, Léontiev, Vygotski) pour étudier, comme le fait par exemple la didactique professionnelle, les apprentissages réalisés par des adultes en situation professionnelle. Les sensibilités sont alors différentes selon la référence théorique choisie.

Pour être plus précis, dans les travaux classiques, trois options semblent exister pour expliquer le développement : il se réalise par l'**activité** (thèse de Léontiev, 1984), par le **langage et l'interaction** (Vygotski, 1985) ou par la **pensée** (Piaget, 1970). On retrouve ces débats dans les sensibilités aujourd'hui présentes : la clinique de l'activité (Clot, Prot, Werthe, 2001) emprunte aux thèses de Léontiev et Vygotskiquand la didactique professionnelle (Pastré, Vergnaud, Mayen, 2006) est influencée par la thèse constructiviste de Piaget,…

Pour notre part, nous proposons la distinction suivante entre professionnalisation et développement professionnel (distinction largement développée dans Wittorski, 2007) :

- la professionnalisation aurait à voir avec une intention sociale de professionnalisation émanant d'une institution ou d'une organisation ; soit l'injonction faite aux individus en formation, d'engager un processus de construction-transformation de leurs activités au service d'une efficacité et d'une lisibilité plus grandes du travail. Cette intention se traduit par la proposition de dispositifs de travail et de formation particuliers ;
- le développement professionnel concernerait un mouvement conjoint de développement d'une activité, de stratégies et de dynamiques identitaires accompagnant la construction-transformation d'une professionnalité au niveau d'un individu agissant.

Notre groupe de recherche considère que le processus de développement professionnel des enseignants s'effectue dans les espaces de la formation (en IUFM), de la transition formation-travail (IUFM et établissement scolaire) et du travail investi (en établissement scolaire). Il correspond à trois dynamiques :

- une dynamique de transformation des connaissances et des compétences ;
- une dynamique de transaction individuelle entre une offre institutionnelle de professionnalisation, un mode de contrôle des professionnalités et des modalités effectives de construction de celles-ci comme enseignant stagiaire puis titulaire ;
- une dynamique de transaction individuelle entre l'identité héritée et l'identité vécue ou en construction/transformation.

1.3 Problématique générale de la recherche

Nous considérons ainsi que les enseignants en formation, en livrant, via un questionnaire, leur vécu et ressenti à propos de leur parcours de formation et de travail, révèlent dans le même temps la proximité ou la distance qu'ils ont vis à vis de l'offre-injonction institutionnelle de professionnalisation (discours de la formation et du terrain). Il y aurait ainsi une tension entre le projet de l'institution et les vécus des enseignants en formation et au travail. Le projet de l'institution est lisible à la fois dans les textes qui régissent la formation (de ce point de vue la prescription est forte), mais

aussi dans les pratiques de formation proposées et les modalités d'évaluation professionnelle (visites de formateurs et/ou d'inspecteurs notamment). Pour leur part, les enseignants vivent des expériences personnelles contrastées dans leur établissement d'exercice, ils sont confrontés à la fois à des discours venant de leurs pairs qui n'entrent pas toujours en résonance avec le projet de l'institution et à des réactions des élèves et des parents qui leur renvoient une image d'eux-mêmes et de leur métier qui les interroge. Dès lors, une tension est ressentie par les enseignants en formation qui est à l'origine du développement de stratégies identitaires qui nous semblent être au cœur du processus de développement professionnel des nouveaux enseignants. Nous faisons l'hypothèse que les réponses au questionnaire constituent des traces de ces stratégies.

En comparant les résultats issus de deux campagnes de collecte de données réalisées respectivement en 2007 et 2012, auprès de deux promotions ayant suivi des formations conçues sur la base de logiques de formation différentes, nous faisons l'hypothèse que les représentations globales du métier ne seront pas structurées de la même façon relativement à ces deux corpus de données.

De plus, nous chercherons à valider cette hypothèse sur une question particulière ; la place des supports d'enseignement dans ces représentations. Au cœur de l'activité enseignante, tant dans la phase de conception des séquences didactiques que dans celle de leurs pratiques de classe (Leroyer, 2010), il nous semble possible d'identifier ici des différences de conception du métier liées aux modalités de formation.

1.4 Méthodologie mise en œuvre

1.4.1 Le questionnaire

Le cahier des charges de la formation des enseignants (2007) définit les compétences attendues des enseignants en général, que ceux-ci exercent dans le premier degré ou le second degré. Ces compétences sont regroupées en dix grands domaines. Les 28 items[7] qui structurent le questionnaire renvoient à ce référentiel, avec la répartition suivante :

Concevoir son enseignement : 5 items (8-10-11[8]-13-14)
Travail en partenariat : 5 items (1-7-24-25-26)
Gérer la classe : 4 items (15-16-17-28)
Prendre en compte la diversité des élèves : 4 items (2-3-12-18)
Évaluer : 3 items (19-20-21)
TIC : 2 items (22-23)
Agir en fonctionnaire de l'État de manière éthique et responsable : 2 items (4-5)
Recherche et innovation : 1 item (27)
Maîtrise de la discipline : 1 item (6)
Maîtrise de la langue française : 1 item (9)

Pour chacun des items, trois jugements successifs sur une échelle en 6 points sont demandés (1 = non, pas du tout, 6 = oui, tout à fait). La compétence évoquée dans l'item est-elle importante ? Appartient-elle au domaine du « faisable » ? Est-elle en construction dans la pratique ? Chaque stagiaire (statut des répondants en 2007) ou

[7] Voir annexes 1 et 2.

[8] C'est l'item 11 qui est centré sur la place des supports d'enseignement ; « Être enseignant, pour vous, c'est… avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) »

étudiant (statut des répondants en 2010) était donc amené à produire une liste de 84 jugements (28 x 3). En outre, il est demandé avec qui ces compétences se construisent. Deux choix sont permis parmi six possibles : les autres professeurs débutants, les formateurs, les tuteurs terrain ou conseillers pédagogiques, les collègues de l'établissement, la famille ou les amis, seul dans la classe. La dernière page du questionnaire est consacrée aux variables qu'on peut désigner comme « supplémentaires », caractérisant les répondants : genre, ancienneté, parcours universitaire, disciplines enseignée pour les PLC[9], etc. L'objectif n'est en aucun cas de demander aux enseignants stagiaires/étudiants d'évaluer la formation qu'ils ont suivie ni même de comprendre quels apprentissages particuliers ils ont développé, mais plutôt d'analyser leur perception de ce qui permet de devenir un professionnel de l'enseignement d'un double point de vue :

- perception des compétences requises, « atteignables » et plus ou moins acquises à l'issue de l'année,
- modalités d'apprentissage perçues comme étant les plus efficaces pour acquérir ces compétences.

Les jugements sur l'importance des items dessinent une image du métier idéalisé. Leur analyse permet de mesurer l'impact de l'injonction de professionnalisation. En demandant des jugements sur la « faisabilité » ; nous accédons aux représentations du métier possible, en relation avec les expériences personnelles dans les établissements d'exercice. Les jugements sur des constructions de compétences nous renseignent sur la perception d'apprentissages réalisés au cours de l'année. Les écarts entre ces trois séries de jugements doivent permettre de repérer où portent, où agissent les tensions ressenties par les enseignants en formation, confrontés à des discours partiellement dissonants.

1.4.2 Le recueil des données

Un double recueil de données a été mené. Les effectifs se répartissent ainsi :

- Première phase de recueil des données (mai 2007) : PE[10] : 496 ; PLC : 350 ; PLP[11] : 54, soit un total de 900 stagiaires ayant répondu au questionnaire.
- Seconde phase de recueil des données (mai 2012) : PE : 150 ; PLC/PLP : 75.

Nous ne nous intéresserons dans ce texte qu'aux résultats concernant les PE, les effectifs PLC/PLP étant trop faibles en 2012.

2 Un résultat majeur

Nous nous centrerons dans ce texte sur le seul point de vue « Important ». L'analyse statistique implicative nous permet-elle de révéler des logiques d'acteurs différentes entre 2007 et 2012 ?

---

9 Professeur des Lycées et Collèges
10 Professeur des Écoles
11 Professeur de Lycée Professionnel

On trouvera les graphes implicatifs issus des analyses menées en 2007 et 2012 respectivement dans les annexes 3 et 4[12]. Le graphe de 2007 est très clairement structuré en deux réseaux disjoints. Celui qui figure sur la gauche de la figure (respectivement sur la droite) regroupe des items choisis en position 5 ou 6 (respectivement des items choisis en position 1 ou 2). nous interprétons ce graphe comme révélant deux rapport différents au projet institutionnel de professionnalisation, présent à travers les items du questionnaire : l'adhésion à ce projet (choix 5 ou 6) d'un côté, le rejet de ce projet de l'autre choix 1 ou 2). Il nous faut apporter ici une importante précision relative à la lecture du graphe : la lisibilité a été privilégié par le logiciel CHIC (Couturier, Almouloud, 2013 ; Ratsimba-Rajohn, 2013) lors de la réalisation du graphe, il ne faut donc pas tenir compte de la position des items, ni de la « hauteur » des réseaux, sur l'axe vertical pour leur accorder un poids correspondant à leur position sur cet axe. Pour information les items du « réseau du rejet » (à droite de la figure) ont des poids (pourcentage de choix) compris entre 4 et 6 % alors que ceux du « réseau de l'adhésion » ont des poids compris entre 30 et 93 % (variable attractrice 28 : être capable d'établir une relation de confiance avec les élèves). Il est intéressant de noter qu'à l'intérieur même du réseau de l'adhésion, il est possible de mettre en évidence deux sous-réseaux (voir annexe 5) dont l'un est structuré par les items du domaine didactique[13] (R2), l'autre réseau (R1) regroupant des items relevant d'autres problématiques plus transversales (citoyenneté, autorité, différenciation, dimension collective du travail).

Le réseau R2 est articulé autour des items suivants, qu'on peut qualifier de variables « nodales »[14] :

- item 20 (ajuster sa progression en fonction des résultats des évaluations) ;
- item 9 (participer à l'apprentissage de la langue française chez ses élèves à travers son enseignement) ;
- item 16 (susciter l'implication des élèves dans le travail) ;
- item 8 (aider les élèves à acquérir des méthodes de travail) ;
- item 15 (créer les conditions favorables à la mise au travail des élèves).

Les variables 11 (avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) et 21 (consacrer un temps, avant chaque évaluation, à l'explicitation des attentes et des critères) sont aussi, à un degré moindre (parce qu'alimentées chacune par des implications issues de l'autre sous-réseau) des variables nodales de ce sous-réseau R2.

Le graphe de 2012, à la différence du précédent au même seuil, ne fait pas apparaître de réseaux disjoints mais des chemins implicatifs entremêlés, aboutissant sur dix (!) variables attractrices et dont les origines sont, pour certains, des positions de rejet des items choisis ! Nous interprétons ce résultat comme la trace d'une représentation du métier très éclatée, non encore structurée suite à une année de formation professionnelle

[12] Dans ces graphes, il faut lire les noms des variables de la façon suivante : vi12.27 : item 27 choisi en position 1 ou 2, vi34.27 : item 27 en position 3 ou 4, vi56.27 : item 27 en position 5 ou 6.

[13] Nous qualifions de « domaine didactique » le regroupement des items suivants : 6, 8, 9, 10, 11, 12, 14, 15, 16, 19, 20 et 21.

[14] Voir Lahanier-Reuter, Gras, Bailleul dans cet ouvrage.

qui n'a pas permis aux étudiants interrogés de percevoir la complexité du métier autrement que comme une juxtaposition de préoccupations.

3 La variable 11, variable nodale des deux graphes

3.1 La variable 11 dans le graphe de 2007

Nous avons explicité dans la figure ci-dessous les items constituant le « cône 11 » pour l'année 2007. La participation à des collectifs de réflexion contribue fortement à l'analyse critique des supports d'enseignement qui constitue une ressource pour la prise en compte des items suivants :

- ajuster sa progression en fonction des résultats des évaluations (à travers la constitution d'une « banque de scénarios » potentiellement mobilisables à différents degrés de maîtrise des connaissances visées),
- créer les conditions favorables à la mise au travail des élèves (en proposant aux élèves des scénarios motivants et adaptés à leurs profils),
- participer à l'apprentissage de la langue française chez ses élèves à travers son enseignement (en s'appropriant des scénarios construits pas des « spécialistes » : auteurs de manuels, créateurs de ressources en ligne, etc.),
- être capable d'établir une relation de confiance avec les élèves (en assurant ainsi, d'une certaine façon, la « qualité » de sa prestation didactique, ce qui permet alors de « faire.autorité »)

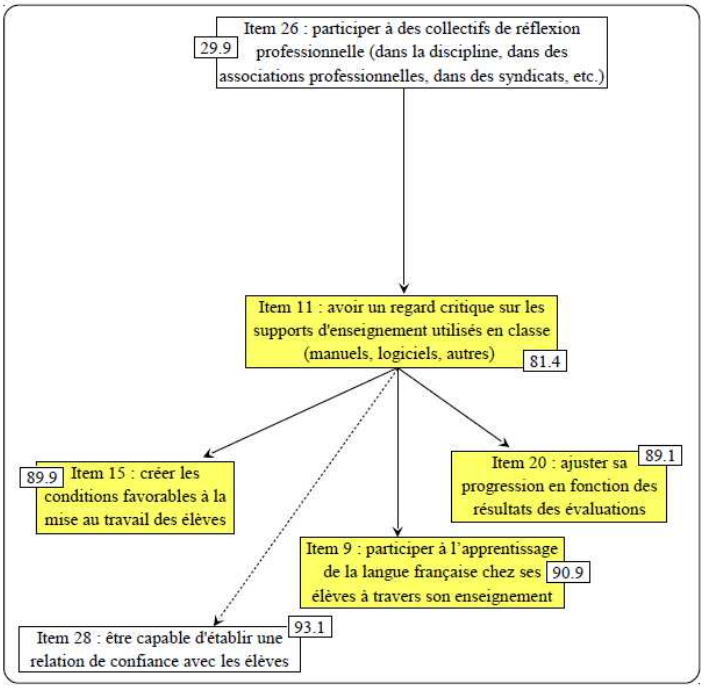


Figure 1 : La variable nodale 11 (avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) dans le graphe de 2007 NB : le grisé sur une case marque son appartenance au domaine didactique

Les pointillés entre les variables 11 et 28, dans le mode cône du logiciel CHIC, indiquent que cette implication « passe » par des chemins transitifs (11 → 9 → 28 ; 11

→ 15 → 28 ; 11 → 20 → 28).Notons le rôle essentiellement didactique de la variable 11.

3.2 La variable 11 dans le graphe de 2012

Nous avons explicité dans la figure de la page suivante les items constituant le « cône 11 » pour l'année 2012. Par simple perception visuelle, une différence apparaît très nettement entre les deux figures : l'item 11 a perdu, entre 2007 et 2012, de son importance didactique puisque seul un autre item de ce domaine (l'item 8 : aider les élèves à acquérir des méthodes de travail) figure encore dans le cône en 2012.

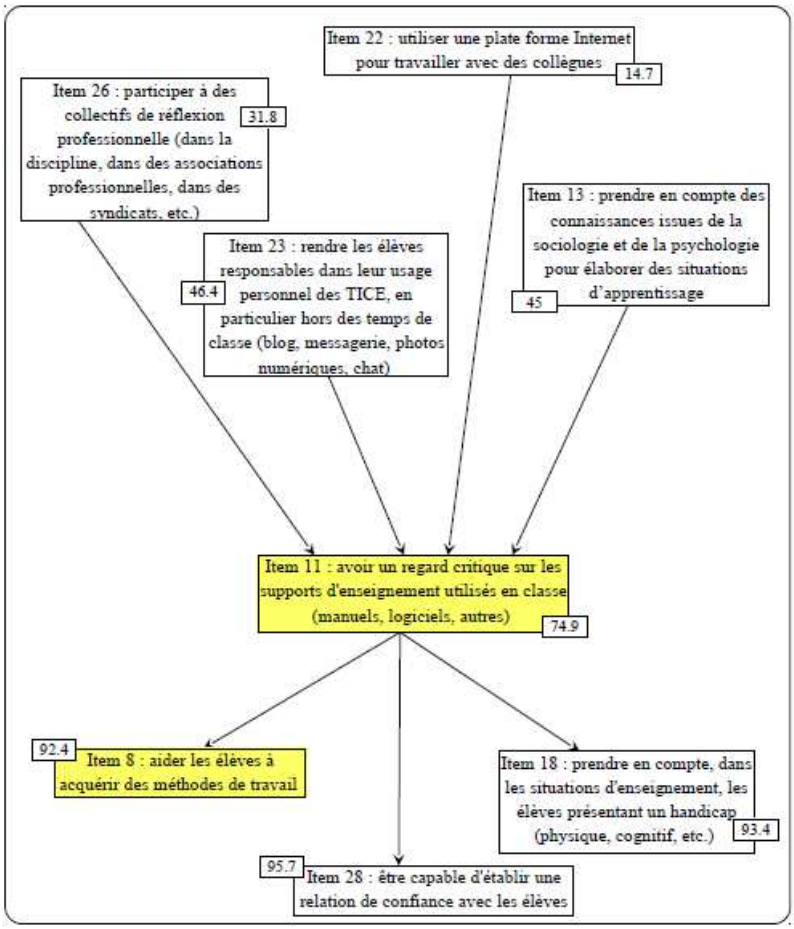


Figure 2 : La variable nodale 11 (avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) dans le graphe de 2012 NB : le grisé sur une case marque son appartenance au domaine didactique

3.3 Regard complémentaire sur les différences

Pour aller au-delà de cette différence perceptive, nous avons mobilisé le concept de variance implicative (Gras & Régnier, 2013). Nous avons calculé les variances implicatives intra-grappe de tous les cônes des deux graphes (11 pour le graphe de 2007 et 9 pour celui de 2012). Rappelons que la variance implicative n'étant pas un indice compris entre 0 et 1, nous ne pouvons que lui attribuer une valeur relative. Néanmoins nous pouvons comparer les places obtenues, pour chacune des deux années, par les cônes ayant pour variable nodale la variable 11, relativement à l'ensemble des autres

cônes. En 2007, le « cône 11 » est classé en position 6, sur 11 alors qu'il est en position 8 sur 9 en 2012. Non seulement les cônes étudiés ne concernent pas les mêmes items mais, de plus, le flux implicatif qui traverse chacun des deux cônes est donc plus significatif en 2007 qu'en 2012.

3.4 Discussion des résultats

Construire sa propre représentation de l'activité enseignante n'est pas chose simple pour un professeur en formation. Nous considérons l'activité comme la résultante d'un système de ressources et de contraintes composé de trois sous-systèmes en interaction que sont : l'enseignant exerçant, le contexte dans lequel il exerce et le genre professionnel (Leroyer, 2013). L'enseignant exerçant est considéré comme un sujet caractérisé par des connaissances, des valeurs et croyances, des émotions, des capacités physiques, etc. Le contexte est composé d'un environnement organisationnel, relationnel et matériel (Chatigny et Vézina, 2008). Enfin, le genre professionnel est constitué des formes communes de la vie professionnelle (Clot, Faïta, 2000).

L'existence en 2007, la non existence en 2012, d'une organisation des réponses au questionnaire à travers un graphe implicatif fortement structuré en 2007, et passablement éclaté en 2012, nous donnent à voir les effets de deux logiques de formation complètement différentes. Mai 2012 voit sortir des IUFM la première promotion qui a été formée selon les modalités de formation mises en œuvre suite à la réforme dite « de la mastérisation ». Une des principales différences concernait la confrontation au terrain à l'occasion des stages. Elle n'avait pas du tout la même importance : plus développée en 2007 qu'en 2012 et aussi beaucoup plus soucieuse de la mise en synergie théorie-pratique. Nous pensons que c'est cette synergie des propos entre les formateurs universitaires et les formateurs de terrain, à travers des co-interventions par exemple, qui permet aux professeurs débutants de mettre en cohérence les différents éléments d'un discours qui apparaît alors comme unique et représentant les attentes de l'institution. L'éviction, sur la période 2010-2012, des formateurs de terrain des interventions à l'IUFM et leur « cantonnement » aux classes, révélateurs d'une juxtaposition des savoirs et savoir-faire et non de leur inter dépendance, peut être avancée comme explication à l'éclatement de la représentation du métier. Les conditions de découverte du système de ressources et de contraintes évoqué plus haut sont, en 2012 et à la différence de 2007, défavorables puisque favorisant la construction d'une image « biface » du métier, assez souvent relayées par le discours des médias grand public.

L'étude, dans ce contexte, du rôle particulier de la variable 11 (avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) nous a semblé être prioritaire dans la mesure où, le temps consacré à la dimension *stricto sensu* professionnelle de la formation diminuant, il est urgent d'outiller les professeurs débutants de savoirs et de savoir faire centrés sur l'analyse critique des supports d'enseignement. Les résultats confirment, pour ce micro point de vue, l'analyse générale ci-dessus. La problématique du choix et de l'utilisation des supports d'enseignement n'est plus intégrée à une réflexion didactique elle-même partie prenante de l'activité enseignante. Elle n'apparaît, en 2012, que comme un des éléments qu'il est possible de mobiliser pour atteindre certains des multiples, ou du moins perçus comme tels, objectifs assignés aux professeurs en classe. La comparaison des variances

implicatives des cônes extraits des deux graphes confirme, pour cet item, la détérioration de la structuration des représentations du métier entre 2007 et 2012.

4 Conclusion

Nous avons, dans ce texte, comparé les résultats de deux recherches menées à cinq ans d'intervalle, avec un même support d'enquête et un même outil d'analyse statistique, sur deux promotions de professeurs des écoles en fin de formation, promotions qui ont suivi des parcours différents suite à la mise en place de la réforme dite « de la masterisation » de la formation des enseignants.

Le premier niveau de comparaison a été global : les deux graphes implicatifs issus des analyses menées avec le logiciel CHIC sont très différents, l'un fortement structuré, l'autre non.

Le deuxième niveau que nous avons étudié était un micro point de vue : le rôle d'une variable spécifique parce que perçue par nous comme particulièrement importante dans le contexte actuel de l'évolution de la formation des enseignants, l'item de l'enquête ciblé sur le trait de compétence « avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres) ». Là encore, un effet différenciateur a été mis en évidence rendant compte d'une détérioration de la perception de cet élément comme central dans une réflexion didactique au cœur de la pratique.

D'autres membres de l'équipe ont travaillé à une autre dimension mise en avant par l'institution comme une compétence clé des futurs enseignants : le travail en équipe (Tavignot, Thémines, Buhot, 2014).

L'organisation de la formation des enseignants a, une nouvelle fois, été remaniée avec la création des Écoles Supérieures du Professorat et de l'Éducation (ESPE) en 2013. C'est pourquoi nous allons reconduire cette enquête en mai 2015, faisant l'hypothèse de nouvelles transformations des représentations induites par les modalités de formation auxquelles sont confrontés les professeurs débutants. L'histoire se poursuivra donc dans un prochain colloque A.S.I. 9

Références

[1] Bourdoncle, R., 1993. La professionnalisation des enseignants : les limites d'un mythe. *Revue Française de Pédagogie*, 105, 83-114.

[2] Chatigny, C., & Vezina, N. 2008. L'analyse ergonomique de l'activité de travail : un outil pour développer les dispositifs de formation et d'enseignement. In Y. Lenoir & P. Pastré (dir.), *Didactique professionnelle et didactiques disciplinaires en débats, un enjeu pour la professionnalisation des enseignants,* Toulouse : Octarès, 269-284.

[3] Clot, Y., & Faïta, D. (2000). Genres et styles en analyse du travail. Concepts et méthodes,Vol. 4, Travailler.

[4] Clot Y., Prot B., Werthe C., 2001. Clinique de l'activité et pouvoir d'agir, *Éducation Permanente,* 46, 12-37.

[5] Couturier, R., Ag Almouloud, S., 2013. Historique et fonctionnalités de CH.I.C. in Gras, R. (dir.) *Analyse Statistique Implicative, Méthode exploratoire et confirmatoire pour la recherche de causalités*, 2ème édition revue et augmentée, RNTI E-16, Toulouse : Cépaduès éditions, 313-326.

[6] Gras, R. & Régnier, J.-C., 2013. Qualité d'un graphe implicatif : variance implicative, in Gras, R. (dir.) *Analyse Statistique Implicative, Méthode exploratoire et confirmatoire pour la recherche de causalités*, 2ème édition revue et augmentée, RNTI E-16, Toulouse : Cépaduès éditions, 207-218.

[7] Leontiev, A. 1984. *Activité, conscience, personnalité*. Moscou : Editions du progrès.

[8] Leroyer, L., 2013. Le rapport au support dans le travail de préparation en mathématiques des enseignants du premier degré, *Education & didactiques*, vol. 7, n° 1, 147-164.

[9] Ministère de l'Education Nationale, Cahier des charges de la formation des maîtres en institut universitaire de formation des maîtres, A. du 19-12-2006 JO du 28-12-2006, BOEN n°1, 4 janvier 2007.

[10] **Pastré** P., Mayen P., **Vergnaud** G. 2006. La **didactique professionnelle**. *Revue française de pédagogie*, 154, 145-198.

[11] Piaget, J., 1970. *L'épistémologie génétique*, Paris : PUF.

[12] Ratsimba-Rajohn, H., 2013. Guide d'utilisation des principales fonctionnalités du logiciel CHIC, Gras, R. (dir.) *Analyse Statistique Implicative, Méthode exploratoire et confirmatoire pour la recherche de causalités*, 2ème édition revue et augmentée, RNTI E-16, Toulouse : Cépaduès éditions, 327-348.

[13] Tardif, M. et Lessard, C., 2000. *Le travail enseignant au quotidien*. Louvain : De Boeck.

[14] Tavignot, P., Thémines, J.-F., Buhot, E. 2014. Intentions de formation et dimension collective : enseignants débutants, in *Actes du colloque Convisciencia*, Toulouse : ENFA.

[15] Vygotski L.S., 1985. *Pensée et Langage*, Terrains Editions Sociales.

[16] Wittorski, R. 2007. *Professionnalisation et développement professionnel*. Paris : L'Harmattan.

Annexe 1 : Le questionnaire, extrait de la première page.

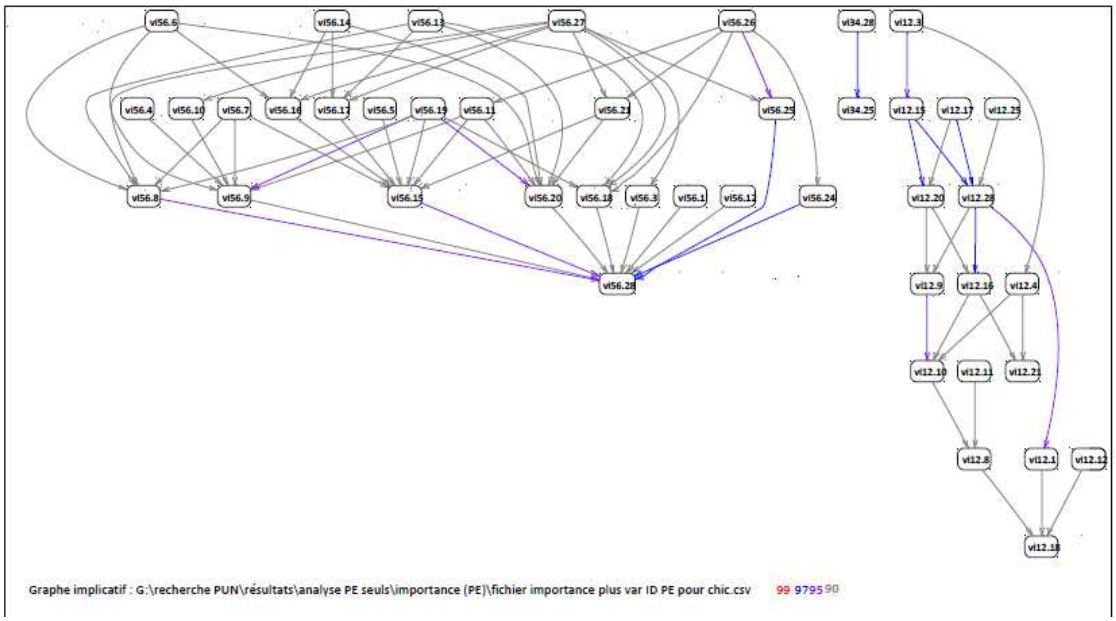
Être enseignant, pour vous, c'est…		Pour vous, la **construction** de cette compétence est…			Selon vous, cette compétence **se construit avec qui** ? (Cochez deux cases maximum par item)						
		importante	faisable	en cours dans votre pratique	avec vos collègues stagiaires	avec vos conseillers pédagogiques et tuteurs	avec vos formateurs	avec vos collègues dans l'établissement	seul dans l'établissement	avec des proches (parents, amis, …)	
Item 1	… pouvoir discuter avec des collègues des difficultés rencontrées dans la classe	-- ++	-- ++	-- ++							
Item 2	… consacrer du temps à s'entretenir avec les élèves en dehors de la classe	-- ++	-- ++	-- ++							
Item 3	… aider l'élève à se construire en tant que personne	-- ++	-- ++	-- ++							

Annexe 2 : Les 28 items du questionnaire

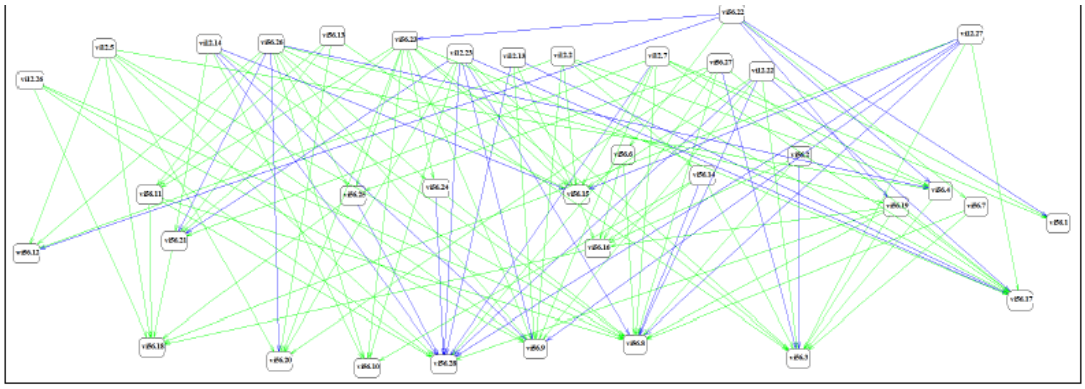
Être enseignant, pour vous, c'est…

Item 1	… pouvoir discuter avec des collègues des difficultés rencontrées dans la classe
Item 2	… consacrer du temps à s'entretenir avec les élèves en dehors de la classe
Item 3	… aider l'élève à se construire en tant que personne
Item 4	… former les élèves à une citoyenneté active
Item 5	… assumer une fonction d'autorité auprès des élèves
Item 6	… s'informer de l'actualité scientifique et / ou didactique dans sa ou ses disciplines d'enseignement
Item 7	… aider les élèves à faire leurs choix d'orientation
Item 8	… aider les élèves à acquérir des méthodes de travail
Item 9	… participer à l'apprentissage de la langue française chez ses élèves à travers son enseignement
Item 10	… concevoir une progression dans les notions à acquérir
Item 11	… avoir un regard critique sur les supports d'enseignement utilisés en classe (manuels, logiciels, autres)
Item 12	… aménager le programme pour s'adapter aux élèves
Item 13	… prendre en compte des connaissances issues de la sociologie et de la psychologie pour élaborer des situations d'apprentissage
Item 14	… développer des approches pluridisciplinaires et transversales (IDD, TPE, Education à..., PPCP, ECJS, etc.)
Item 15	… créer les conditions favorables à la mise au travail des élèves
Item 16	… susciter l'implication des élèves dans le travail
Item 17	… instaurer un climat de classe propice à la coopération entre élèves
Item 18	… prendre en compte, dans les situations d'enseignement, les élèves présentant un handicap (physique, cognitif, etc.)
Item 19	… diversifier les formes d'appréciation du travail des élèves
Item 20	… ajuster sa progression en fonction des résultats des évaluations
Item 21	… consacrer un temps, avant chaque évaluation, à l'explicitation des attentes et des critères
Item 22	… utiliser une plate forme Internet pour travailler avec des collègues
Item 23	… rendre les élèves responsables dans leur usage personnel des TICE, en particulier hors des temps de classe (blog, messagerie, photos numériques, chat)
Item 24	… être à l'écoute des parents qui rencontrent des difficultés dans l'éducation de leurs enfants
Item 25	… participer à des temps réguliers d'échanges entre professeurs dans l'établissement
Item 26	… participer à des collectifs de réflexion professionnelle (dans la discipline, dans des associations professionnelles, dans des syndicats, etc.)
Item 27	… se tenir informé des résultats de la recherche en éducation
Item 28	… être capable d'établir une relation de confiance avec les élèves

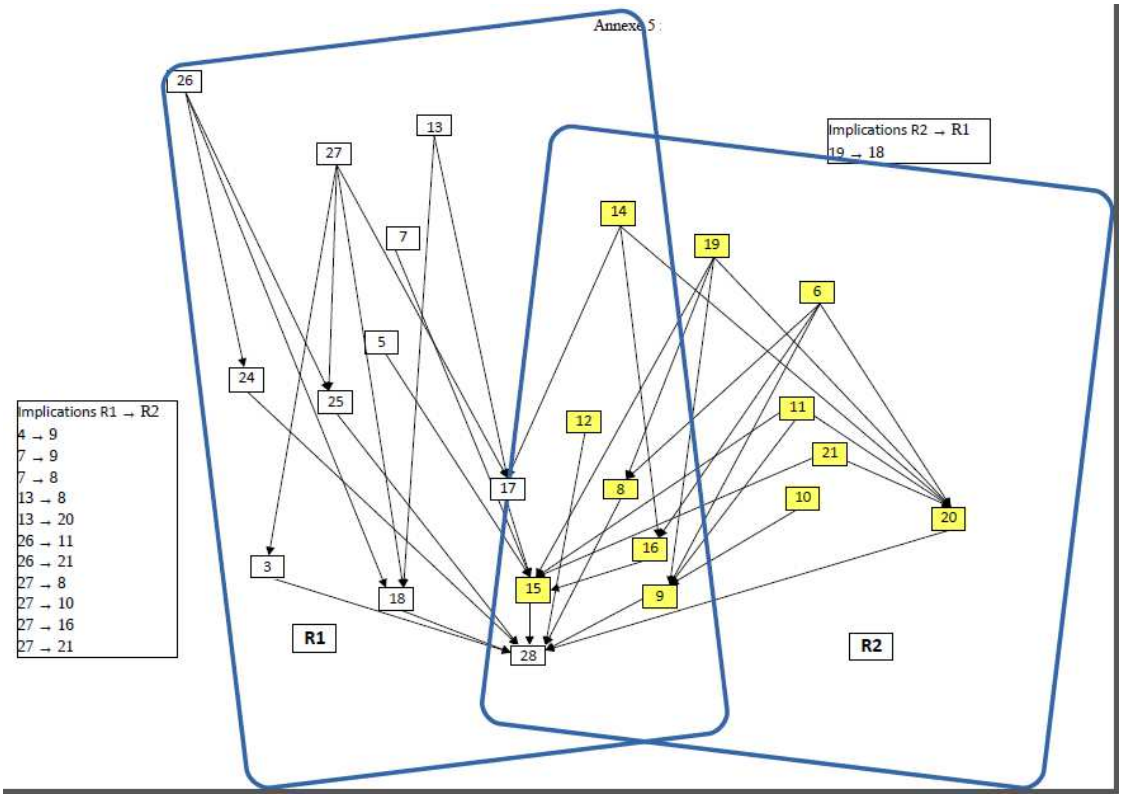
Annexe 3 : 2007, le graphe implicatif PE au seuil de .90



Annexe 4 : 2012, le graphe implicatif PE au seuil de .90



Annexe 5 : 2007, le réseau de l'adhésion et ses deux sous-réseaux



UMA ANÁLISE DAS DIFERENTES PRAXEOLOGIAS NO ENSINO DAS EQUAÇÕES DE SEGUNDO GRAU. UM OLHAR A PARTIR DA QUADRO TEORICO DA A.S.I[1]

Marcus Bessa DE MENEZES[2], Marcelo Câmara DOS SANTOS[3]

UNE ANALYSE DES PRAXÉOLOGIES DIFFÉRENTES DANS L'ENSEIGNEMENT DES ÉQUATIONS DU SECOND DEGRÉ. UN REGARD A PARTIR DE L'A.S.I.

AN ANALYSIS OF THE DIFFERENT PRAXEOLOGY IN THE TEACHING OF SECOND DEGREE EQUATIONS, A LOOK FROM THE S.I.A.

RESUMO

Essa investigação caracteriza-se por um novo olhar no trabalho de Tese defendida pelo proponente dessa pesquisa. A tese de doutoramento em questão tratou das diferentes praxeologias apresentadas pelo professor e pelos alunos durante o ensino das equações de segundo grau. Durante a investigação, pudemos observar como se comportam os saberes em sala de aula, o saber-fazer, que valoriza a técnica de resolução, e um saber de cunho mais científico que avança na consolidação do conhecimento. No entanto, algumas perguntas não foram respondidas por não ser o foco da pesquisa ou por falta de ferramentas que pudessem nos elencar informações sobre os eventos ocorridos. Com isso, em nossa nova pesquisa, propomos buscar elementos que respondam algumas perguntas abertas e promover um novo olhar sobre os dados que foram coletados a partir do quadro da A.S.I. Acreditamos que esse trabalho poderá fomentar discussões importantes para a formação do professor de Matemática quanto à gestão do saber em sala de aula.

Palavras chave: *Gestão da sala de aula; Fenômenos Didáticos; Teoria Antropológica do Didático; ASI.*

RÉSUMÉ

Cette recherche se caractérise par un nouveau regard sur le travail de thèse de l'auteur, sous la direction du coauteur de ce texte. La thèse de doctorat en question a traité des différentes praxéologies présentées par l'enseignant et ses étudiants pendant l'enseignement des équations du second degré. Au cours de la recherche, nous avons vu comment se comportent les connaissances dans la classe, le savoir-faire qui valorise la technique de résolution et une connaissance du point de vue plus scientifique, qui avance dans la consolidation des connaissances. Toutefois, certaines questions n'ont pas été abordées, soit pour n'être pas objet de la thèse, soit par le manque d'outils qui pouvaient nous fournir des informations sur les événements qui ont eu lieu dans le cours. Ainsi, dans cette nouvelle recherche, on a récupèré des éléments qui puissent permettre de répondre à des questions encore ouvertes et promouvoir un nouveau regard sur les données qui ont été recueillies,

[1] Esta pesquisa contou com o apoio da CAPES através da bolsa PVE do programa PPGEC-UFRPE com o Professor Jean-Claude Régnier UMR 5191 ICAR – Université Lyon2

[2] Universidade Federal de Campina Grande, Campus de Sumé-PB, Rua Luiz Grande S/N – Frei Damião, Sumé-PB – CEP 58540-000, marcusbessa@gmail.com

[3] Universidade Federal de Pernambuco, Colégio de Aplicação da UFPE, Avenida dos Funcionários, S/N - Cidade Universitária, Recife – PE – CEP 50740-550, marcelocamaraufpe@yahoo.com.br

dans le cadre de l'A.S.I.. Nous croyons que ce travail peut stimuler des discussions importantes pour la formation de l'enseignant de mathématiques au sujet de la gestion des connaissances dans la classe.

***Mots-clés**: Gestion de la classe ; Phénomènes didactiques ; Théorie anthropologique du didactique ; A.S.I.*

ABSTRACT

This investigation is characterized by a new look on work Theses defended by the proposer of this research. The doctoral thesis in question dealt with the different praxeologias presented by the teacher and the students during the teaching of second degree equations. During the investigation, we have seen how to behave in the classroom knowledge, know-how, who values the resolution technique, and a knowledge of more scientific slant that advances in the consolidation of knowledge. However, some questions have not been answered by not being the focus of research or by lack of tools that could list information about the events that occurred. With that, in our new survey, we get elements that respond some open-ended questions and promote a new look at the data that was collected from the frame of the S.I.A. We believe that this work can stimulate important discussions for the formation of the math teacher regarding the management of knowledge in the classroom.

***Keywords:** Classroom management; Didactic Phenomena; Anthropological theory of Didactic; S.I.A.*

1. Introdução

Procurando avançar em algumas questões abertas na tese intitulada: “Praxeologia do professor e do aluno: uma análise das diferenças no ensino de equações do segundo grau”, estamos propondo um novo olhar para os dados que foram coletados durante a tese. Para isso, utilizaremos o quadro metodológico da A.S.I. como uma ferramenta de análise, a partir do software CHIC[4].

Hoje em dia, a grande maioria das escolas apresenta um discurso construtivista para a construção do conhecimento. A ideia construtivista na Educação é a de que o conhecimento é construído a partir de um processo de ensino-aprendizagem infinito, ou seja, a relação com o objeto de saber (rapport au savoir) vai se modificando com o tempo, e não se restringe aos conteúdos trabalhados em sala de aula, pois há que se considerar o meio social, o contexto, a cultura local. César Coll (2006) afirma que o Construtivismo não é uma teoria, mas uma referência explicativa, um auxílio na reflexão sobre a prática docente. Pois há a necessidade de compreender que o aluno é um aprendiz social e o professor um agente mediador entre o indivíduo e a sociedade.

Apesar desses discursos, percebemos em sala de aula que muitos alunos não conseguem explicar os mecanismos ou as técnicas que utilizam para resolverem as atividades ou tarefas a eles propostas. É como se soubessem fazer, no entanto, não sabem muito ao certo porque fazem, ou seja, repetem o que o professor fez, o que está no Livro Didático que é adotado pela escola ou, até mesmo, a forma como um parente ou amigo realiza essas tarefas. Enfim, percebemos a existência de um saber-fazer em detrimento de saber de cunho mais científico, o qual forneceria condições aos alunos de explicarem exatamente suas escolhas para a realização de suas tarefas.

[4] Classification Hiérarchique Implicative et Cohésitive.

Em nossa tese, observamos que alguns alunos realizam suas atividades ou tarefas de formas diferentes das que seus professores apresentam, porém, não sabem justificar suas escolhas, ficando limitados ao saber fazer. Todavia, não era o foco da tese saber se existia uma estabilidade por parte desses alunos na realização dessas tarefas e quais os motivos que levariam a essa invariabilidade, ficando assim como uma questão em aberto. Será a partir dessa questão que iremos propor nosso trabalho de pesquisa, o qual irá buscar elementos que indiquem a estabilidade dos alunos na resolução de suas tarefas e qual a influência dos professores para que isso ocorra. Utilizaremos o quadro metodológico da A.S.I.[5] para nos apontar essa possibilidade (ou não) de estabilidade nas escolhas dos alunos para a realização de uma tarefa.

Com os dados encontrados nessa pesquisa, acreditamos que poderemos levantar uma discussão nos Cursos de formação de professores de Matemática: qual o saber que é valorizado em sala de aula pelo professor? Pensando no construtivismo como a concepção de ensino que é defendida pelos pesquisadores em Educação Matemática, essa discussão é extremamente importante para a formação do aluno, pois, no enfoque construtivista, se atribui ao sujeito um papel ativo, o aluno é responsável por seu próprio processo de aprendizagem, cabendo ao professor criar situações e implementar condições para que se desenrole o processo de construção.

Acreditamos ser necessário a apresentação de algumas teorias que orientaram nosso olhar para a sala de aula durante o desenvolvimento dessa pesquisa.

2. Fundamentação Teórica

2.1 Um breve histórico sobre a A.S.I.

É a partir dos trabalhos de Régis Gras na década de 1970 que se inicia e impulsiona desenvolvimento da análise estatística implicativa, sendo esses trabalhos amplamente apoiados pelas reuniões internacionais realizadas sucessivamente no IUFM de Caen (França), em 2000 e 2012; na PUC de São Paulo (Brasil), em 2003; na Universidade de Palermo (Itália), em 2005 e 2010; e na Universidade de Castellón (Espanha), em 2007. O pesquisador Saddo Ag Almouloud, nos aponta que “o desenvolvimento da A.S.I. vem, antes de tudo, do fruto da interação prévia e posterior das aplicações em diversas áreas, tais como as Ciências da Educação, Psicologia, Sociologia, Didática, Bioinformática, Medicina, História da Arte etc.” (ALMOULOUD, 2014, p. 624).

Ainda sobre a A.S.I, o pesquisador afirma o seguinte: “a Análise Estatística Implicativa (A.S.I.) visa à extração de conhecimentos, invariantes, regras indutivas não simétricas consistentes, e atribui uma medida para proposições do tipo ‘quando a é escolhido, tende-se a escolher b’.” (ALMOULOUD, 2014, p. 625).

Com intuito de buscar a explicitação de uma forma mais simples quanto possível, a intenção e o significado da análise estatística implicativa pode ser enunciada afirmando que a A.S.I. é um marco teórico para a análise de dados com base em uma relação não-simétrica. Sobre esse olhar, Gras & Regnier (2009) avançam:

> (...) um campo teórico, centrado no conceito de implicação estatística ou, mais especificamente, sobre o conceito de quase-implicação para

[5] Analyse Statistique Implicative.

> distingui-lo da implicação lógica das áreas de lógica e matemática. O estudo do conceito de quase-implicação como objeto matemático, nas áreas de probabilidade e estatística, permitiu construir ferramentas teóricas que instrumentam um método de análise de dados (GRAS, REGNIER, 2009, p. 12).

Em todos os colóquios realizados sobre a A.S.I., a abordagem foi realizada a partir de uma ferramenta fundamental, o software CHIC.

2.1.1 A ferramenta fundamental: o software CHIC

De acordo com Saddo Ag Almouloud (2014) O software, cuja sigla CHIC significa Classificação Hierárquica Implicativa e Coesitiva, foi elaborado na década de 1980 por Régis Gras, nos anos 90 por Saddo Ag Almouloud; posteriormente, por Harrison Ratsimba-Rajon, e, finalmente, por Raphaël Couturier nos dias de hoje.

O CHIC encontra-se na versão 6.d, podendo ser trabalhado em cinco idiomas: inglês, espanhol, francês, italiano, e português. Além de permitir, segundo Almouloud (2014): - tratar diferentes tipos de variáveis (binária, modais, frequênciasfrequências vetoriais, intervalos, fuzzy), - quantificar a significância dos valores atribuídos à qualidade, consistência da regra associada à relação implicativa entre variáveis, entre classes ordenadas de regras, a tipicidade e a contribuição de sujeitos ou categorias de sujeitos a determinadas regras, - representar por um gráfico, para um limiar de qualidade escolhido, os caminhos de regras e, por hierarquia, regras de entre regras (também chamadas de regras generalizadas), - excluir, adicionar, fazer conjunção de variáveis, - representar por uma hierarquia ascendente as classes de similaridade de variáveis.

Em sua última versão, um novo recurso, denominado '*Redução'*, foi introduzido, e é parte integrante da teoria da implicação estatística (A.S.I.) de R. Gras. Ainda segundo Almouloud, com esse novo recurso, "o CHIC permitirá reduzir um conjunto muito grande de variáveis a um subconjunto menor equivalente ao primeiro, no sentido de A.S.I., que permita ao usuário controle do limiar de redução, na tentativa de minimizar as informações ocultadas" (ALMOULOUD, 2014, p. 626).

2.2 Transposição Didática

A produção e a comunicação dos saberes de referência são necessidades sociais. O pesquisador, no mundo acadêmico/científico, sofre pressões internas e externas (ARSAC, 1989) para que comunique suas 'descobertas', suas 'teses'. As pressões internas aparecem quando a própria comunidade científica exige que tais saberes sejam comunicados, pois, a partir deles, novos saberes serão produzidos.

Por outro lado, existem, também, as pressões externas para a apresentação desse saber à sociedade. Os saberes comunicados, inicialmente no mundo acadêmico e científico, necessitam de um novo tratamento, no sentido de que sua roupagem mais acadêmica seja retirada e que ele possa, após essa primeira "transformação", ser comunicado, compreendido e, se possível, utilizado socialmente num período breve.

Acreditamos ser importante a identificação das diferenças entre os saberes que estarão envolvidos em nossa pesquisa, o primeiro deles será o saber científico. O saber

científico está associado à vida acadêmica, porém, devemos lembrar que não são todas as produções acadêmicas que serão saberes científicos. O saber científico é um saber criado nas universidades que irá servir de parâmetro para os saberes que irão chegar ao ensino básico, mas não está necessariamente vinculado a ele (ensino básico). A linguagem é uma das diferenças entre o saber científico e os outros saberes. Ela possui características diferentes nos outros saberes, visto que atende a um público específico, a comunidade científica, e assim sendo, não poderíamos levar esse tipo de linguagem para a sala de aula, pois dificilmente conseguiríamos auxiliar na compreensão e entendimento de nossos alunos.

Para identificarmos o próximo saber, o saber a ser ensinado, também chamado saber escolar, recorreremos a Luiz Carlos Pais (2001) que avança na seguinte explicação:

> O saber escolar representa o conjunto dos conteúdos previstos na estrutura curricular das várias disciplinas escolares valorizadas no contexto da história da educação. Por exemplo, no ensino da matemática, uma parte dos conteúdos tem suas raízes na matemática grega, de onde provém boa parte de sua caracterização. (PAIS, 2001, p 22)

Assim sendo, poderemos entender o saber a ser ensinado como todos os saberes eleitos para comporem a grade curricular de uma determinada disciplina. Será na "produção" do saber a ser ensinado que irão ser evidenciadas as diferenças, como avança Pais (2001), ao afirmar que na passagem do saber científico ao saber previsto na educação escolar ocorre a criação de vários recursos didáticos, cujo resultado prático ultrapassa os limites conceituais do saber matemático. A partir do surgimento desses recursos, surgem também as criações didáticas que fornecem o essencial da intenção de ensino da disciplina.

Outro ponto de diferença entre os saberes até aqui apresentados, está no seu aspecto. Enquanto o saber científico aparece a partir de artigos, teses, livros e relatórios o saber a ser ensinado se apresenta por meio de livros didáticos, programas e de outros materiais, o que ratifica a necessidade de uma linguagem diferente entre eles, tendo em vista o público ao qual são apresentados.

Segundo Ravel (2003), o saber preparado é o saber apresentado no plano de aula do professor, um saber que está envolvido com as expectativas que este professor tem com relação aos alunos, e ao saber a ser ensinado. Esse saber terá uma particularidade de que, normalmente, se apresenta de forma própria, pois as expectativas poderão ser diferentes para cada professor em relação ao grupo de alunos, que estão envolvidos no cenário didático.

O saber ensinado resulta das mudanças ocorridas durante a aplicação do que estava previsto no plano de aula (saber preparado) para o que efetivamente ocorre na sala de aula, ou seja, a realização, ou não, das expectativas. Esse saber será impregnado, principalmente, pela relação existente entre o professor e o saber a ser ensinado, o que irá orientar as mudanças que ocorrerão no processo de "produção" deste saber (saber ensinado), como avança Bessa de Menezes (2004):

> Um outro ponto está nas expectativas que os professores tinham em relação ao saber, fazendo, assim, com que esse objeto de ensino recebesse uma maior ou menor relevância no seu discurso em sala de aula, criando, desta forma, discursos diferentes para esses saberes em

> função dessas expectativas, as quais se apresentaram distintas para cada professor (BESSA DE MENEZES, 2004, p 131)

O saber aprendido seria o último saber dentro desse processo de apropriação do saber que ocorre em sala de aula. Diferente do que o nome dado a este saber possa parecer, principalmente para a área da psicologia da educação, iremos definir este saber como sendo todo e qualquer saber "retornado" pelo aluno, após esse saber ter sido "apresentado" em sala de aula. Sabemos que o mesmo (saber aprendido) não é somente formado pelo que é apresentado em sala de aula, ou seja, somente através do que é "ensinado"; temos consciência de que outras relações fora da sala de aula, na família, na comunidade em que vive, nos clubes, enfim, em outros locais onde pode aparecer esse saber em jogo, fazem com que nossos alunos tenham outras fontes para transformar este saber.

Para cada mudança no saber, nesse processo de intencionalidade do ensino, iremos ter as fases (ou etapas) da transposição didática, a saber: transposição didática externa e a transposição didática interna.

2.3 A Teoria Antropológica do Didático (TAD)

Segundo Chevallard, a sua teorização proposta na Teoria Antropológica do Didático (TAD) deve "(...) ser encarada como um desenvolvimento e uma articulação das noções cuja elaboração visa permitir pensar de maneira unificada um grande número de fenômenos didáticos, que surgem no final de múltiplas análises." (1998, p. 92)

Assim, podemos ver a TAD funcionando como uma forma de explicar a transposição didática (TD) no ecossistema[6] da sala de aula, ou melhor dizendo, um prolongamento da teoria da transposição didática, no momento em que amplia estes ecossistemas para relações, entre objetos de ensino, que irão além da sala de aula.

> Na prática, as primeiras análises propostas em *la transposition didactique*[7] limitavam-se a distinguir objetos «matemáticos», «paramatemáticas» e «protomatemáticos». O alargamento do quadro, levado a cabo por necessidades de análise, conduziu-me a propor uma teorização em que qualquer «objeto» pudesse aparecer : a função logarítmica é, evidentemente, um objeto («matemático»), mas existe igualmente o objeto «escola», o objeto «professor», o objeto «aprender», o objeto «saber», o objeto «dor de dente», o objeto «fazer xixi», etc. (CHEVALLARD, 1998, p.92)

O autor afirma que para começar sua teorização são necessários três conceitos primitivos: os objetos O, as pessoas X e as instituições I; e que outros virão a ser acrescentados subsequentemente.

O objeto O tomará uma posição privilegiada em relação aos outros temas, em virtude do mesmo ser o "material de base" da construção teórica. Tudo será objeto. Chevallard faz uma analogia com o universo matemático contemporâneo, o qual é fundado na teoria dos conjuntos, tudo é um conjunto. Assim também será na sua teoria,

[6] Entendemos ecossistema como sendo o local onde se desenvolve um determinado sistema que possui uma ecologia própria, no caso em estudo, o sistema didático.

[7] Ver Chevallard 1991.

"todas as coisas serão objetos", as pessoas X e as instituições I também são objetos, assim como as outras entidades que serão introduzidas.

O objeto irá existir no momento em que for reconhecido como existente por uma pessoa X ou instituição I. Com isso, aparecerão a relação pessoal de X com O, que será denotada por R(X, O), e a relação institucional de I com O, R(I, O). Ou seja, o objeto irá existir caso seja reconhecido por, pelo menos, uma pessoa X ou instituição I.

> Do ponto de vista da «semântica» da teoria, qualquer coisa pode ser um objeto. Um objeto existe a partir do momento em que uma pessoa X ou uma instituição I o reconhece como *existente* (para ela). Mais precisamente, podemos dizer que o objeto O existe para X (respectivamente, para I) se existir um objeto, que denotarei por R(X, O) (resp. R_1(O)), a que chamarei de *relação pessoal de X com O* (resp. *relação institucional de I com O*)[8]. (CHEVALLARD, 1998, p 93)

Chegamos a um ponto em que necessitamos evidenciar o que são as instituições. Segundo Chevallard (1998), "(...) uma instituição pode ser quase o que quer que seja". Devido à natureza da palavra, poderíamos dar uma conotação própria a esse personagem, ou seja: "Associação ou organização de caráter social, educativo, religioso, de ensino, etc." (Kury, 2002), porém, não devemos nos surpreender ao vermos, em certos momentos, objetos tomarem o status de instituição. Uma escola é certamente uma instituição, que possui outras instituições a ela agregada, uma sala de aula, por exemplo.

O conceito de Instituição pode ser explicitado como sendo um dispositivo social, total ou parcial, que impõe aos seus sujeitos formas de fazer e de pensar, que são próprias a cada "tipo ou forma" de instituição. Para avançarmos ainda mais sobre o conceito de instituição I, devemos percebê-la não como uma estrutura homogênea, mas sim heterogênea, em que existem várias relações de pessoas X com objetos O que pertencem a I.

Mas de que forma se relacionam os objetos O e instituição I? A cada instituição I está associado um conjunto de objetos O que são conhecidos por I, ou seja, existe uma relação institucional R(I, O).

> (...) A cada instituição I está associado um conjunto de objetos 0_1, chamado conjunto dos objetos *institucionais* (para I), que é o conjunto dos objetos O que I conhece, ou seja, para os quais *existe* uma relação institucional R_1(O). Um objeto O é institucional para I ou, dito de outro modo, existe para I, quando I define uma relação (institucional) com O.[9] (CHEVALLARD, 1999, p 225)

O objeto O se relaciona com a instituição I através de suas características próprias, por exemplo, a noção de porcentagem para uma instituição financeira (um banco) pode representar taxas e lucros, enquanto para a engenharia civil pode representar proporcionalidade entre partes de uma mistura (um traço de concreto). Assim sendo, o objeto O pode estabelecer diferentes formas de relações de acordo com a instituição R1(O), R2(O), R3(O), etc. Da mesma forma, seu desenvolvimento, dentro destas instituições, pode vir a ser modificado com o passar do tempo, ou seja, evoluir, envelhecer ou até mesmo desaparecer.

---

[8] Grifos do autor.
[9] Grifos do autor.

Essas relações são permeadas por outro fenômeno didático que surge nas relações dos sujeitos X com os objetos O da instituição I, fenômeno este que se estabelece devido às expectativas que existem dentro das relações, o contrato didático.

Apesar de já termos diversas vezes citado o próximo conceito fundamental da TAD, a pessoa, e não termos, ainda, definido o seu conceito, chegou a sua vez. Para isso, iniciaremos diferenciando alguns estágios deste conceito, a saber: o indivíduo, o sujeito e a pessoa. Podemos dizer que o estágio mais primitivo seria o de Indivíduo, visto que, ele não se sujeita, nem muda com as relações cotidianas com objetos e instituições. Chevallard afirma que:

> Bem entendido, no curso do tempo, o sistema das relações pessoais de X evolui; objetos que não existem para ele passam a existir; outros deixam de existir; para outros enfim a relação pessoal de X muda. Nesta evolução, o invariante é o indivíduo; o que muda é a pessoa (CHEVALLARD, 1999, 226).

O indivíduo se torna um sujeito quando se relaciona com uma Instituição I qualquer, ou melhor dizendo, quando se sujeita a uma Instituição I, sob suas demandas, hábitos, formas; enfim, se sujeitando a esta relação.

É por meio das várias relações que o indivíduo tem com instituições diferentes que se constitui a pessoa, ou seja, o conjunto de sujeitos do indivíduo é que forma a *pessoa* X, a qual irá mudando conforme estabelece suas relações com as instituições, as quais toma conhecimento com o passar do tempo.

> Uma pessoa X está sujeita a uma série de instituições. Introduzo aqui o axioma segundo o qual uma pessoa não é, na realidade, mais do que *a emergência de um complexo de sujeições institucionais*. Aquilo que se chama de «liberdade» da pessoa surge então com o efeito obtido em conseqüência *de uma ou de várias sujeições institucionais contra outras*.[10] (CHEVALLARD, 1999, p. 227)

Uma pessoa X entra para uma instituição I, e existe um objeto O para I, que é chamado de objeto institucional. Assim, X ao entrar em I, começa a viver uma relação com O sob a influência da relação institucional, ou seja, a relação R(X, O) irá se alterar ou se construir mediante a relação R(I, O), e, de forma mais ampliada, sob o constrangimento do contrato institucional C.

Devemos deixar claro que O poderia ou não existir para X antes de sua entrada para I (que analogamente podemos sugerir como conjunto vazio, sem existência), porém, independente desse fato, a relação R(X, O) irá alterar-se Daí, então, Chevallard dirá que há aprendizagem de X em relação a O. Ou seja, havendo alteração em R(X, O) então haverá aprendizagem da pessoa X sobre o objeto O. De forma análoga, caso R(X, O) não se altere, podemos afirmar que nada aprendeu. Devemos observar que não há nada de didático até agora, pois a instituição I não se manifestou com intencionalidade de fazer com que R(X, O) se altere ou modifique.

Para que a instituição I manifeste uma intencionalidade de fazer uma modificação ou uma alteração na relação R(X, O), é necessário que se introduza uma nova noção

[10] Grifos do autor.

primitiva, a do sujeito adequado. Com isso, uma pessoa X se tornará um sujeito da instituição I, relativamente ao objeto O, quando as relações R(X, O) e R(I, O) estão em conformidade. Ou seja, o sujeito está cumprindo as expectativas desejadas pela Instituição, está conforme "deseja" a Instituição. Caso isso não esteja ocorrendo, é considerado que o sujeito está inadequado em relação ao contrato institucional C.

Assim, entra em cena um desenvolvimento relativo à avaliação institucional. Segundo Chevallard (1999), essa avaliação é um dos mecanismos segundo os quais I é levada a pronunciar, através de alguns dos seus agentes, um veredicto de conformidade (ou de não conformidade) R(X, O) com R(I, O).". Ainda sobre esse assunto, o autor afirma que:

> A este respeito, as instituições são sempre «vigarizadas» (trapaceadas) pelos seus sujeitos. Quando esperam encontrar *sujeitos puros*, que julgam ser inteiramente moldados por elas, deparam-se com *pessoas*, que lhes aparecem sempre, de uma forma ou de outra, como *sujeitos desadequados*. (CHEVALLARD, 1999, p. 227)

É preciso ressaltar que, em nosso trabalho, desejamos observar a Instituição Sistema Didático, pois será lá o local em que poderemos verificar as transposições que serão efetuadas pelo aluno enquanto pertencente a esta instituição. Em outra instituição, com outros contratos estabelecidos, em virtude dos ecossistemas serem distintos, identificaríamos variantes que não poderíamos analisar no âmbito desta tese.

Assim sendo, podemos pensar que a instituição sala de aula – a qual chamaremos, a partir deste momento, de I1 – tem em seus sujeitos X1 – os alunos –, objetos O1 – saberes em jogo – e seus agentes que irão regular a conformidade, ou a não conformidade, com a instituição I1, de acordo com a intencionalidade estabelecida – são os professores, o contrato didático e o contrato institucional estabelecidos, as avaliações, entre outros, que aparecerão de acordo com o momento necessário.

A avaliação, como um dos elementos controladores da conformidade, ou não conformidade, na Instituição I1, pode, nesse sentido, vir (ao contrário do que se espera) a podar todo esse interesse pelo objeto O1, fazendo com que o sujeito X1 se preocupe somente com a conformidade, ou seja, quais são os conjuntos de sequência que deve realizar com intuito de ter a "adequação" esperada. Não podemos esquecer que essa avaliação é estabelecida por meio de um contrato pedagógico e um contrato didático definidos, que, de certa forma, dá sua importância dentro de I1. Assim sendo, isso poderá comprometer a formação dos conceitos desse objeto O1 em jogo no cenário didático.

Essas alterações nas relações entre o sujeito X1 e o objeto O1, (ou as transformações que o aluno faz do saber, quando tratamos da questão da Transposição Discente), vão muito além de uma questão epistemológica do objeto O1 (saber) ou de uma questão metodológica. Elas partem, também, de uma intencionalidade vinculada ao contrato que é estabelecido. Não significa que deixamos de fora esses outros fatores. Porém, é extremamente necessário, quando olhamos para o saber aprendido, a relação entre os contratos (pedagógico e didático) estabelecidos que têm, se assim podemos dizer, um peso maior nas escolhas realizadas pelos sujeitos X1 (alunos).

Algumas relações entre sujeitos, objetos e instituição são permeadas por intencionalidades diversas, tanto por parte dos sujeitos como, também, das instituições perante os objetos em jogo nessa relação. Fazendo um paralelo com a sala de aula,

podemos identificar vários fenômenos didáticos que ocorrem devido a essas intencionalidades, mediante as relações entre alunos e professores diante do saber a ser ensinado.

2.3.1 A Organização Praxeológica ou Praxeologia

Podemos entender uma organização praxeológica, ou praxeologia, como a realização de certo tipo de tarefas (T) através de um modo de fazer, que Chevallard (1999) chama de técnica (t). Essa associação tarefa-técnica (T-t) irá definir um saber-fazer próprio para esse tipo de tarefa. Porém, ela (T-t), não se mantém em estado isolado, ou seja, não se sustentará por si só. A T-t necessita de um amparo tecnológico-teórico (ou saber), que é formado por uma tecnologia (θ), que irá dar uma racionalidade e uma sustentação inteligível à técnica (t) aplicada, e uma teoria (Θ) que irá justificar e esclarecer a tecnologia (θ).

Assim sendo, a organização praxeológica ou praxeologia (que a partir desse momento iremos tratar somente como praxeologia) será composta por quatro elementos, a saber: tipo de tarefa (T), técnica (t), tecnologia (θ) e teoria (Θ); articulados a partir de um bloco prático-técnico (gerando o saber-fazer) e um bloco tecnológico-teórico (amparado no saber).

Segundo Chevallard (1998), a noção de tarefa, ou de tipos de tarefas, se encontra na raiz da noção de praxeologia. Podemos entender como tarefa (T), de acordo com a TAD, como todo e qualquer objeto que não encontramos sua existência diretamente na natureza, ou seja, será necessário realizar procedimentos próprios, no caso de nosso estudo: matemáticos, para encontrá-lo. Quando uma tarefa (t) ressalta de um tipo de tarefa T, escreveremos então: (t) ϵ T.

Para Chevallard (1998) podemos, ainda, diferenciar o gênero de tarefa do tipo de tarefa ou tarefa propriamente dita. O gênero de tarefa seria caracterizado por um verbo, como, por exemplo, montar, levar, calcular, etc., sendo expresso de forma mais ampla e conteúdo não definido. Já o tipo de tarefa, ou tarefa, tem seu conteúdo estritamente especificado. Assim sendo, para exemplificarmos um tipo de tarefa, temos resolver uma equação de primeiro grau, encontrar a altura de um triângulo isósceles, etc.

Chevallard (1998) observa ainda que um determinado tipo de técnica (t) não é universal para todas as instituições I. Em certos casos, algumas instituições não estão em conformidade com determinados tipos de técnicas, e assim sendo, não reconhecerão e contestarão a validade desta técnica (t).

Para poder dar um suporte racional e justificar a técnica (t) aplicada para a realização de uma tarefa (T) é necessário a introdução da noção de tecnologia (θ), a qual é definida por Chevallard (1998) como sendo:

> (...) um discurso racional (*logos*) sobre a técnica – a tekhnê – t, discurso tendo por objetivo primeiro de justificar 'racionalmente' a técnica t, e nos assegurar que ela permite o bom cumprimento das tarefas do tipo T, isto quer dizer realizar o que é pretendido. (CHEVALLARD, 1998, p 93)

Ainda sobre tecnologia (θ), Chevallard afirma que em dada instituição I uma técnica (t) para a realização de um tipo de tarefa (T) vem, frequentemente, acompanhada de vestígios ou embriões de tecnologia (θ), e, em diversos casos, na técnica (t), certos elementos tecnológicos vêm incorporados. O autor avança ainda, ao afirmar que quando em uma instituição I existe, em princípio, somente uma técnica (t) que é reverenciada, reconhecida e empregada, esta técnica adquire um papel de "autotecnológica", ou seja, não irá necessitar de justificativas, pois esta é a melhor maneira de se fazer nesta instituição I.

Para assegurarmos o funcionamento regular de uma tecnologia (θ) em uma instituição I, necessitamos de uma nova noção que explique e justifique esta tecnologia (θ). Esse fato nos leva com a noção de Teoria (Θ) que é a especulação abstrata da tecnologia; no plano teórico encontram-se as definições, os teoremas, as noções mais abrangentes que servem para explicar, justificar e produzir novas tecnologias.

Segundo Chevallard (1998), poderíamos chegar a uma regressão absurda, na qual sempre teríamos que justificar uma coisa atrás da outra, ou seja, a técnica justificada por uma tecnologia, que é justificada por uma teoria, que seria justificada por outra teoria, por outra e outra... Porém, o autor afirma que "(...) a descrição em três níveis (técnica/tecnologia/teoria), em geral, é o suficiente para dar conta da atividade a analisar" (CHEVALLARD, 1998, p 94).

Para analisarmos a prática docente, devemos observar as seguintes questões: Como realizar a tarefa do tipo T? Ou ainda, Como realizar melhor esta tarefa? Essas questões invocam uma produção de técnicas, e, portanto, de praxeologias.

Sendo os tipos de tarefa T, acima citados, objetos matemáticos O para serem tratados em uma instituição I (uma sala de aula qualquer), podemos considerar essa análise em duas classes distintas: a) observando o primeiro questionamento com um viés pela realidade matemática, poderemos construir uma realidade como uma praxeologia matemática ou organização matemática, a qual denominaremos como OM; b) ao observarmos o segundo questionamento, teremos um olhar sobre a didática, ou seja, de que forma encaminharemos a realidade matemática estabelecida na OM. Assim, essa realidade se denominará uma praxeologia didática ou uma organização didática OD.

Chamaremos de praxeologia matemática ou organização matemática, toda realidade matemática que está envolvida na resolução de um tipo de tarefa matemática T. Para isso, serão exigidas técnicas t, amparadas por um conjunto teórico-tecnológico [θ; Θ].

A organização matemática tem sua origem nas análises, efetuadas pelos professores[11], dos documentos oficiais existentes (tais como programas e manuais escolares, além do livro didático), dos quais saem os saberes matemáticos escolhidos a serem ensinados.

Outro ponto da prática docente será de como conduzir esta praxeologia matemática, agora estabelecida, para a sala de aula. Isto é, como transpor da realidade matemática para a realidade didática. Segundo Chevallard (1999), a construção da praxeologia se inicia em uma falta de técnica para a resolução de um determinado tipo de tarefa. Assim sendo, podemos pensar no exemplo dado anteriormente: "Como encontrar as raízes de

[11] Lembramos que, nesse momento em particular, estamos fazendo um olhar pela prática docente.

uma equação de 2º grau?", e fazer agora a seguinte questão: "Como ensinar a encontrar as raízes de uma equação de 2º grau?" Dar resposta a esta nova questão nos leva a elaborar um novo tipo de praxeologia, a praxeologia didática.

A organização (ou praxeologia) didática surge na intenção de pôr em prática, ou de conduzir, uma organização matemática qualquer. Será ela, a OD, que irá dar conta da (re)construção ou transposição de uma determinada OM. Assim como toda praxeologia, a OD é composta de tipos de tarefas que serão resolvidas por técnicas, as quais serão explicadas pelas tecnologias e justificadas por teorias.

Para identificarmos as praxeologias, do professor e dos alunos, e fazermos a análise das diferenças, e elencar elementos para análise a partir do quadro metodológico da ASI, foram realizadas algumas etapas, as quais apresentaremos a seguir na metodologia utilizada.

3 Metodologia

3.1 Caracterização dos sujeitos e ações seguidas

Tomamos como sujeito um professor do 9º ano, com formação em Licenciatura em Matemática, e seus 24 alunos. O saber matemático que será observado são as equações do segundo grau. Isso se deve por ser nesse ano (9º) que essas equações são introduzidas formalmente no domínio algébrico no Brasil. A escolha pela álgebra se dá pelo fato de considerarmos que a transição do domínio aritmético para o algébrico marca uma das mais importantes rupturas no ensino de matemática: de uma matemática mais 'concreta' a um campo que exige um nível maior de abstração e generalização.

Uma questão que achamos pertinente destacar aqui, diz respeito à escolha da equação do segundo grau como conteúdo matemático a ser contemplado nesse estudo. As pesquisas voltadas para a Educação Matemática apontam para uma ruptura existente na passagem da aritmética à álgebra (VERGNAUD e CORTES, 1986; VERGNAUD, CORTES e FAVRE-ARTIGUE, 1987; KIERAN, 1992; BOOTH, 1995; USISKIN, 1995, dentre outros). Entretanto, entendemos que dentro da própria álgebra o aluno também se depara com outra ruptura, ao passar das equações de primeiro grau, para as equações de segundo grau. Enquanto que, no primeiro tipo de equações (1º grau), o aluno elege o procedimento de resolução (transposição de um membro para outro da igualdade, realizando a operação inversa, por exemplo), nas equações de 2º grau ele precisa lançar mão de outros procedimentos, como fatorar a equação ou mesmo utilizar a fórmula de Bhaskara.

Durante a coleta de dados para a tese (que é a base dessa nossa nova pesquisa), tivemos como referencial para observar a dinâmica da sala de aula a teoria antropológica do didático (TAD) (CHEVALLARD, 1999), a metodologia consistiu de uma análise das atividades propostas pelo professor em sala de aula, quando analisamos a sua prática sob o olhar da praxeologia, e comparamos com as atividades realizadas pelos alunos, também sob a ótica da praxeologia. Nas análises buscamos evidenciar os elementos que nos apontaram para as diferenças nas praxeológias do professor e dos alunos (técnicas, tecnologias e teorias) que vieram a aparecer.

Para identificarmos a praxeologia do professor e do aluno tivemos que realizar algumas ações. Primeiramente, realizamos uma descrição e análise do livro didático utilizado, a primeira ação. Com isso, buscamos estabelecer contato com o saber a ensinar, a partir da análise do livro didático utilizado, especificamente conteúdo de equação do segundo grau. Considerando que este tem se mostrado como uma espécie de "texto do saber", conforme a conceituação de Chevallard (1991). Posteriormente, observamos o professor, a segunda ação a ser realizada foi de identificar os tipos de tarefas que foram propostas. Para identificarmos esses tipos de tarefas, filmamos as aulas do professor sobre o conteúdo escolhido, equações de segundo grau, além de identificarmos elementos ostensivos e não-ostensivos que apareceram durante essa filmagem. A partir daí, partimos para a terceira ação que foi fazer uma análise da Organização Matemática, na qual identificamos as técnicas, os elementos tecnológicos e teóricos que se apresentam para a realização da tarefa. Com isso, identificamos a Organização Matemática proposta pelo professor.

Após a identificação dos elementos componentes da Organização Matemática, realizamos a quarta ação que consistiu em identificar quais são as técnicas, tecnologias e teorias mobilizadas pelo professor em sala de aula, também através das filmagens feitas anteriormente, para a realização do tipo de tarefa proposta, ou seja, a Organização Didática. Diferentemente da Organização Matemática, que tem um olhar sobre a realidade matemática envolvida, a Organização Didática foca em como pôr em prática o conteúdo matemático, nesse caso a equação do segundo grau. Com a intenção de facilitar o entendimento da Organização Didática realizada pelo professor, realizamos uma quinta ação que foi uma entrevista semiestruturada com o professor, na qual buscamos identificar suas escolhas na OD. A escolha por esse tipo de entrevista se deu porque ela pode nos permitir que, através do conteúdo manifesto da fala, alcançar respostas mais ricas e complexas, o que acreditamos ser mais proveitoso para o nosso trabalho, uma vez que reconhecemos que apenas com as observações não iríamos conseguir as informações necessárias, relativas ao interesse pela álgebra, especificamente no conteúdo de equações de segundo grau, o que como já citamos anteriormente, é um elemento importante para a formação do saber ensinado. Visto que, pode revelar indícios de sua relação (do professor) com o saber em jogo (equação do segundo grau). Para a identificarmos os elementos da praxeologia do aluno, elaboramos uma lista de atividades que continha os mesmos subtipos de tarefa que foram apresentados pelo professor em sala de aula. Nessa lista, buscamos as técnicas, tecnologias e teorias que foram mobilizadas pelos alunos participantes da pesquisa na realização desses subtipos de tarefas. Essa foi a nossa sexta ação. Será a partir desses dados coletados que iniciamos nossa pesquisa com a A.S.I. Diante das devolutivas dos alunos na atividade proposta, elaboramos nossas variáveis e geramos as tabelas necessárias para a análise com o software CHIC. A lista de atividades possui nove itens, os quais serviram de variáveis para análise do tipo de técnicas que foram utilizadas para a resolução de cada item, bem como, para identificarmos a praxeologia que foi utilizada pelos alunos. Os itens 'c' e 'h' da lista, não foram observados na pesquisa por erros de digitação na sua confecção, assim sendo, foram analisados sete itens.

A seguir, apresentamos a Lista de Atividades que foi aplicada aos alunos.

Ficha de Atividades – Equações do 2° Grau

Encontre o (s) valor (es) de "x" nas equações abaixo:

a) $2x^2 - 98a^6 = 0$
b) $2X^2 - 8x = 0$
c) $4x^2 - 20x = -35$
d) $x^2 - x - 6 = 0$
e) $(x - 8).(x + 4) = 0$
f) $(2x + 5)^2 = 0$
g) $(x + 4).(x - 7) = 5x + 5$
h) $3x^2 + 6x = -4$
i) $\frac{6x + 10}{2} - 3x^2 = 3x - 103$

Figura 1 – Lista de Atividades aplicada aos alunos.

3.2 Tarefa em jogo, Variáveis e as delimitações para o uso do quadro metodológico da A.SI..

A tarefa T em jogo foi de *Resolver equações de segundo grau*. Durante as aulas foram apresentadas, pelo professor, 07 (sete) subtipos de tarefas relativas a T, nas quais identificamos as seguintes estruturas:

Subtipos de tarefas identificadas

$T_1: ax^2 + c = 0$
$T_2: ax^2 + bx = 0$
$T_3: (ax + c)^2$
$T_4: (x + a).(x + b) = 0$
$T_5: (x + a).(x + b) = cx + d$
$T_6: \frac{(ax + b)^2}{c} + dx^2 = ex + f$
$T_7: ax^2 + bx + c = 0$

Figura 2 – Subtipos de tarefas identificadas nas durante as aulas

Na tarefa T, em nenhum momento, se propõe a identificação ou a conceituação de equações do segundo grau. Todo o trabalho do professor gira em torno da resolução destas equações. Isso é percebido, em sala de aula, no momento em que as aplicações se voltam a simplesmente resolver tipos de equações. O trabalho do professor finda em sala de aula com a resolução de exercícios. O professor apresenta alguns subtipos de tarefa, os quais identificamos sete, anteriormente enumerados em T_1, T_2, T_3... T_7, e os resolve utilizando-se de técnicas por ele (o professor) apresentadas. Podemos colocar, em linhas gerais, que o procedimento do professor em sala de aula foi de apresentar técnicas para os subtipos da tarefa T que foram expostas em sala de aula e fazer aplicação dessas técnicas com base em exercícios. O professor reafirma essa hipótese durante a entrevista, quando anuncia que,

> **Prof**: (...) Então, eles iam começar a fazer sem usar a fórmula de Bhaskara, usando apenas as propriedades, quando eles... quando não tinham o valor de "b", resolviam por radicais, e quando tinham o valor de "b" e não tinham o de "c", faziam por fatoração. A gente passou bastante tempo fazendo isso para

> eles reconhecerem melhor. Aí a gente, depois, entrou no trinômio quadrado perfeito, para eles identificarem quando elas forem completas, aí resolveram completando quadrados. Aí, depois que eles trabalharam nestas fases, aí eu entrei na equação do segundo grau, com a fórmula de Báskara. (...) A partir disso foi identificar a fórmula de Báskara saber o que é o valor de "a", de "b" e de "c", e resolver, parte mecânica mesmo, só substituindo, aí depois, nós fomos para a parte de problemas".

Como dito anteriormente, pedimos aos alunos que resolvessem uma lista de atividade a qual continha os sete subtipos de tarefas. Em posse da devolutiva dos alunos, pudemos identificar as técnicas utilizadas pelos alunos, assim como, as praxeologias utilizadas.

Quanto às técnicas utilizadas, foram identificadas as seguintes técnicas:

Nomenclatura da Técnica	**Sigla utilizada no CHIC**
Não Fez	NF
Bháskara	BHAS
Propriedade do Produto Nulo	PN
Transposição dos Termos (Conjunto dos Números Reais)	TT
Fatoração[12]	FAT
Radiciação	RAD
Tentativa	TENT

Quadro 1 – Técnicas utilizadas pelos alunos para resolução da Lista de Atividades

Para cada item da atividade foram elaboradas as possibilidades de cada evento, assim sendo, tivemos as seguintes variáveis: **V1NF** (Item 1 – Não Fez); **V1BHAS** (Item 1 – Bháskara); **V1PN**(Item 1 – Propriedade do Produto Nulo); **V1TT**(Item 1 – Transposição dos Termos); **V1FAT**(Item 1 – Fatoração); **V1RAD**(Item 1 – Radiciação); **V1TENT** (Item 1 – Tentativa); **V2NF** (Item 2 – Não Fez); **V2BHAS** (Item 2 – Bháskara); e assim sucessivamente nos sete itens que foram analisados.

Quanto às praxeologias, elaboramos duas possibilidades para cada item da Lista de Atividades, ou o aluno seguia a praxeologia que o professor apresenta em sala de aula, a qual denominamos Praxeologia do Professor, ou utilizou uma praxeologia própria, a qual iremos tratar como Praxeologia do Aluno. Assim sendo, temos as seguintes variáveis quanto às praxeologias: **PROF_T1** (Item 1 – Praxeologia do Professor); **ALUNO_T1** (Item 1 – Praxeologia do Aluno); **PROF_T2** (Item 2 – Praxeologia do Professor); **ALUNO_T2** (Item 2 – Praxeologia do Aluno); **PROF_T3** (Item 3 – Praxeologia do Professor); **ALUNO_T3** (Item 3 – Praxeologia do Aluno); **PROF_T4** (Item 4 – Praxeologia do Professor); **ALUNO_T4** (Item 4 – Praxeologia do Aluno); **PROF_T5** (Item 5 – Praxeologia do Professor); **ALUNO_T5** (Item 5 – Praxeologia do Aluno); **PROF_T6** (Item 6 – Praxeologia do Professor); **ALUNO_T6** (Item 6 – Praxeologia do Aluno); e **PROF_T7** (Item 7 – Praxeologia do Professor); **ALUNO_T7** (Item 7 – Praxeologia do Aluno).

[12] Estamos chamando de *Fatoração* a regra da soma e do produto das raízes da equação, e não o conceito matemático. Fizemos essa opção porque alguns professores se reportam a regra da soma e do produto como fatoração, assim como o professor dessa pesquisa também o fez.

Quanto às delimitações do uso do quadro metodológico da A.S.I., para a análise dos dados coletados na devolutiva dos alunos com o software CHIC, utilizamos como *Tipo de Implicação* segundo ***Método Entrópico*** (GRAS ET AL, 2009, 2013), por melhor satisfazer ao objetivo de modelagem da inclusão conjuntista, além de ser mais severa no diz respeito à intensidade de implicação. Quanto ao *Tipo de Lei* fizemos a opção pela ***Lei Binomial***, tendo em vista termos um número baixo para as tentativas (n=24 alunos). Caso tivéssemos uma amostra igual ou superior a 30 (n≥30), poderíamos usar a Lei de Poisson como uma aproximação da distribuição Binomial.

Faremos, a partir de agora, uma análise dos sete subtipos de tarefa que apareceram durante as aulas. Mais adiante, iremos apresentar uma síntese com as expectativas, diante as apresentações das técnicas pelo professor, de resolução para cada item da Lista de Atividades.

3.3 Subtipos de Tarefas

3.3.1 Subtipo de Tarefa T_1: $ax^2 + c = 0$

No primeiro subtipo de tarefa T1 foram reunidos os exercícios que levavam os alunos a resolver equações de segundo grau apresentadas na forma $ax^2 + c = 0$. Parece-nos que a intenção do professor, nesse subtipo de tarefa, era de aproximar ao máximo das resoluções de equações de primeiro grau.

Assim sendo, para realizar esse subtipo de tarefa, o professor começa fazendo uma breve explanação sobre os graus da equação. Para isso, escreve no quadro a seguinte equação: $x^3 + y^x + 12 + 10 = 0$, e faz o seguinte comentário: "(...) se for em relação a 'x' a equação será do segundo grau, em relação a 'y' será de grau 'x'". Em seguida, o professor utiliza um exercício do livro adotado pela escola: $3x^2 - 75a^4 = 0$. Durante a resolução, o professor isolou o $3x^2$, transpondo o termo $75a^4$ invertendo as operações. Essa técnica, transpor os termos e inverter as operações, é a escolhida pelo professor como a técnica principal para a resolução de exercícios desse subtipo de tarefa T_1: $ax^2 + c = 0$. Após a transposição do termo $75a^4$, o professor passou o número três dividindo e reduziu a expressão para $x^2 = 25a^4$, nesse momento fez uma referência ao conteúdo de radicais: "(...) estão lembrados do que trabalhamos com radicais? Quando passamos o 2 (índice) para o outro lado... teremos a raiz! (os alunos repetiram junto com o professor)" extraiu a raiz e encontrou o valor da incógnita $5a^2$. É importante salientar que o professor não considera a raiz negativa $(-5a^2)$.

A subtécnica adotada pelo professor, para a resolução desse subtipo de tarefa, foi desenvolver ou reduzir expressões, técnicas que os alunos aprenderam para resolver equações de primeiro grau. Pudemos identificar, também, apesar de não serem explicitados, os elementos tecnológicos que deram a suporte à técnica. Foram eles: as propriedades das operações inversas em R (conjunto dos números reais) ou leis de transposição de termos e as propriedades da radiciação.

De certa forma, o professor pretendia começar por um subtipo mais próximo ao que os alunos já tinham trabalhado (equações de primeiro grau) até chegar a um subtipo que não seria mais possível, com as técnicas "até agora conhecidas", chegar a uma solução

da equação. No entanto, o professor não deixa claro que estava buscando uma "evolução das técnicas". A reação dos alunos era de passividade, repetiam uníssonos as "deixas" do professor, observando cada passo na resolução do exercício feito no quadro.

3.3.2 Subtipo de tarefa T_2: $ax^2 + bx = 0$

Neste subtipo de tarefa o professor utiliza como técnica principal, ou primária, a fatoração, deixando a técnica do produto do produto nulo e de transposição dos termos como subtécnicas (auxiliares ou secundárias). Os elementos tecnológicos, novamente, não foram evidenciados pelo professor. Porém, pudemos identificar as propriedades distributiva da multiplicação, do produto nulo e das operações inversas em R (conjunto dos números reais) ou leis da transposição de termos, que serviram para dar uma sustentação inteligível a técnica e as subtécnicas utilizadas nesse subtipo de tarefa.

O exercício utilizado pelo professor para esse subtipo foi a equação $3x^2 - 24x = 0$. O diálogo que ocorre na sala de aula é o seguinte:

> **Prof**: *Como é que iremos resolver essa?* (apontando para a equação escrita no quadro)
> Os alunos começam a conversar entre si e o professor aguarda o silêncio da turma. Quando o silêncio é estabelecido, o professor se volta para a turma e indaga:
> **Prof**: *E aí?* (alguns alunos tentam fornecer algumas ideias, porém falam todos juntos sem que haja uma sincronia. Então o professor diz) *Vamos lá! Isolo o 'x' e tenho: 'x' vezes* $(3x - 24)$ *igual a zero (fatoração - técnica principal). Quando tenho um produto de dois números que dá zero, então é porque um deles é zero ou os dois são zero. (produto nulo - subtécnica)".*

O professor se volta para o quadro e iguala os termos a zero desenvolvendo as expressões, encontrando as raízes zero e oito. A partir da utilização da técnica de fatoração, por parte do professor, os alunos permanecem em silêncio acompanhando todo o desenvolvimento da resolução do exercício.

3.3.3 Subtipo de tarefa T_3: $(ax + c)^2$

O subtipo T3 apresenta uma pequena variação dos subtipos anteriores. Para sua resolução o professor extraiu a raiz de zero, isolou a incógnita e inverteu as operações. Para este subtipo, o professor não considerou a possibilidade de resolver a partir da propriedade da potenciação $(a^2 = a.a)$ e, em seguida, utilizar a subtécnica do produto nulo, utilizada anteriormente. Caso os alunos desenvolvessem o produto de $(ax + c).(ax + c) = 0$, aplicando a propriedade distributiva da multiplicação, poderia causar um bloqueio, pois eles ainda não possuíam técnicas que dessem conta, visto que, os alunos, teoricamente, ainda não foram apresentados à fórmula de Bhaskara. Um outro elemento que não foi considerado pelo professor, foi ressaltar a existência de duas raízes, apesar de serem iguais.

E, por fim, o professor não demonstrou que a técnica aplicada também "funciona" quando a equação é igualada a outro número qualquer diferente de zero, tendo em vista que em todos os exercícios apresentados, neste subtipo de tarefa, a expressão aparece igualada a zero. Podemos analisar este fato com um elemento indicador de transposição

didática interna, já que no livro utilizado, esse subtipo de tarefa, apresenta equações igualadas a números diferentes de zero. Acreditamos que a estratégia do professor era, de certa forma, apresentar as equações em condições, teoricamente, mais simples para a resolução, focando, principalmente, na aplicação da técnica principal.

Neste subtipo de tarefa o professor utilizou o seguinte exercício do Livro: $(z + 5)^2 = 0$. O professor inicia a resolução perguntando a turma:

> **Prof**: *Como é que iremos resolver essa equação aqui?* (Aguarda durante um tempo alguma resposta dos alunos, que estão agitados, com vários falando ao mesmo tempo, até que ele continua)
> **Prof**: *Eu posso fazer desse jeito aqui!* (a turma faz silêncio e ouve o professor) *'z' mais cinco igual... Se é potencia passa pra lá como?.*
> **Alunos**: *Raiz*! (O professor se volta para o quadro e escreve o sinal da raiz no zero. É nesse momento que o professor apresenta a técnica principal, a extração da raiz quadrada. Após, ele volta a falar)
> **Prof**: *z mais cinco é igual... Quanto é a raiz de zero?*
> **Alunos**: *Zero*! (O professor termina de resolver a equação dizendo que a raiz é cinco, para isso, utilizou como subtécnica a transposição dos termos, invertendo as operações. Como foi dito anteriormente, o professor não fez nenhuma alusão às raízes iguais).

Em nenhum dos subtipos de tarefa os elementos tecnológicos são evidenciados de forma clara. Em um momento, ou outro, se faz uma explicação da técnica ou subtécnica utilizada, mas sem explicitar a tecnologia e a teoria que, de certa forma, deram uma racionalidade a essas técnicas e subtécnicas. Porém, em nossa tese, buscamos identificar, em todos os subtipos de tarefa, os elementos tecnológicos que compõem o bloco teórico-tecnológico. Para esse subtipo de tarefa, as propriedades da radiciação e das operações inversas em R (conjunto dos números reais) ou leis da transposição de termos, como elementos que constituem esse bloco teórico-tecnológico.

3.3.4 Subtipo de tarefa T_4: $(x + a).(x + b) = 0$

O subtipo de tarefa T4, diferentemente de T3, gera duas raízes distintas. Nesse subtipo de tarefa, a técnica utilizada é a do produto nulo que ganha um status, nesse subtipo de tarefa, de técnica principal ou primária, deixando para a transposição de termos e desenvolvimento da expressão o papel de auxiliares ou secundárias. Os elementos tecnológicos observados foram as propriedades do produto nulo e as operações inversas em R (conjunto dos números reais) ou leis da transposição de termos.

A equação $(x - 17).(x + 11) = 0$ é utilizada como exemplo para este subtipo de tarefa. Ao iniciar a resolução o professor relembra aos alunos que a técnica do produto nulo já foi aplicada, afirmando que: "Na aula passada, nós já fizemos algumas questões parecidas com essa, que era o seguinte: Eu tenho aqui dois números que quando eu multiplico dá zero, e aí?", os alunos respondem: "Um deles é zero!".

As questões parecidas, mencionadas pelo professor, são as do subtipo de tarefa T_2 $(ax^2 + bx = 0)$, nas quais ele, como descrito anteriormente, utiliza-se da fatoração como técnica principal (colocando em evidência o termo comum "x"), para depois

aplicar as subtécnicas do produto nulo (igualando os termos do produto a zero), da transposição dos termos e do desenvolvimento das expressões.

Continuando a resolução, o professor avança: "É isso aí! Um dos dois vai ser zero!". O professor começa a escrever no quadro igualando os termos a zero, encontrando as duas raízes da equação: 17 e -11. Após a transposição dos termos e o desenvolvimento das expressões, o professor alerta para a existência das duas raízes dizendo: "Então, as raízes dessa equação será 17 e -11".

Assim como nos subtipos de tarefas anteriores, os alunos se portam de modo indiferente, observando a resolução do professor e respondendo juntos as indagações do professor.

3.3.5 Subtipo de tarefa $\mathbf{T_5}: (\mathbf{x}+\mathbf{a}).(\mathbf{x}+\mathbf{b}) = \mathbf{cx}+\mathbf{d}$

Da mesma forma que foi feito no subtipo de tarefa anterior, no subtipo T_5, o professor utiliza subtécnicas preliminares para chegar a uma equação de 2º grau, são elas, desenvolvimento ou redução de expressões e a transposição de termos invertendo as operações, chegando a um subtipo de tarefa já conhecido pelos alunos T_1: $(ax^2 + b = 0)$. Com isso, tivemos como técnica principal uma das subtécnicas preliminares, a de desenvolvimento ou redução de expressões. Ou seja, a técnica mudou de status no desenvolvimento da resolução do subtipo de tarefa, devido a ser novamente utilizada durante a atividade.

Para esse subtipo de tarefa, identificamos como elementos tecnológicos as propriedades distributiva da multiplicação, das operações inversas em R (conjunto dos números reais) ou leis da transposição de termos e da radiciação.

Para esta resolução, é necessário que o valor de c seja igual à soma de a e b $(c = a + b)$, para que possa eliminar o termo que é multiplicado por x. E assim, voltar a um modelo T_1: $(ax^2 + b = 0)$. Novamente, a equação é preparada para se chegar a uma equação incompleta. Assim, a partir do desenvolvimento da multiplicação entre $(x + a).(x + b)$, da transposição e da redução dos termos, além da inversão das operações, o professor chegará ao valor da incógnita.

A equação utilizada pelo professor foi $(2x + 1).(x + 3) = 7x + 11$. Mais uma vez, o professor aguarda o silêncio da turma para começar a resolver a equação. Assim que é feito o silêncio, o professor pergunta:

> **Prof**.: *E aí? Como vamos resolver essa?* (Os alunos ficam em silêncio, sem darem nenhuma opinião. Então, o professor começa a resolver)
> **Prof**.: *Multiplica isso daqui* (apontando para o primeiro termo da equação), *então vamos lá.* (O professor utiliza a propriedade distributiva da multiplicação, passa os termos do segundo membro $(7x + 11)$ para o primeiro membro mudando o sinal, verbalizando)
> **Prof**.: *Eu já posso passar esses pra cá* (apontando para o segundo termo da equação. Em seguida, ele reduz os termos semelhantes) *Vamos agora agrupar e cortar os iguais,* e continua *e o que temos agora?* $2x^2 - 8 = 0$. (A partir desse momento o professor desenvolve a equação de modo similar ao subtipo de tarefa T1, no qual aplica a técnica de transposição dos termos e desenvolvimento de expressões).

Nesse subtipo de tarefa, o professor, novamente, não considerou a raiz negativa da equação, tratando somente da raiz positiva.

3.3.6 Subtipo de tarefa $T_7: \frac{(ax+b)}{c} + dx^2 = ex + f$

O subtipo de tarefa T_7 demanda a determinação do MMC, o desenvolvimento de expressões e a transposição de termos como subtécnicas preliminares. O professor não considera a técnica da simplificação ou a multiplicação de todos os termos pelo mesmo número para a resolução desse subtipo de tarefa.

Assim como ocorreu no subtipo de tarefa T_6, uma subtécnica muda de status passando para técnica principal. Novamente, a equação é preparada para se chegar em uma equação incompleta, um modelo já conhecido pelos alunos, o $T_1: (ax^2 + b = 0)$, após a utilização das subtécnicas preliminares.

Quanto aos elementos tecnológicos identificamos as seguintes propriedades relativas aos números racionais, as operações inversas em R (conjunto dos números reais) ou leis da transposição de termos e a da radiciação.

O professor utiliza a equação $\frac{(2x+3)}{2} - x^2 = x + 6$ para iniciar esse subtipo de tarefa. O professor começa a resolver a equação chamando a atenção dos alunos

> **Prof**.: *Silêncio! Essa equação aqui, como é que eu começo a resolver? Quando eu tenho essa situação* (ele está falando do denominador) *o que é que eu tenho que fazer? Vou tirar o MMC gente! Toda vez que eu tiver fração... Silêncio! Toda vez que eu tiver fração, eu tenho que tirar o MMC!* (Em seguida, ele se volta para o quadro e começa a narrar todos os passos que está fazendo)
> **Prof**.: *Eu retirei o MMC e agora vou colocar esses termos para o outro lado..., até chegar à equação* $-2x^2 - 9 = 0$. (A partir desse momento, ele desenvolve conforme fez no subtipo de tarefa $T_1: (ax^2 + b = 0)$, transpondo termos e desenvolvendo e reduzindo expressões, sempre voltado para o quadro narrando todas as etapas que são efetuadas).

No final, o professor encontra uma raiz que não existe para o conjunto dos números reais $\sqrt{\frac{-9}{2}}$. O professor pergunta a turma:

> **Prof**.: *Quanto é que é essa raiz aqui?*
> **Alunos**: *É a raiz de quatro vírgula cinco!*
> **Prof**.: *Eu posso achar a raiz desse número aqui? Não! Vocês lembram que quando o índice aqui for par, não pode ser negativo aqui dentro! Então, isso aqui não tem raiz nos reais. Certo?* (Em seguida, ele escreve no quadro) *Não existe raiz nos reais.*

3.3.7 Subtipo de Tarefa $T_7: ax^2 + bx + c = 0$

O subtipo T_8 corresponde às atividades agrupadas com a seguinte forma: $ax^2 + bx + c = 0$. Esse subtipo já vai apresentar a equação de segundo grau

completa e, com isso, o professor começará utilizando como técnica principal ou primária a resolução completando quadrados, até, enfim, introduzir outra técnica que irá substituí-la, a fórmula da Bhaskara. Essa técnica já era uma expectativa de alguns alunos, pois, como afirmou o professor durante a entrevista, alguns já a conheciam. Vale a pena salientar que nenhum dos alunos utilizou a técnica de completar quadrados durante a resolução da lista de exercícios, como iremos ver posteriormente.

A estratégia elaborada pelo professor para mostrar "todos os tipos de formas" de resolução (técnicas) das equações de segundo grau, foi de começar apresentando equações incompletas, passando por casos particulares de equações completas (completar quadrados) culminando no *grand finale* com uma fórmula para todos os tipos de tarefas: a fórmula de Bhaskara. Como pudemos observar durante seu discurso em sala de aula:

> **Prof**.: *Gente! Pronto! É o seguinte, a gente antes estava resolvendo as equações de segundo grau de várias maneiras... Essa daqui foi criada* (apontando para a fórmula de Bhaskara que estava escrita no quadro) *para resolver todos os tipos de equação de segundo grau.*

Após analisarmos todos os subtipos de tarefas, pudemos verificar, durante a resolução dos alunos, que, apesar de o professor cumprir exatamente o que programou para as suas aulas, essa organização matemática levou-os a uma possível dificuldade. Nas tentativas de encontrar as soluções dos exercícios, os alunos, normalmente, encontravam uma única raiz para equação. Acreditamos que esse comportamento dos alunos foi devido à associação feita com as equações de primeiro grau, que apresentam uma única raiz. Outro fator que pode ter contribuído para esse comportamento, foi que o professor, em algumas resoluções, como evidenciamos durante a análise dos subtipos de tarefa, não alertou para a existência das duas raízes.

Outro detalhe que observamos durante as aulas, foi que o professor não apresentava os subtipos de tarefas como novos desafios, de forma que com as técnicas até então "conhecidas", não conseguiríamos realizar a nova atividade proposta, ou seja, deveríamos buscar novas técnicas que iriam superar a(s) anterior(es). Tal fato é importante para que o aluno não fique desestimulado e desista de realizar as tarefas propostas, bem como não fique parado, aguardando que o professor apresente a nova técnica. Sobre essa evolução de técnicas, Chevallard (1999) afirma que quando uma "maneira de fazer" tem êxito somente sobre uma parte de um tipo de tarefa T, a qual é relativa, essa parte se denomina alcance da técnica. Essa técnica será "substituída" por outra que dê conta, se não sobre toda tarefa T, ao menos sobre parte dela. Nessa visão, podemos dizer que uma técnica é superior à outra.

3.4 Expectativas de resolução da Lista de Atividades

Foi elaborado um quadro (2) em que são apresentadas as expectativas de resolução de cada item da Lista de Atividades[13]. Essas Expectativas foram elaboradas diante a apresentação das técnicas durante as aulas do conteúdo de equações de 2º grau, e servirá para identificar se o aluno seguiu ou não as técnicas apresentadas pelo professor o que irá caracterizar sua praxeologia.

[13] Os itens 'c' e 'h' foram descartados para análise, por erro de digitação durante a confecção da Ficha de Atividades. Com isso, os itens não possuem raízes reais, o que não era objeto de estudo da pesquisa.

Variável	Item da Lista	Técnica esperada
1	a) $2x^2 - 98a^6 = 0$	**RAD** – Radiciação
2	b) $2X^2 - 8x = 0$	**FAT** – Fatoração
3	d) $x^2 - x - 6 = 0$	**BHAS** – Bháskara
4	e) $(x - 8).(x + 4) = 0$	**PN** – Produto Nulo
5	f) $(2x + 5)^2 = 0$	**RAD** – Radiciação
6	g) $(x + 4).(x - 7) = 5x + 5$	**TT** – Transposição dos Termos
7	i) $\frac{6x + 10}{2} - 3x^2 = 3x - 103$	**TT** – Transposição dos Termos

Quadro 2 – Expectativas de resolução

Será a partir dessas expectativas que veremos se ocorre ou não uma uniformidade na ação do aluno durante a resolução de atividades em sala de aula. Ou seja, se esses alunos indicam ter uma preferência por determinada técnica ou se seguem as técnicas apresentadas pelo professor. Diante das análises dos gráficos fornecidos polo CHIC e com os dados que foram coletados durante o estudo da tese que serve de base para essa pesquisa, poderemos avançar nas análises apresentadas na conclusão da tese, e levantar novas hipóteses sobre o que aconteceu em sala de aula.

4. Análise com o software CHIC

Para iniciarmos nossas análises com o software CHIC, elaboramos duas tabelas com os dados coletados. Uma denominada Técnica x Aluno (**Tabela 1**), em que identificamos quais as técnicas que os alunos utilizaram em cada um dos itens da Lista de Atividades, e outra denominada Praxeologia x aluno (**Tabela 2**), em que identificamos se o aluno acompanhou ou não a técnica do professor em cada um dos itens da atividade.

	V1NF	V1TT	V1RAD	V2NF	V2BMAB	V2TT	V2TENT	V3NF	V3BMAB	V3TT	V3TENT	V4NF	V4BMAB	V4PN	V4TT	V5NF	V5BMAB	V5PN	V5TT	V5RAD	V6NF	V6BMAB	V6FAT	V6TENT	V7NF	V7BMAB	V7RAD	V7TENT
S01	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
S02	1.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S03	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
S04	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S05	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00
S06	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S07	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S08	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
S09	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S10	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S11	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
S12	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S13	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00
S14	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S15	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S16	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S17	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
S18	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S19	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
S20	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S21	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S22	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00
S23	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00
S24	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00

Tabela 1 – Técnicas - Alunos

	PROF_T1_	ALUNO_T1_	PROF_T2_	ALUNO_T2_	PROF_T3_	ALUNO_T3_	PROF_T4_	ALUNO_T4_	PROF_T5_	ALUNO_T5_	PROF_T6_	ALUNO_T6_	PROF_T7_	ALUNO_T7_
S01	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	1.00	0.00
S02	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S03	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00
S04	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S05	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	1.00	0.00
S06	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00
S08	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00
S09	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S10	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S11	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00
S12	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S13	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
S14	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S15	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S16	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S17	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00
S18	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S19	1.00	0.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00
S20	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00
S21	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S22	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
S23	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
S24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00

Tabela 2 - Praxeologias

Como já foi explicitado anteriormente, quanto ao uso do quadro metodológico da ASI, para a análise dos dados coletados com o software Chic, utilizamos como *Tipo de Implicação* a **Implicação segundo ao método entrópico** e o *Tipo de Lei* foi a **Lei Binomial**. Para a elaboração dos gráficos implicativos utilizamos um nível de

implicação de no mínimo de 0,80, por considerarmos um nível alto para análise, e que nos daria *quase-implicações* com grau altamente significativos.

4.1 Análise das técnicas utilizadas pelos alunos – Tabela 1

Ao analisarmos os gráficos apresentados pelo CHIC, podemos perceber que há uma implicação no uso das mesmas técnicas para resolver determinados itens da Lista de Atividades. Como já foi informado, para a construção do gráfico, utilizamos um nível mínimo de 0,80; o que nos permite apresentar uma estrutura extremamente significativa do ponto de vista da estatística. As implicações com a cor azul até 0.95; a cor verde até 0,90; e a cinza até 0,80.

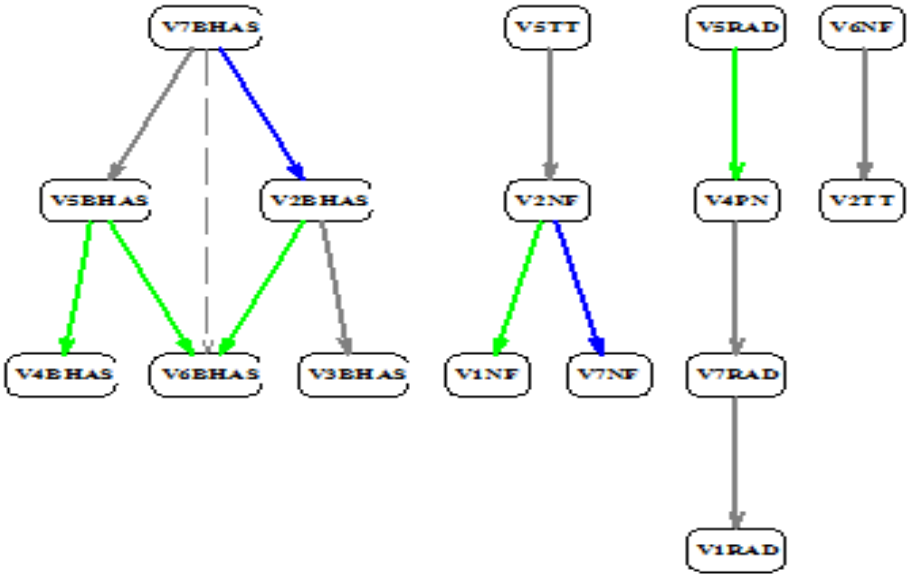


Gráfico 1 – Gráfico implicativo das técnicas x aluno

De um modo geral as variáveis de mesma técnica possuem uma forte implicação como podemos perceber no Gráfico 1, gerando o que chamaremos de classes apresentadas no quadro a seguir.

Classes	Técnicas Implicadas
1ª	V2BHAS – V3BHAS – V4BHAS – V5BHAS – V6BHAS – V7BHAS
2ª	V5TT – V1NF – V2NF – V7NF
3ª	V1RAD – V5RAD – V7RAD – V4PN
4ª	V6NF – V2TT

Quadro 3 – Classes geradas pelo Gráfico 1

No que denominamos de *classe 1*, percebemos uma forte implicação entre a técnicas BHAS. Esse fato reforça a hipótese que apresentamos na tese (BESSA DE MENEZES, 2010), em que os dados nos levam a crer que técnica de Bhaskara ocorra, principalmente, pelo status que a fórmula possui quanto à resolução de equações de 2º grau. Esse status dá a Bhaskara uma garantia na resolução dessas equações, o que coloca o aluno em conformidade com a Instituição Matemática, na qual, historicamente, é valorizado o resultado encontrado.

Ainda refletindo sobre a hipótese, a necessidade de conformidade, por parte dos alunos, com a Instituição Matemática, faz com que eles, apesar de as técnicas propostas pelo professor darem a impressão de serem "*mais simples*" de aplicá-las na resolução das tarefas, não percorram esse caminho, ou seja, nos dão indícios de que eles querem a "*segurança*" que Bhaskara fornece. Os resultados encontrados no tratamento dos dados no CHIC nos reforçam essa ideia.

Na *classe 2*, percebemos a implicação entre os alunos que não realizaram alguns itens (NF), esse fato já era esperado devido os itens apresentarem ou equações incompletas do 2º grau (Itens 1 e 2) ou em forma pouco usual como no item 7. A diferente apresentação pode ter levado aos alunos a não realizarem a tarefa.

Quanto à implicação no item 5 (transposição de termos – V5TT), com os alunos que não realizaram os itens 1, 2 e 7; podemos levantar a hipótese que isso ocorra devido o professor iniciar o conteúdo de equações do 2º grau, a partir dos conhecimentos dos alunos em equações do 1º grau, e no caso particular do item 5, não há o expoente 2 em nenhum valor de '*x*' [$(2x+5)^2 = 0$]. Essa situação nos leva a supor que o aluno entendeu como uma equação de grau 1 e, por isso, se sentiu com possibilidades de resolvê-la, o que não ocorre nos itens 1, 2 e 7, em que se observa a presença do expoente 2 para a incógnita '*x*'.

A *classe 3* apresenta resoluções a partir de técnicas que foram apresentadas pelo professor em sala de aula, o que nos leva a identificar um alinhamento do que foi estabelecido como "*modelo de resolução*" pelo professor, pelos alunos dessa classe.

Quanto à *classe 4*, verificamos que ao desenvolver a equação do item 6, o aluno se depara com uma completa do 2º grau, já no item 2, ele encontra uma incompleta. Tal fato, nos leva ao indício de que os alunos dessa classe, não conseguem resolver uma equação completa, somente a incompleta.

O árvore de similaridade, gerada pelo CHIC, reforça a ideia de um agrupamento no uso de técnicas, como podemos perceber no gráfico 3 e no quadro de similaridade quadro 3). Nele, podemos perceber que os 'galhos' da árvore possuem uma tendência de ter as mesmas técnicas.

Tal fato nos reforça a ideia de que o aluno sempre busca estar em conformidade com a Instituição Matemática, pois utiliza de uma técnica que provavelmente tem confiança ou uma '*boa relação ao saber*'[14] para resolver as tarefas (itens) que são propostos pelo professor em sala de aula.

Essa relação é, em alguns casos, determinante para a escolha de técnicas para resolução de tarefas no dia a dia. Vejamos o gráfico a seguir.

[14] Entendemos como 'boa relação ao saber' a proximidade do indivíduo (I) ao objeto do saber (O). Quanto maior proximidade com o saber, melhor sua relação.

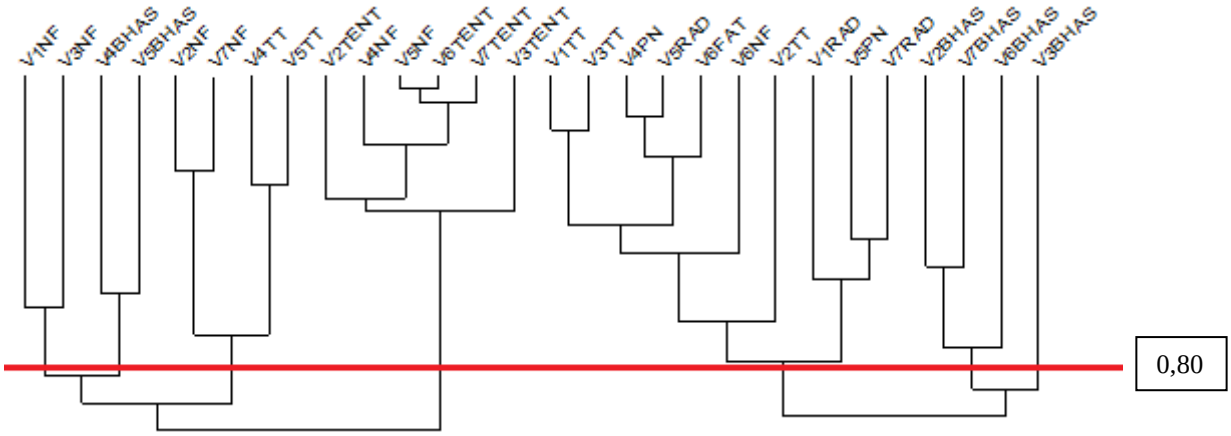


Gráfico 2 – Gráfico de similaridade das técnicas utilizadas pelos alunos

Classificação ao nível: 1 : (V5NF V6TENT) similaridade : 0.999999
Classificação ao nível: 2 : ((V5NF V6TENT) V7TENT) similaridade : 0.999999
Classificação ao nível: 3 : (V4PN V5RAD) similaridade : 0.999898
Classificação ao nível: 4 : (V1TT V3TT) similaridade : 0.999626
Classificação ao nível: 5 : (V4NF ((V5NF V6TENT) V7TENT)) similaridade : 0.998878
Classificação ao nível: 6 : ((V4PN V5RAD) V6FAT) similaridade : 0.993336
Classificação ao nível: 7 : (V2NF V7NF) similaridade : 0.993263
Classificação ao nível: 8 : (V4TT V5TT) similaridade : 0.98964
Classificação ao nível: 9 : (V2TENT (V4NF ((V5NF V6TENT) V7TENT))) similaridade : 0.986716
Classificação ao nível: 10 : ((V2TENT (V4NF ((V5NF V6TENT) V7TENT))) V3TENT) similaridade : 0.983423
Classificação ao nível: 11 : ((V1TT V3TT) ((V4PN V5RAD) V6FAT)) similaridade : 0.98014
Classificação ao nível: 12 : (V5PN V7RAD) similaridade : 0.979072
Classificação ao nível: 13 : (((V1TT V3TT) ((V4PN V5RAD) V6FAT)) V6NF) similaridade : 0.949261
Classificação ao nível: 14 : (V2BHAS V7BHAS) similaridade : 0.905895
Classificação ao nível: 15 : (V1RAD (V5PN V7RAD)) similaridade : 0.905169
Classificação ao nível: 16 : (V4BHAS V5BHAS) similaridade : 0.891594
Classificação ao nível: 17 : (V1NF V3NF) similaridade : 0.891259
Classificação ao nível: 18 : ((((V1TT V3TT) ((V4PN V5RAD) V6FAT)) V6NF) V2TT) similaridade : 0.880819
Classificação ao nível: 19 : ((V2NF V7NF) (V4TT V5TT)) similaridade : 0.870849
Classificação ao nível: 20 : ((V2BHAS V7BHAS) V6BHAS) similaridade : 0.819345
Classificação ao nível: 21 : (((((V1TT V3TT) ((V4PN V5RAD) V6FAT)) V6NF) V2TT) (V1RAD (V5PN V7RAD))) similaridade : 0.803561

Quadro 4 – Classificação ao nível de similaridade das técnicas utilizadas pelos alunos

Outro ponto de análise de nossa pesquisa envolve as praxeologias que pudemos observar na tese, que serve de base para essa pesquisa, explicitada na devolutiva dos alunos na resolução da Lista de Atividades. Foram identificadas as praxeologias do Professor e dos Alunos.

1.1 4.2 Análise das "*praxeologias*"[15] do Professor e do Aluno – Tabela 2

Ao analisarmos os gráficos gerados a partir do CHIC para a tabela 2, em que buscamos identificar se o aluno segue ou não a praxeologias do Professor, identificamos uma linearidade, ou seja, aqueles que optam por seguir a práxis do professor, tendem a acompanhá-lo na resolução dos itens. No entanto, aqueles que optam por um outro caminho (ou outra práxis), também, tendem a fazê-lo na resolução dos itens. Assim como trabalhamos na construção dos gráficos da Tabela 1, utilizamos um nível mínimo

[15] O que iremos observar na realidade são as técnicas que o professor utilizou e quais os alunos utilizaram, o que podem caracterizar uma diferença de praxeologias entre o professor e os alunos.

de 0,80; o que nos permite apresentar uma estrutura extremamente significativa do ponto de vista da estatística. As implicações com a cor azul até 0.95; a cor verde até 0,90; e a cinza até 0,80.

Vejamos o gráfico implicativo da tabela 1, a seguir.

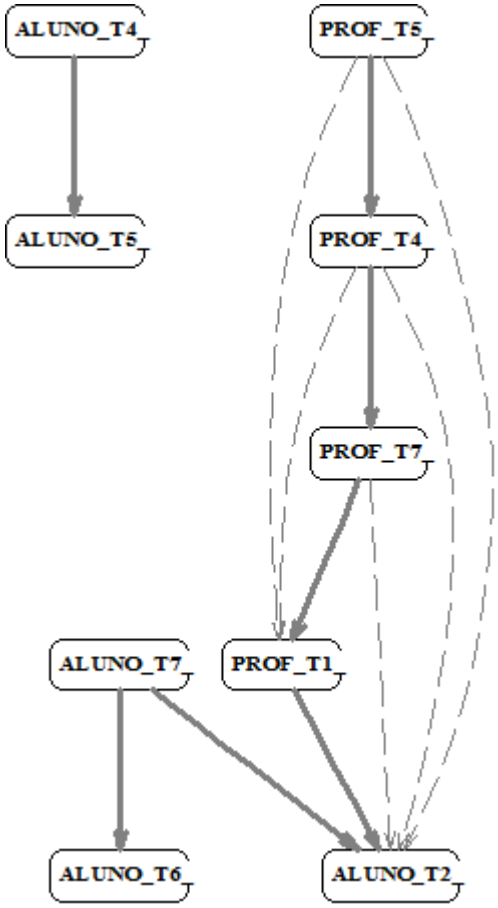


Gráfico 3 – Gráfico implicativo das praxeologias x aluno

De forma natural existe uma expectativa que o aluno siga a técnica de resolução do professor, pois existe uma regra de contrato didático implícito de que "*se fizer como o professor faz está certo*". O gráfico 3 nos dá uma confirmação dessa expectativa. Os itens 1, 4, 5 e 7, não estão apresentados na forma reduzida de uma equação de 2º grau completa ($ax^2 + bx + c = 0$), com isso, devido o professor ter iniciado esse conteúdo a partir de equações incompletas fazendo relação com o assunto trabalhado nas equações do 1º grau, acreditamos que os alunos se sentiram mais seguros em resolver com as técnicas apresentadas pelo professor. Um fato que reforça nossa hipótese é de que a maioria dos alunos encontra somente uma raiz na resolução dos itens em que as equações não estão completas, o que pode estar relacionado diretamente com um efeito da transposição do saber que foi realizada em sala de aula, um possível efeito de transposição.

A implicação que ocorre entre as técnicas do professor (PROF) nos itens 1, 4, 5 e 7 e a técnica ALUNO_T2, pode ser explicada a partir da escolha dos alunos pela transposição de termos (TT), que os aproxima das técnicas utilizadas para resolução de equações de grau 1; como também, pela escolha da fórmula de Bhaskara (BHAS), a

qual já identificamos na tabela 1 como tendo um status muito forte para resolução de equações de grau 2.

Podemos notar que as implicações entre as variáveis ALUNO_T4 e ALUNO_T5 ocorrem devido à escolha dos alunos em realizar as questões T4$(x - 8).(x + 4) = 0$ e T5 $(2x + 5)^2 = 0$, pela fórmula de Bhaskara. Assim sendo, eles desenvolvem o produto entre os fatores e aplicam a fórmula. Novamente, pensamos na ideia de se estar adequado à instituição de ensino ou ao professor por parte do aluno.

O gráfico e a classificação de similaridade nos apontam para o mesmo caminho, ou seja, a uma regularidade nas escolhas das técnicas para a resolução dos itens. Considerando o índice de 0,80 (estatisticamente mais significativo), temos três classes (nível 1, 2 e 3) em todas temos um agrupamento segundo a similaridade das técnicas do professor com professor e do aluno com aluno.

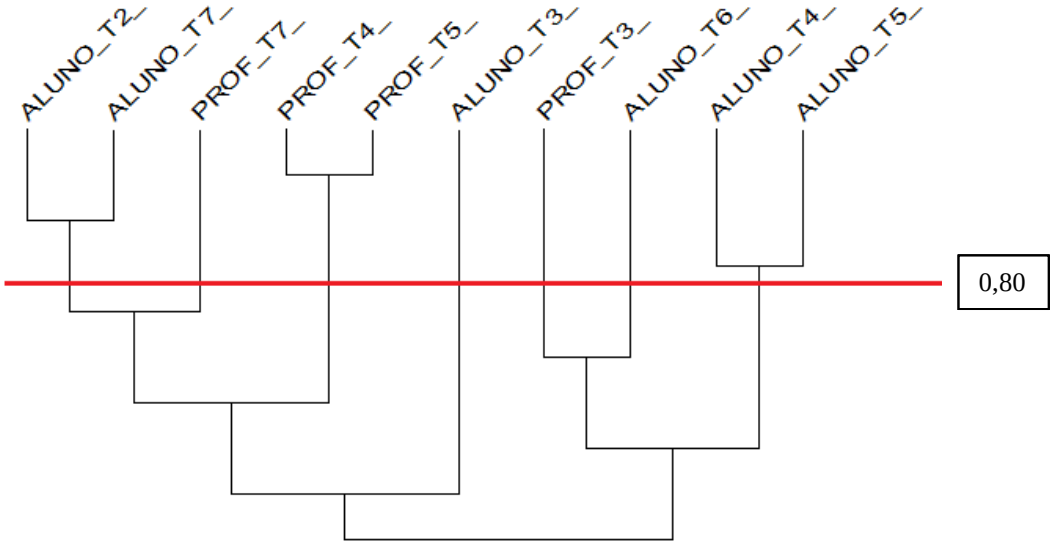


Gráfico 4 – Gráfico de Similaridade das Praxeologias

Classificação ao nível: 1 : (PROF_T4_ PROF_T5_) similaridade : 0.999898 Classificação ao nível: 2 : (ALUNO_T2_ ALUNO_T7_) similaridade : 0.898199 Classificação ao nível: 3 : (ALUNO_T4_ ALUNO_T5_) similaridade : 0.883579 Classificação ao nível: 4 : ((ALUNO_T2_ ALUNO_T7_) PROF_T7_) similaridade : 0.740698 Classificação ao nível: 5 : (PROF_T3_ ALUNO_T6_) similaridade : 0.707813

Quadro 4 - Classificação ao nível de similaridade das Praxeologias

Após a análise dos gráficos gerados a partir do CHIC, das tabelas 1 e 2, buscaremos fazer uma retomada da pesquisa e apontar algumas considerações que foram possíveis a partir do uso da A.S.I.

5 Considerações Finais

Nossa pesquisa tinha a como objetivo avançar em um ponto que não foi observado durante a realização da nossa tese que tratou das diferentes praxeologias apresentadas pelo professor e pelos alunos durante o ensino das equações de segundo grau, que seria a estabilidade na escolha de técnicas na resolução de tarefas propostas em sala de aula. Para isso, recorremos ao uso do quadro metodológico da A.S.I.

A partir da análise dos dados percebemos que há uma estabilidade nas opções dos alunos para resolverem tarefas propostas, seja para seguir as técnicas apresentadas pelo professor, seja para *caminhar com seus próprios passos*.

Quanto a *caminhar com seus próprios passos*, o gráfico de implicação e de similaridade da Tabela 1, gerados segundo o software CHIC, nos aponta para essa estabilidade, e nos permite reforçar algumas hipóteses que foram levantadas durante a tese, porém, não tínhamos elementos para afirmá-las. Pudemos perceber uma implicação significativa na escolha da técnica de Bhaskara (BHAS) para a resolução de atividades que poderiam ser solucionadas por técnicas mais simples com um número menor de passos para se chegar ao resultado. Esse fato pode estar ligado ao fato da busca pela conformidade.

Chevallard (1998) trata o aluno como um sujeito em cada instituição (família, escola, bairro, trabalho...) que ele participa em sua vida. Sujeito no sentido de se sujeitar, de estar adequado, de estar em conformidade, com essas instituições. A ação de se sujeitar à instituição poderá ser de forma indiferente ou ativa. Na forma indiferente, o sujeito irá se adequar às normas e regras da instituição sem contestá-las. No entanto, quando se sujeita de forma ativa, ele se transformará e será transformado pela instituição. O aluno, como sujeito de várias instituições, carrega em si elementos de cada uma dessas relações institucionais, as quais irão transformá-lo na pessoa que é. Com isso, a partir dessas relações, o aluno reconstrói o conhecimento para si, dando a ele características particulares, que ele traz das diversas instituições a que pertence.

Essa conformidade está presente, no momento em que os alunos buscam estratégias que os mantenham seguros quanto à resolução de um tipo de tarefa. Podemos pensar que os alunos não queriam errar, e simplesmente ter a certeza de que resolveriam a questão, e, com isso, se manterem em conformidade com as instituições escolares. O próprio livro didático, utilizado pelos alunos, valoriza a técnica de Bhaskara, ao citá-la como “Essa célebre fórmula aplica-se a todas as equações...”.

Quanto a *acompanhar as técnicas utilizadas pelo professor*, também remonta o que foi dito anteriormente sobre a conformidade. Foi possível perceber apoiado pelo gráfico de implicação e similaridade da Tabela 2. No entanto, sobre outro olhar: o do contrato didático que é estabelecido.

O professor, de certa maneira, impõe aos sujeitos (alunos) formas de fazer e de pensar, que são próprias a cada professor. Quanto a estas formas de fazer e de pensar, muitos pesquisadores já avançaram a partir do momento em que perceberam que o professor “cria” um metatexto do saber a ser ensinado em sala de aula (BESSA DE MENEZES, 2004; CÂMARA DOS SANTOS, 2002; BRITO DE MENEZES, 2006). Com isso, podemos, sim, entender o professor como uma Instituição durante sua atividade docente. Ou, na pior das hipóteses, adquirir status de Instituição durante sua atividade em sala de aula. Com isso, o aluno tende a repetir as técnicas que são utilizadas pelo professor, e, assim, garantem que não errarão, assegurando a sua conformidade com a instituição escolar ou a Instituição Professor.

A nossa pesquisa gera uma reflexão sobre a prática docente. O professor deverá estar atento à busca pela conformidade institucional, por parte do discente, a partir de escolhas das técnicas de resolução de tarefas, seja por influência do professor ou por caminhos próprios. A conformidade pode não representar um aprendizado significativo.

Referências

[1] Almouloud, S., Gras. R. & Régnier, J-C. (2014), *Especial ASI.* Revista Educação Matemática Pesquisa, Programa de Estudos Pós-graduados em Educação Matemática – PUCSP. São Paulo, v.16, n.3, pp.623-1087.

[2] Arsac, G.; Develay. M. & Tiberghien, A. (1989), *La transposition didactique en mathématiques, en physique, en biologie*. Lyon: IREM.

[3] Bessa de Menezes, M. (2004), *Investigando o processo de transposição didática interna: o caso dos quadriláteros*. 184f. Dissertação (Mestrado em Educação) – Universidade Federal de Pernambuco, Recife.

[4] Bessa de Menezes, M. (2010), *Praxeologia do Professor e do Aluno: Um análise das diferenças no ensino de Equações do Segundo Grau*. 177f. Tese (Doutorado em Educação) – Universidade Federal de Pernambuco, Recife.

[5] Booth, L. (1995), Dificuldades das crianças que se iniciam em álgebra. Em: Coxford, A. & Shulte, A (Orgs.), 1995. *As Idéias da Álgebra*. São Paulo, SP: Atual Editora. p. 23-37.

[6] Brito Menezes, A.P. (2006), *Contrato Didático e Transposição Didática: Inter-relações entre os fenômenos didáticos na iniciação à álgebra na 6ª série do Ensino Fundamental*. 410f. Tese (Doutorado em Educação). Universidade Federal de Pernambuco, Recife.

[7] Câmara dos Santos, M. (2002), Algumas concepções sobre o ensino-aprendizagem de matemática. In: *Educação Matemática em Revista.* nº 12. São Paulo: SBEM.

[8] Chevallard, Y. (1991), *La transposition didactique.* Grenoble: La pensée Sauvage.

[9] Chevallard, Y. (1998) Analyse des pratiques enseignantes et didactique des mathématiques: l'approche anthropologique. In : *L'UNIVERSITE D'ETE*, p.91-118. Actes de l'Université d'été La Rochelle, IREM, Clermont-Ferrand, France, 1998.

[10] Chevallard, Y. (1999), L'analyse des pratiques enseignantes en Théorie Anthropologie Didactique. In: *Recherches en Didactiques des Mathématiques*, p. 221-266.

[11] Coll, C. et All. (2006), *O construtivismo na sala de aula*. São Paulo: Ática.

[12] Gras R., Briand H., Peter P., Philippe J. (1997) : *Implicative statistical analysis, Proceedings of International Congress I.F.C.S.*, 96, Kobe, Springer-Verlag, Tokyo.

[13] Gras R., Régnier J.-C., Guillet F. (Eds) (2009) *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. RNTI-E-16 Toulouse: Cépaduès.

[14] Gras R., Régnier J.-C., Marinica C., Guillet F. (Eds) (2013) *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.* Toulouse: Cépaduès.

[15] Kieran, C. (1992), *The learning and teaching of school algebra. Handbook of research on mathematics teaching and learning*. National Council of Teachers of Mathematics; New York. p. 390-419.

[16] Kury, M. G. (2002), *Minidicionário Gama **Kury** da Língua Portuguesa*. 1ª Edição. São Paulo: FTD.

[17] Pais, L. C. (2001) *Didática da Matemática: Uma Análise da Influência Francesa*. Belo Horizonte: Autêntica.

[18] Ravel, L. (2003) *Des programmes a la classse: etude de la transposition didactique interne.* Tese de Doutorado não-publicada. Université Joseph Fourier – Grenoble I.

[19] Usiskin, Z. (1995), Concepções sobre a álgebra da escola média e utilizações das variáveis. In: Coxford, A. & Shulte, A (Orgs.), 1995. *As Ideias da Álgebra*. São Paulo, SP: Atual Editora. p. 9-22.

[20] Vergnaud, G., Cortes, A. (1986), *Introducing Algebra to "Low-level" Eighth and Nineth graders*. Proceedings of the Xth International Conference of Psychology of Mathematics Education. London. p. 319-324.

[21] Vergnaud, G., Cortes, A. & Favre-Artigue, P. (1987), Introduction de l'algèbre auprès de débutants faibles: problèmes épistémologiques et didactiques. In: Vergnaud, G., Brousseau, G. & Hulin, M. (Orgs.) 1987. *Didatique et aquisition des connaissances scientifiques: Actes du Colloque de Sèvres*. Sèvres, La Pensée Sauvage. p. 259-280.

PESQUISAS SOBRE PROFESSORES QUE ENSINAM MATEMÁTICA NOS ANOS INICIAIS DO ENSINO FUNDAMENTAL NO BRASIL ESTUDOS DAS TENDÊNCIAS DE PESQUISA NO BRASIL (2006 – 2014)[1] À LUZ DA ANÁLISE ESTATÍSTICA IMPLICATIVA

Luciana SILVA DOS SANTOS SOUZA[2], Marcelo CÂMARA DOS SANTOS[3], Nadja Maria ACIOLY-RÉGNIER[4] e Edênia Maria RIBEIRO DO AMARAL[5]

PROFESSEURS QUI ENSEIGNENT LES MATHÉMATIQUES DANS PREMIERES ANNEES DE L'ECOLE PRIMAIRE. ANALYSE DES TENDANCES DE RECHERCHE AU BRESIL (2006 - 2014) A LA LUMIERE DE L'ANALYSE STATISTIQUE IMPLICATIVE

RESUMO

Este artigo consiste no trabalho de conclusão da disciplina de Análise de Tendências de Pesquisa em Ensino de Ciências e Matemática, ministrada no Curso de Doutorado no PPGEC – UFRPE. Esta pesquisa objetiva delinear o panorama das pesquisas, realizadas entre 2006 e 2014, sobre os professores que ensinam matemática nos anos iniciais do ensino Fundamental. Com base em um levantamento realizado entre os artigos publicados e difundidos no Brasil, no âmbito da Educação Matemática, adotamos como estratégia metodológica a análise e a seleção de artigos em sítios virtuais dos periódicos-Qualis A e B (áreas: ensino e educação) e, em anais de eventos nacionais e internacionais, realizados pela Sociedade Brasileira de Educação Matemática – SBEM. Os dados oriundos do corpus latente na internet foram categorizados e analisados, com base nos tratamentos estatísticos, realizados com o suporte do software CHIC 6.0. Para tanto, nos respaldamos nos pressupostos da análise estatística implicativa (ASI) objetivando definir as principais tendências das pesquisas realizadas no Brasil acerca dos professores que atuam no início da escolarização básica. Os resultados revelam que há poucas pesquisas cujo foco é a formação matemática dos professores que atuam nos anos iniciais do ensino fundamental. Por outro lado, as já divulgadas no período investigado tratam dos saberes relativos à prática docente e, não da relação que os professores estabelecem com a Matemática, a ser ensinada por eles nesta modalidade de ensino.

Palavras-chave: *educação matemática, formação de professores dos anos iniciais, saberes docentes, interfaces da relação ao saber matemático, análise estatística implicativa.*

RÉSUMÉ

Cet article vise à décrire le panorama des recherches menées entre 2006 et 2014, s'intéressant à l'enseignement des mathématiques à l'école primaire. Nous nous sommes basés sur une recherche réalisée à partir d'articles publiés et diffusés au Brésil, dans le champ de l'Éducation Mathématique. Nous avons adopté comme stratégie méthodologique la sélection et l'analyse des articles sur l'éducation mathématique à partir de la base numérique des données on-line , portant sur ceux classés dans des revues qualifiantes au Brésil et dans des actes d'évènements nationaux et internationaux de la Société Brésilienne

[1] Esta pesquisa contou com o apoio da CAPES através da bolsa PVE para o programa PPGEC-UFRPE com o Professor Jean-Claude Régnier UMR 5191 ICAR – Université Lyon2

[2] Universidade Federal Rural de Pernambuco (UFRPE) em cotutela com Université Lumière/Lyon 2.E-mail: lucianasantos08@gmail.com.

[3] Universidade Federal de Pernambuco E-mail: marcelocamara@yahoo.com.br

[4] ESPE-École Supérieure du Professorat /Université Claude Bernard – Lyon 1. E-mail: nadja.acioly-regnier@univ-lyon1.fr

[5] Universidade Federal Rural de Pernambuco (UFRPE). E- mail: edsamaral@uol.com.br

de l'Éducation Mathématique – SBEM. Ces données ont été catégorisées et analysées à l'aide du logiciel CHIC 6.0. Nous avons adopté les principes de l'Analyse Statistique Implicative pour identifier les principales tendances des recherches effectuées au Brésil auprès des professeurs des écoles. D'une part, les résultats révèlent qu'il existe peu des recherches axées sur la formation mathématique des professeurs des écoles. D'autre part, ils mettent en évidence que, dans la période étudiée, les recherches déjà publiées abordent notamment les savoirs relatifs à la pratique enseignante et non sur le rapport qu'entretiennent ces enseignants au savoir mathématique.

Mots-clés : *éducation mathématique, formation des professeurs d'école primaire, savoirs des enseignants, rapport au savoir, analyse statistique implicative.*

1. Introdução

Nos últimos anos pesquisadores do âmbito educacional têm empreendido esforços homéricos objetivando a investigação do processo de profissionalização dos professores que atuam nas modalidades de ensino da educação básica ao ensino superior. Os focos de interesse destas investigações são diversificados, mas nos parece que ao menos o cenário é bem definido. Referimo-nos ao fato de que os contextos das pesquisas estão diretamente associados à formação inicial ou continuada do professor. Tais iniciativas descendem da necessidade de estar continuamente procurando esclarecer os meandros característicos da profissionalização e da construção da identidade docente. Neste sentido, destacamos a relevância para a comunidade científica internacional das pesquisas sobre a formação docente, desenvolvidas na última década, por Nóvoa (1995), Tardif (2000, 2011), Imbernón (2009, 2010) e seus colaboradores.

A formação de professores é uma das linhas de investigação dos pesquisadores em Educação Matemática. Os relatórios periódicos elaborados pelo grupo de trabalho 7 (GT-7), denominado professores que ensinam matemática, indicam o crescente interesse dos pesquisadores brasileiros pelas temáticas inerentes à construção do repertório de saberes da prática profissional; as concepções dos professores acerca da matemática e do seu ensino; dos entraves à construção de conceitos matemáticos, dos significados a estes associados ou à transposição didática dos conteúdos curriculares; embora, haja grupos voltados à formação matemática e aos aspectos didáticos ou metodológicos no ensino da matemática na educação básica.

Todavia, ao realizarmos a revisão bibliográfica preliminar da tese6, em desenvolvimento no Programa de Ensino das Ciências e Matemática (PPGEC) da Universidade Federal Rural de Pernambuco, nos deparamos com a escassez de estudos que versam sobre a formação matemática dos professores que atuam nos anos iniciais do ensino fundamental. Nesse sentido, destacamos que há uma quantidade menor ainda, de pesquisas sobre a relação que estes professores estabelecem com os saberes matemáticos que ensinam nesta modalidade de ensino. Mas, antes de investigarmos os modos de relação ao saber e as características psíquicas, sociológicas ou didáticas, que configuram as interfaces do fenômeno pesquisado na tese, tais constatações instigaram a nossa curiosidade fazendo emergir questões que possibilitaram atingir o objetivo de traçar o panorama das tendências de pesquisa acerca do professor que ensina matemática nos anos iniciais.

[6] Relação ao saber matemático de professores brasileiros e franceses que atuam nos primeiros anos da escolaridade. Projeto de tese em cotutela com a Université Lumière – Lyon 2.

Considerando, pois que as pesquisas já realizadas no âmbito da Educação Matemática (no período de 2006 a 2014), têm como contexto da investigação a formação (inicial ou continuada) do professor que ensina matemática nos anos iniciais do ensino fundamental, nós indagamos:

i) *Quais são os principais objetos de investigação eleitos pelos pesquisadores brasileiros*?

ii) *Como podemos caracterizar as pesquisas em função das implicações estatísticas (considerando o tipo de estudo, o método utilizado na(s) análise(s) e os instrumentos usados na construção dos dados)?*

iii) *As interfaces da relação ao saber matemático do professor estão contempladas na construção argumentativa dos autores dos artigos selecionados?*

O presente artigo consiste visa, portanto, esclarecer tais questões mediante o compartilhamento de um levantamento minucioso das publicações difundidas em periódicos nacionais (Qualis A e B) e, nos anais de eventos nacionais e internacionais em Educação Matemática, que apresentaram o professor dos anos iniciais do ensino fundamental como protagonista da investigação. A revisão bibliográfica, por nós realizada, também acaba por situar geograficamente a origem das publicações e a contribuição anual da comunidade acadêmica.

A construção dos dados, cujas análises apresentamos neste artigo, teve suporte em corpus latente na internet (bancos de dados virtuais on line) em virtude das dificuldades de acesso aos cd's-rom dos eventos da Sociedade Brasileira de Educação Matemática - SBEM. Para tanto, nos detivemos nos anais do Encontro Nacional de Educação Matemática – ENEM, do Congresso Internacional de Educação Matemática - CIAEM e do Simpósio Internacional de Educação Matemática – SIPEM acessíveis no site da referida sociedade. Assim como, revisitamos quatro periódicos, classificados pela Capes como Qualis A ou B, os quais encontram-se indexados à plataforma Scientific Eletronic Library Online - Scielo, entre eles: o Boletim de Educação Matemática – BOLEMA; a Zetetiké - Revista de Educação Matemática; a Educação em Revista e Ciência & Educação;

O nosso banco de dados foi construído com base em artigos selecionados segundo dois critérios: (a) ter como contexto a formação (inicial ou continuada); (b) ter como sujeito participante o professor (em formação ou atuante) nos anos iniciais do ensino fundamental. Posteriormente, procedemos a categorização destas informações em função de variáveis essenciais e suplementares (descritas adiante no metodológicos). As análises dos dados apresentadas neste artigo encontram respaldo em elementos qualitativos e quantitativos, seguindo as orientações metodológicas da análise do conteúdo (Bardin, 2011) e da análise estatística implicativa (ASI). Neste sentido, o software CHIC 6.0 foi a ferramenta utilizada para produzir o grafo implicativo, as árvores coesitivas e de similaridade que possibilitou a organização e a análise dos resultados obtidos nesta pesquisa.

2. Explicitação do objeto da pesquisa

A tese em desenvolvimento no curso de doutorado da Universidade Federal Rural de Pernambuco em cotutela com a Université Lumière – Lyon 2, tem como objeto de pesquisa o fenômeno da relação ao saber matemático dos professores que ensinam matemática no 5º ano do ensino fundamental (no Brasil) e no Cours Moyen 2 da école primaire (na França).

A noção de relação ao saber pode ser definida em diferentes quadros teóricos em função das múltiplas interfaces que a constituem e, vem sendo desenvolvida desde o início de 1960 por Jacky Beillerot entre outros pesquisadores que lhe sucederam, tais como: Charlot, Chevallard, Blanchard-Laville, Mosconi, por exemplo.

Câmara dos Santos (1997) define a relação ao saber matemático como a disposição de cada indivíduo em relação ao objeto do saber. Tal relação pode ser considerada um fenômeno multifacetado, uma vez que suas características constitutivas, nem sempre perceptíveis a olho nu, podem ser extraídas em aspectos psíquicos (conscientes e inconscientes), sociológicos e didáticos que permeiam o universo pessoal/profissional e da personalidade dos professores.

Em outras palavras, a relação ao saber pressupõe a existência de uma distância entre o sujeito e o objeto do saber. Esse distanciamento é relativo, pois a proximidade revela-se algo circunstancial e vulnerável às influências psíquicas (sentimentos e atitudes), sociológicas (eu's sociais, sujeições, relações interpessoais) e didáticas (praxeologias, mecanismos de ensino, etc.). Essa relação determina, em larga escala, de que maneira ele organiza as situações de ensino e que conhecimentos ele irá, ou não, privilegiar no momento da ação didática, ou seja, de que forma ele irá negociar o contrato didático e a *cara* que ele dará aos conhecimentos no momento da transposição didática. (Câmara dos Santos, 2014, p.639)

Charlot (2008, p.58) nos diz que as relações com o saber (ou os "aprenderes") devem ser privilegiadas nas pesquisas considerando-se suas especificidades epistemológicas, cognitivas e didáticas. Mais adiante, afirma que estas "pesquisas poderiam ser preciosas para aprofundar a questão da relação ao saber. De fato, se os princípios da especificidade dos objetos de saber e da normatividade das atividades que permitem a um sujeito apropriar-se deles foram postos, as pesquisas até agora não avançaram muito".

A primeira revisão bibliográfica que realizamos coaduna os argumentos de Charlot, pois embora os pesquisadores brasileiros e franceses tenham contribuído significativamente para ampliar o conhecimento acerca da relação ao saber matemático, ainda não encontramos modelos teóricos construídos para explicitar a relação ao saber dos professores que ensinam matemática no início da escolaridade em ambos países. Portanto, a tese que estamos desenvolvendo visa fomentar o debate e trazer elaborações teóricas que forneçam elementos para a construção de um modelo da relação ao saber matemático dos professores que atuam nesta modalidade de ensino. Para tanto, dedicamos os primeiros esforços de pesquisa para mapear a produção científica, nos países supracitados, acerca deste fenômeno.

Fizemos uma ampla revisão bibliográfica, seguida da revisão em portais de periódicos e nos acervos dos bancos de teses disponíveis na internet. E, nos deparamos

com a escassez de pesquisas sobre o professor que atua nos anos iniciais do ensino fundamental e no CM2 da école primaire. As dissertações e teses que conseguimos identificar direcionaram os focos de suas lentes investigativas acerca da relação ao saber dos (futuros) professores de matemática ou dos alunos dos anos finais do ensino fundamental ou médio.

Por esta razão, ampliamos o campo de buscas no corpus latente na internet e o intervalo temporal de 5 para 8 anos (2006 a 2014). Na tentativa de elucidar as questões de partida, nossas estratégias de busca, nos conduziram às comunicações científicas publicadas nos anais dos principais eventos da Sociedade Brasileira de Educação Matemática - SBEM e, aos artigos científicos publicados em periódicos (Qualis A e B nas áreas de Ensino e Educação), os quais encontram-se indexados à plataforma Scielo.

Incialmente, utilizamos como palavras-chave os seguintes termos: *relação ao saber* e, em francês, *rapport au savoir*, porém as buscas revelaram a escassez de estudos relativos ao fenômeno que estamos investigando. Todavia, constatamos que algumas produções trazem o termo no título, mas, uma leitura mais cuidadosa indica que a discussão fomentada pelos autores evidencia que a relação ao saber do professor não é o principal argumento. Esta, serve apenas como um plano de fundo para tratar dos saberes que permeiam a prática docente dos professores que ensinam matemática nos primeiros anos do ensino fundamental.

Melhorar a compreensão acerca da relação ao saber matemático dos professores é algo essencial e significa compreender a relação epistemológica dos profissionais que inserem no contexto da sala de aula os primeiros conceitos, significados e relações entre os objetos de conhecimento da matemática escolar. Entender o(s) modo(s) como estes se relacionam com os saberes a serem ensinados descortina um universo de possibilidades de conhecer entre outras coisas como este fator interfere nos modos de relação ao saber dos alunos, na construção do espaço psíquico em sala de aula, da proposição/diversificação das situações didáticas que envolvem a matemática, na seleção dos saberes em cena no jogo didático, na efetiva gestão dos tempos didáticos e de aprendizagem, por exemplo. No entanto, antes de tentarmos nos aproximar destes e, de outros aspectos, há uma premente necessidade de conhecer as reais motivações que incentivam as pesquisas, realizadas no Brasil, sobre o professor que ensina matemática nos anos iniciais.

3. Delimitação da amostra

Diante destas dificuldades e dos desafios aos quais nos referimos optamos pelo redirecionamento das buscas relativas à revisão de literatura, acionando como palavras-chave os sintagmas: *professores dos anos iniciais* e *formação docente*. Consequentemente, esta estratégia metodológica, resultou no aumento exponencial dos resultados das buscas.

A etapa seguinte desta revisão foi concentrada na análise de artigos disponibilizados nos anais de eventos que estão ativos e acessíveis na plataforma on line da SBEM. Nesse sentido, consideramos as três últimas edições do Encontro Nacional de Educação

Matemática (ENEM)[7], do Seminário Internacional de Pesquisa em Educação Matemática (SIPEM)[8]. E, a última edição, da Conferência Interamericana de Educação Matemática (CIAEM) realizada em 2011.

Para agregar mais dados ao banco pré-existente, pesquisamos em periódicos indexados ao Scielo, utilizando como pré-requisito a classificação da Capes, no Qualis A ou B das áreas de ensino e educação. Cabe ressaltar que analisamos os volumes de 12 periódicos, mas logramos êxito apenas nas pesquisas realizadas em volumes do Boletim de Educação Matemática – BOLEMA (5 artigos), da revista ZETETIKÉ (2 artigos), da Revista Ciência & Educação (1 artigo) e da Educação em Revista (1 artigo).

Os eixos de discussão e a concentração das publicações (artigos) nos fóruns sobre Educação Matemática são muito diversificados, poderíamos citar inúmeros deles: formação de professores (inicial ou continuada), ensino e aprendizagem da matemática (nos Anos Iniciais ou Finais do Ensino Fundamental e Médio), por exemplo. No entanto, os eventos mencionados têm uma matriz mais alinhada, organizada em grupos de trabalhos, os chamados GT's.

Desse modo, direcionamos nosso olhar para o GT 7 que discute e concentra as publicações relativas à formação inicial e continuada de Professores no âmbito da Educação Matemática. Os trabalhos desse eixo veem a formação, seja inicial ou continuada, como a "espinha dorsal" dos processos de construção da identidade docente daqueles que, independentemente da modalidade (Educação Básica ou Ensino Superior), ensinam matemática. Neste caso, são focos de interesse do GT 7 correspondem à(s):

a. *Formação inicial* em cursos de licenciatura em matemática, pedagogia, e de outros cursos que lhe permitam atuar como docente.
b. *Formação/desenvolvimento profissional do formador de professores* que lecionam Matemática.
c. *Formação/desenvolvimento profissional do professor que leciona Matemática no Ensino Superior* (não necessariamente daqueles que atuam nas licenciaturas em matemática ou pedagogia).
d. *Experiências formais e informais de formação continuada*, o desenvolvimento profissional e as aprendizagens docentes envolvendo grupos de professores da escola e/ou futuros professores e/ou professores da universidade.
e. *Os saberes docentes*: quais são, como se constroem/desenvolvem, etc.
f. *Os processos de constituição da identidade profissional* (percursos profissionais) e a profissionalidade docente.

Na constituição do quadro de variáveis (categorização) que possibilitou as análises dos artigos por nós selecionados, consideramos as temáticas discutidas no GT7 em encontros nacionais e internacionais promovidos pela SBEM. Referimo-nos aos itens a), b), d), e) e f) acima relacionados.

[7] Realizados nos anos de: 2007 (Belo Horizonte), 2010 (Salvador) e 2013 (Curitiba).
[8] Realizados nos anos de: 2006 (Curitiba), 2009 (Brasília) e 2012 (Rio de Janeiro).

Ao adotar todas as medidas e decisões, anteriormente justificadas, obtivemos uma população composta por 57 publicações. Sendo que, deste espaço amostral, 48 artigos foram divulgados na categoria comunicação científica (CC) e, os 9 remanescentes, correspondem aos artigos publicados em periódicos (P) entre os anos de 2006 e 2014. O gráfico, ilustrado na Figura 1 mostra a dispersão das publicações científicas ao longo do intervalo temporal considerado nesta pesquisa (2006-2014).

Figura 1: Produção acadêmica (2006 - 2014)

4. Percurso metodológico

Ao construir o banco de dados nas fontes referenciadas anteriormente procedemos a elaboração de uma categorização por meio da eleição de variáveis binomiais e suplementares para que pudéssemos realizar a análise estatística implicativa (ASI). Gras & Couturier (2013) afirmam que estão convencidos de que a dialética "quantidade-qualidade" conduz a favor dos limiares quantitativamente cruzados (população, variáveis, índice) para dar saltos qualitativos no entendimento do “todo”, comparáveis à passagem da água do estado líquido ao sólido, quando diminuímos a temperatura. Esta é a organização das peças no sistema, as relações entre elas que constituem a fonte explicativa da emergência da nova compreensão “todo”.

Além disso, Acioly-Régnier & Régnier (2010, p. 577) afirmam que as exigências concretas àquelas que são confrontadas pelos pesquisadores não especialistas, ao utilizar-se do software C.H.I.C., os conduzem à necessidade de uma elevação do nível de conceitualização que pode se realizar pela própria consciência dos conceitos imbricados na atividade a partir de sua análise. Todavia, isto conduz o pesquisador a novos questionamentos sobre a própria concepção do software, ao oferecer uma interface interessante que possibilita ao mesmo as condições necessárias e suficientes para poder se apoiar sobre uma mediação didática suscetível de subsidiar a formação do pesquisador não especialista.

Em outras palavras a ASI, mediada pelo uso do CHIC, oferece aos pesquisadores não especialistas possibilidade de realizar múltiplos cruzamentos entre sujeitos, que

neste caso correspondem aos artigos científicos, e as variáveis que aqui são as características da pesquisa relatada pelos autores destas produções. Estes cruzamentos resultam em elementos gráficos que ilustram não apenas as implicações, as similaridades e a cadeia hierárquica entre os sujeitos e as variáveis consideradas na pesquisa, mas também, propiciam ao pesquisador entender o processo que revela o fenômeno investigado como um todo. Nos Quadros 1a e 1b relacionamos e descrevemos as sete categorias e as vinte e quatro subcategorias que constituem as variáveis das planilhas as analisadas com o uso do CHIC 6.0 .

Quadro 1a: Categorização dos dados

CATEGORIAS	SUBCATEGORIAS	COD ASI	DESCRIÇÃO
Vetor variável 1 (V01) TIPO DE PUBLICAÇÃO	**Comunicação Científica**	CC	Artigo apresentado na categoria comunicação científica e publicado nos anais dos eventos ou, em periódicos Qualis A ou B da área de ensino/educação, selecionados neste estudo.
	Periódicos	P	
Vetor variável 2 (V02) CONTEXTO	**Formação inicial**	V2A	Compreende as pesquisas realizadas com professores em formação dos cursos de licenciatura em pedagogia.
	Formação continuada	V2B	Pesquisas realizadas com professores em formação continuada estabelecida pela legislação em vigor e oferecidas pelas secretarias de educação, parcerias ou por instituições formadoras (universidades, ong's e institutos).
Vetor Variável 3 (V03) OBJETO DA INVESTIGAÇÃO	**Saberes docente**	V3A	Segundo a classificação proposta por TARDIF (2000) estes saberes podem ser experienciais, curriculares, disciplinares ou pedagógicos.
	Metodologia de ensino	V3B	A resolução de problemas, a modelagem no ensino e a aprendizagem da matemática escolar.
	Formação matemática	V3C	A formação matemática do professor que atua nos anos iniciais do EF, os impactos nas práxis e no ensino da matemática.
	Conhecimento matemático	V3D	Refere-se ao conhecimento específico acerca dos objetos de ensino do currículo dos anos iniciais do EF (frações, números decimais, sistema de numeração, etc.).
	Outro foco de interesse	V3E	Contributo das disciplinas (metodologia, prática de ensino, etc.), para a formação da identidade docente ou práxis docentes; implicações sociais da matemática; avaliação do modelo formativo; diversificação de tarefas matemáticas.

Quadro 1b: Categorização dos dados

CATEGORIAS	SUBCATEGORIAS	COD ASI	DESCRIÇÃO
Vetor Variável 4 (V04) TIPIFICAÇÃO DA PESQUISA.	**Estudo exploratório**	V4A	Resultados parciais ou destinados ao redimensionamento das variáveis que compõem pesquisas mais amplas.
	Estudo de caso	V4B	Resultados de pesquisas sobre fenômenos específicos sem a intenção de generalização.
	Pesquisa participante ou **intervencionista**	V4C	Resultados de estudos oriundos de proposições ou projetos de intervenção pontual em que o pesquisador esteve envolvido diretamente
	Pesquisa ação	V4D	Resultados de pesquisas que propõem, o planejamento, a reflexão e a ação sobre a prática docente com mais regularidade.
	Outro	V4E	Estudo fenomenológico, colaborativo, documental, etc.
Vetor Variável 5 (V05) MÉTODO	**Análise do discurso**	V5A	Análise pautada no teor da fala em narrativas, da enunciação, de situações de comunicação captadas em áudio e videogravações, por exemplo.
	Análise do conteúdo	V5B	Análise pautada no teor de documentos, textos, postagens, etc.
	Múltiplas análises	V5C	Análise do conteúdo associada à análise do discurso; ou à etnografia, por exemplo)
	Não identificável	V5D	O método não é enunciado/explicitado na pesquisa.
Variável 6 (V06) ABORDAGEM da relação ao saber matemático do professor	**Psicanalítica**	V6A	O viés psicanalítico, sociológico ou didático é utilizado como plano de fundo na argumentação acerca da relação ao saber matemático do professor.
	Sociológica	V6B	
	Didática	V6C	
	Imperceptível	V6D	Associações atreladas às interfaces da relação ao saber do professor não identificadas
Variável 7 (V07) INSTRUMENTOS utilizados na construção dos dados	**Questionário**	V7A	Utilização de um único instrumento para produzir os dados.
	Entrevista	V7B	
	Produção textual	V7C	Relatos escritos, transcrições de narrativas, conteúdo de postagens em fóruns de discussão (moodle), elaboração de projetos, situações problema ou atividades matemáticas.
	Múltiplos instrumentos	V7D	Observação (diário de campo), associadas ao uso de questionário e/ou entrevistas; ou ainda, à relatos de experiência em sala de aula ou formação do professor.
	Outros instrumentos	V7E	Análise de documentos, currículos, ementas de disciplinas acadêmicas, corpus latente da internet, conteúdo de fóruns – plataforma moodle, entre outros.
	Análise de estratégias	V7F	Análise de estratégias de resolução de problemas matemáticos em avaliações ou testes.

Portanto, para as categorias e subcategorias fossem transformadas em objeto de análise (variáveis binárias) houve a conversão da linguagem textual em computacional. Ou seja, em se tratando de variáveis binomiais codificamos com 1(um) para indicar

presença, ou com 0 (zero) para indicar a ausência, em função dos elementos da categorização proposta nos Quadros 1a e 1b.

Salientamos que, neste caso, adotamos duas variáveis suplementares a primeira refere-se à *origem do artigo* e, a segunda, corresponde ao *ano* em que a publicação ocorreu. Portanto, optamos pela verificação do contributo da comunidade científica em função da localização geográfica da instituição vinculada ao autor em uma das cinco regiões brasileiras. Bem como, da verificação da frequência média de artigos publicados entre 2006 e 2014 que se enquadram nos critérios de seleção deste estudo. Ou seja, os 57 artigos selecionados correspondem aos indivíduos desta investigação. Esta medida, por sua vez é a condição sine qua non para viabilizar a análise estatística implicativa (ASI) realizada com o suporte do software CHIC - Classificação Hierárquica Implicativa e Coesiva.

De acordo com Couturier, Bodin & Gras (2004) as principais do software CHIC consistem em extrair de um conjunto de informações, mediante o cruzamento entre sujeitos ou objetos e as variáveis (características, atributos, por exemplo), em função das regras de associação entre variáveis (lei binomial ou de Poisson), fornece um índice de qualidade dessas associações e uma representação estruturada (grafo implicativo, árvore das similaridades e árvore coesitivas) das variáveis obtida por meio destas regras. Na discussão dos resultados as representações gráficas supracitadas forneceram as condições necessárias e suficientes para elaboração das análises de similaridades, das implicações obtidos e da hierarquização coesitiva observada neste estudo.

5. Discussão dos resultados na perspectiva da ASI

Conforme mencionamos anteriormente, a ferramenta utilizada para realizar as análises está ancorada nos pressupostos da ASI por meio dos recursos disponíveis na interface do CHIC 6.0. De acordo com Couturier, Bodin & Gras (2004) este potente instrumento permite a realização de análises de dados a partir de variáveis binárias, modal, frequencial ou de intervalos.

Estas variáveis podem ser principais ou secundárias também denominadas por suplementares. Nesta pesquisa utilizamos apenas variáveis binárias principais e suplementares (características da pesquisa) que foram associadas aos sujeitos (artigos científicos) com o intuito de traçar o panorama das pesquisas realizadas no Brasil acerca dos professores que ensinam matemática nos anos iniciais do ensino fundamental.

Basicamente, os critérios dos tratamentos realizados no CHIC tratam da lei de implicação clássica, segundo esta regra uma relação binária se estabelece quando: dados dois conjuntos não vazios, V (variáveis) e S (sujeitos), o produto cartesiano de V por S o conjunto formado por todos os pares ordenados (a, b) com a em V e b em S, temos uma relação binária de V em S. Assim sendo, a implicação obedece a seguinte regra: $se\ a \wedge\ b\ e\ b \wedge c, então\ a\ \wedge c$.

Outro critério para o tratamento dos dados, à escolha do pesquisador, consiste em optar pelas leis binomial ou de Poisson. Nós adotamos a lei binomial, segundo a qual: seja ε uma experiência aleatória e A um acontecimento associado a ε tal que a probabilidade P (A) = p, em que $(p \in]\,0, 1[\,)$.

6. Implicações encontradas

Para construção do grafo implicativo adotamos um índice de confiança ≥ 0,8. Vale salientar que utilizamos o recurso das cores para ilustrar melhor os diferentes índices de confiança das correlações encontradas. Portanto, no grafo implicativo os segmentos de reta representados na cor cinza corresponde aos índices ≥ 0,8. Enquanto os segmentos grafados na cor verde referem-se aos índices ≥ 0,9.

É relevante dizer que nesse primeiro passo, limitamo-nos a considerar as correlações que apresentam um grau de confiança maior porque estas correlações são as mais significativas do ponto de vista da ASI[9]. Por outro lado, ao escolhermos um índice de confiança maior acreditamos que as correlações encontradas respondem mais objetivamente às questões de partida que norteiam este estudo acerca das tendências de pesquisa, adotadas por pesquisadores brasileiros, em investigações que envolvem a participação dos professores graduandos ou já habilitados para o ensino do componente curricular matemática nos anos iniciais do ensino fundamental do Brasil.

Tais correlações, ilustradas na Figura 2, nos permite verificar que há caminhos mais extensos envolvendo três variáveis e, outros menos extensos, compostos por duas variáveis. Estes subconjuntos estão diretamente relacionados às subclasses ilustradas na árvore das similaridades (Figura 3). Vale salientar que os tratamentos no CHIC não revelaram a existência de transitividades entre os referidos conjuntos de variáveis apresentados neste estudo.

[9] Para facilitar a compreensão das análises utilizaremos as cores gray (cinza) para indicar o índice de confiança superior a 0,80 e green (verde) para indicar um índice nas setas que representam o caminho das implicações.

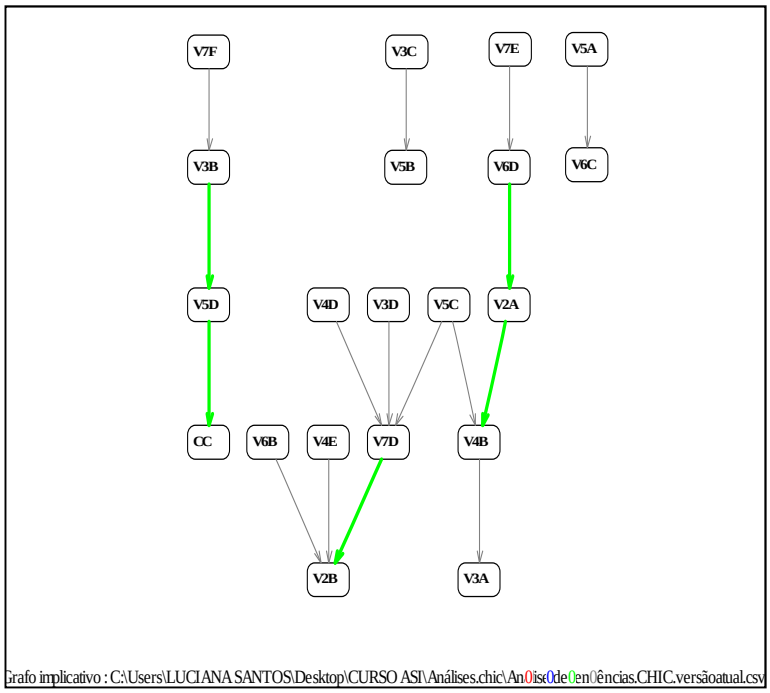


Figura 2: Grafo implicativo

Dentre as trajetórias reveladas pelo CHIC, em função das variáveis binárias, há dois caminhos mais extensos, com índice de implicação ≥ 0,90 (cujas setas implicativas encontram-se na cor verde). A seguir, relacionamos nossas inferências quanto a estes conjuntos de variáveis, incluindo aquelas de menor potencial. São elas:

a. (V7F) →V3B → V5D → CC

Com base na categorização (Quadro 1), podemos afirmar que as comunicações científicas (CC) com maior contributo das pesquisas apresentadas na X e XI edições do Encontro Nacional de Educação Matemática, que representa o grupo optimal de 6. Estas publicações têm como objeto de investigação as questões relativas às metodologias de ensino da matemática, equivalente à variável V3B (resolução de problemas e modelagem), outro aspecto característico é que os autores não enunciaram o método utilizado na análise dos dados (V5D). Interessante é perceber que associado a este subconjunto a variável V7F que indica que os dados das pesquisas foram construídos com base em protocolos de resolução de problemas (teste ou avaliações) e na construção de modelos. A tipicalidade deste caminho fica por conta cinco trabalhos publicados na IX edição do ENEM e 1 artigo (BOLEMA2) do periódico Boletim de Educação Matemática.

b. (V7E) → V6D → V2A → V4B→(V3A)

Neste caminho, o maior índice de implicação corresponde à variável V6D até a variável V4B. Isto indica que a base argumentativa não possibilitou o estabelecimento de associações acerca das interfaces da relação ao saber (V6D). O grupo optimal compreende 6 pesquisas publicadas nas duas últimas edições do ENEM e do Simpósio Internacional de Educação Matemática (SIPEM). Ou seja, estes são os principais contribuintes desta implicação. Nestas os autores utilizaram como contexto a formação

inicial dos futuros professores dos anos iniciais – licenciandos de cursos de pedagogia (V2A). Outras características comuns às referidas publicações é que estas se baseiam nas premissas da análise do conteúdo (V4B) e configuram-se como estudos exploratórios (V3A) que objetivam a produção de dados preliminares de pesquisas mais amplas coordenadas pelo(s) autor(es) ou a (re)adequação das variáveis presentes nas produções analisadas. A tipicalidade deste caminho corresponde a um grupo optimal com 17 elementos que envolvem pesquisas em periódicos: Ciência & Educação (C & E) e o Boletim de Educação Matemática (BOLEMA); além daqueles pulicados no SIPEM e no CIAEM.

c. (V5C) →V7D → V2B

Tomando como ponto de partida a variável V5C, referente ao emprego de múltiplas análises no tratamento dos dados das pesquisas publicadas, existem dois caminhos: o mais curto induz à variável (V4B), tal implicação significa que, independentemente do tipo de associação, a análise do conteúdo é uma das estratégias metodológicas adotadas. O segundo caminho, mais longo, é coerente porque nos leva a concluir que o uso de mais de um método na análise dos dados implica na utilização de múltiplos instrumentos de pesquisa (V7D) nas investigações que usaram como contexto os espaços destinados à formação continuada (V2A) dos professores que ensinam matemática nos anos iniciais. Ainda em relação ao contexto outras duas implicações são notadas (V4E) e (V6B). Ou seja, o tipo de pesquisa apresenta um caráter fenomenológico e colaborativo. Além disso, a base argumentativa trazida pelos autores, traz como plano de fundo elementos da interface sociológica (influências das interações sociais nas práxis ou na identidade docente) da relação ao saber dos participantes das pesquisas analisadas. Observa-se que a tipicalidade neste caso, refere-se há 22 publicações em periódicos e nos eventos, sendo o mais representativo os do V SIPEM.

d. (V3C) →(V5B) e **(V5A) →(V6C)**

No grafo implicativo (Figura 2) percebe-se dois pares de variáveis isoladas dos demais esquemas. A primeira associação denota que as pesquisas analisadas que trazem como objeto de estudo a formação matemática do professor (V3C) implicam na utilização dos pressupostos da análise do conteúdo (V5B). Enquanto o segundo caminho indica que o emprego de métodos de análise do discurso do professor (V5A) implica que a base argumentativa do(s) autor(es) contém elementos que trazem como plano de fundo a interface didática da relação ao saber do professor. Mas, precisamente no que diz respeito à relação pessoal ou institucional com os saberes matemáticos que estes ensinam.

7. Similaridades percebidas: descrição e inferências

De acordo com Gras & Ag Almouloud (2012) a similaridade se define a partir do cruzamento do conjunto V das variáveis com um conjunto S de sujeitos. Este tipo de análise permite ao usuário estudar e interpretar, em termos de tipologia e de semelhança (e não semelhança) decrescente, classes de variáveis, constituídas significativamente a certos níveis da árvore e se opondo a outras nestes mesmos níveis. Diante do exposto estes autores afirmam que o *critério de similaridade* se exprime, no caso das variáveis

binárias, portanto para duas variáveis a e b, satisfeitas respectivamente por dois subconjuntos A e B de S, são muito semelhantes quando o número (*k*) dos sujeitos de A ∧ B é importante de um lado, pelo que teria sido no caso da ausência de ligação entre a e b. Por outro lado, com relação aos cardinais de E, A e B.

A referida semelhança é medida pela probabilidade de que (*p)* seja superior ao número aleatório esperado nesta situação. A Figura 3 ilustra seis subconjuntos, os quais denominamos como subclasses, oriundos da classe geral de variáveis desta pesquisa. As subclasses (i), (ii) e (iii) formam a Classe A e as subclasses (iv),(v) e (vi) integram a Classe B. Estes agrupamentos estão inter-relacionados entre si e, juntamente com a variável 6A que ficou isolada das referidas subclasses, constituem a árvore das similaridades. A variável que mencionamos corresponde a interface psicanalítica da relação ao saber.

Independentemente, do objeto de estudo eleito pelos autores das publicações que analisamos, estas não contemplam nas discussões ou nos resultados apresentados o caráter individual e subjetivo da relação que os professores dos anos iniciais do ensino fundamental, estabelecem com a matemática que ensinam. Em certa medida, a tese que estamos desenvolvendo no PPGEC da UFRPE em cooperação com a Université Lumière-Lyon 2, poderá vir a contribuir neste sentido. A seguir relacionamos e interpretamos as similaridades entre as variáveis adotadas em nossa pesquisa, ilustradas na Figura 3:

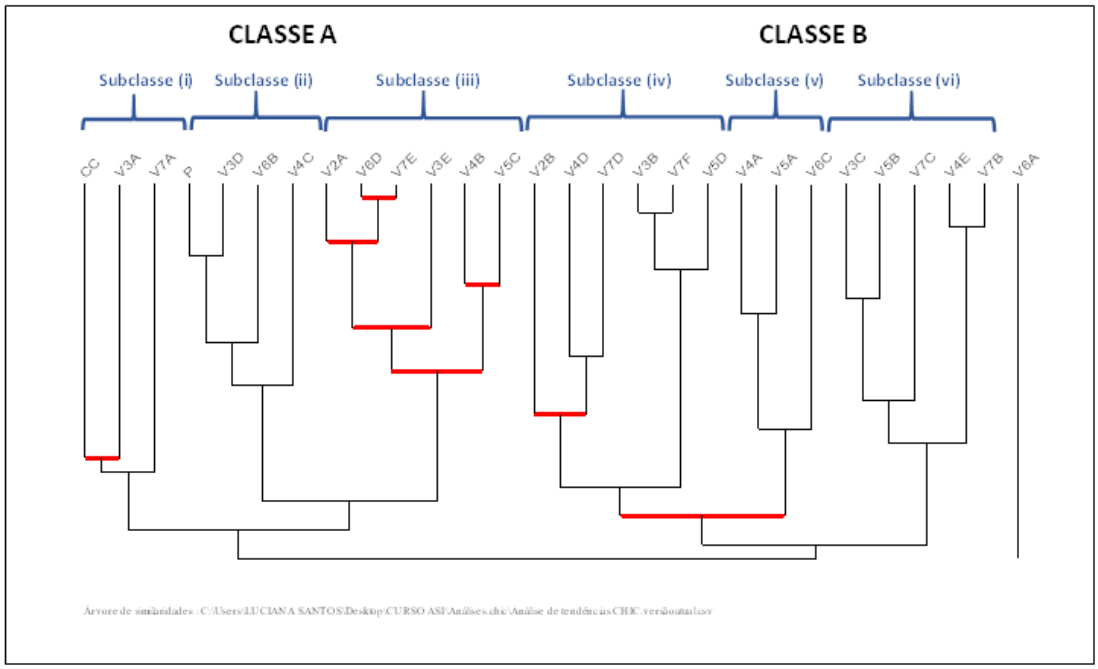


Figura 3: Árvore das similaridades

7.1 Similaridades da CLASSE A

Subclasse (i) ((CC-3A)-7A) – Nesta subclasse encontram-se as produções científicas difundidas em eventos na modalidade comunicação científica. A característica comum, o nó mais significativo, pode ser interpretado ao tipo de produção associada ao estudo

de dois tipos de saberes docentes, mais precisamente aos que Schulman (1986) denominou como saberes específicos do conteúdo e os saberes pedagógico do conteúdo. Ou seja, as pesquisas que selecionamos têm investigado sobre o que os professores dos anos iniciais sabem acerca da matemática escolar que ensinam nesta modalidade e, também como eles abordam e trabalham esse conteúdo em sala de aula. Para construir dados sobre estes saberes docentes, tais pesquisas se apoiam em um único instrumento: o questionário.

Subclasse (ii) ((P-3D)-6B)-4C – As produções dos pesquisadores, divulgadas em periódicos Qualis A e B (área de ensino e educação) referem-se ao estudo do conhecimento matemático do professor acerca dos objetos de ensino, presentes no currículo dos anos iniciais do Ensino Fundamental. Entre os conteúdos pesquisados, destacamos frações, as operações envolvendo números naturais ou decimais, sistema de numeração decimal, noções de probabilidade e geometria. Estas pesquisas têm um caráter intervencionistas, promovidas fundamentalmente entre as universidades e as secretarias de educação. O interessante, neste caso, é que embora o foco destes estudos não seja a relação ao saber matemático dos professores, os autores tentam estabelecer vínculos entre as ações formativas, as interações entre os participantes, a (re)aprendizagem dos professores e/ou as mudanças na praxeologia dele. Portanto, apontam a influência de aspectos sociológicos na relação que o professor estabelece com a matemática que ele ensina.

Subclasse (iii) ((2A-(6D-7E))-3E)-(4B-5C) Os nós mais significativos da árvore das similaridades estão neste agrupamento. Esta associação compreende as pesquisas realizadas na formação inicial dos professores que atuarão nos anos iniciais do ensino fundamental. Ou seja, os pesquisadores que produziram estes estudos estão interessados nos cursos de licenciatura em pedagogia. Neste caso, não é possível associar as bases argumentativas destas produções as interfaces da relação ao saber dos futuros pedagogos. Dentre os instrumentos utilizados na construção dos dados estão: a análise documental dos currículos dos cursos superiores de formação de professores, ementas de disciplinas (metodologia de ensino da matemática, por exemplo). Ou ainda, com base nas postagens em fóruns de discussão (plataforma moodle). Pois, tais pesquisas visam principalmente, identificar os contributos das disciplinas acadêmicas (metodologia, prática de ensino, etc.), ou as adequações necessárias no currículo de cursos superiores de pedagogia que possam ajudar na formação da identidade ou das práxis docentes; implicações sociais da matemática; avaliação do modelo formativo; diversificação de tarefas matemáticas.

7.2 Similaridades da CLASSE B

Subclasse (iv) ((2B-(4D-7D))-(3B-7F)-5D) Neste caso, verificamos que as pesquisas foram realizadas no contexto da formação continuada dos professores que já atuam nos anos iniciais do ensino fundamental. Apresentam características predominantes da pesquisa-ação e os autores construíram os dados mediante o cruzamento de informações oriundas de múltiplos instrumentos (questionários associados a observações, entrevistas, narrativas, relatos, por exemplo. O foco destas investigações corresponde a metodologia

de ensino da matemática, principalmente em relação à adoção da resolução de problemas pelos professores que atuam no 4º e 5º anos do ensino fundamental). Mas, dentre as pesquisas que integram este agrupamento, há algumas, cujos dados resultam da análise das estratégias de resolução de problemas por estes professores. Nos exemplares, que integram esta subclasse, os autores não enunciam o método adotado na análise dos dados. Este agrupamento compõe um nó significativo com a subclasse (v) que descrevemos a seguir.

Subclasse (v) ((4A-5A)-6C) Esta subclasse refere-se às pesquisas correspondem à estudos exploratórios, portanto os autores visavam ajustar as diretrizes de outras pesquisas mais amplas (projetos de doutoramento, mestrado, etc.). Neste caso, as premissas da análise do discurso nortearam a análise dos dados. E, a interface didática da relação ao saber matemático dos professores dos anos iniciais do ensino fundamental está presente nas discussões dos resultados destas pesquisas.

Subclasse (vi) (((3C-5B)-7C)–(4E-7B)) Nesta subclasse verificamos que as pesquisas referem-se à formação matemática dos professores dos anos iniciais do ensino fundamental. Para analisar os dados, os autores das publicações, adotaram as técnicas características da análise do conteúdo. Os instrumentos que propiciaram a construção dos dados foram as narrativas dos professores nos encontros formativos, ou postagens em fóruns de discussão on line, elaboração de projetos, situações problema ou atividades matemáticas. Em parte das pesquisas que compõem esta subclasse, predomina a natureza é fenomenológica e o viés colaborativo (parcerias entre as redes de ensino e as universidades, por exemplo). Neste último caso, o único instrumento utilizado na construção dos dados foi a entrevista com os participantes.

8. Classificações hierárquicas coesitivas observadas neste estudo

Observa-se na árvore coesitiva, apresentada na Figura 4, que há quantitativo considerável de classes implicativas formadas por grupos contendo duas, três e, até quatro, variáveis consideradas nesta investigação.

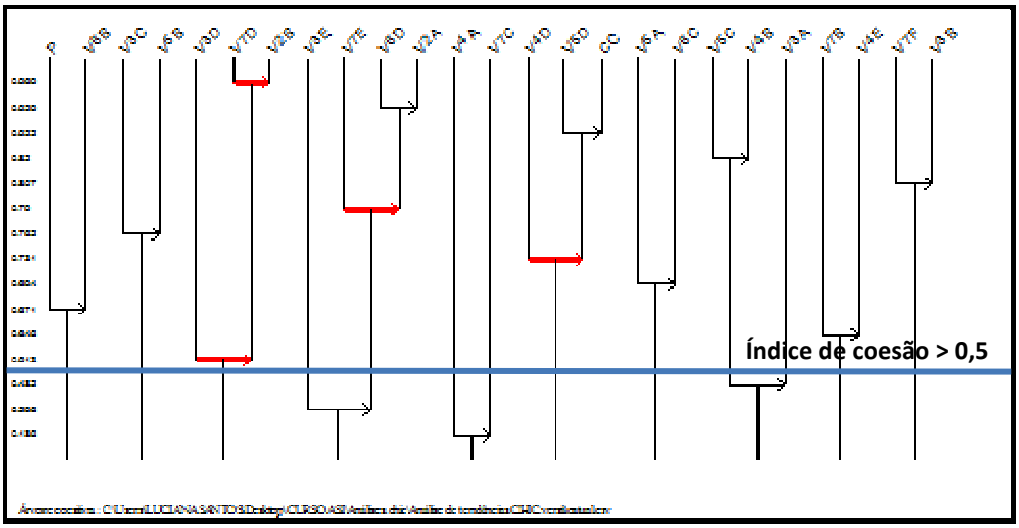


Figura 4: Árvore coesitiva

Para fins de análise destas classes, consideraremos apenas aquelas que apresentam índice qualitativo de coesão superior a 0,5. A seguir, delinearemos brevemente as principais implicações verificadas:

3D→(7D→2B): As pesquisas que versam sobre o conhecimento matemático implicam na aplicação diretamente na utilização de vários instrumentos para construir os dados. O cenário escolhido pelos pesquisadores que assinam a autoria das produções foram os encontros destinados à formação continuada dos professores dos anos iniciais do ensino fundamental. Esta implicação apresenta índice de coesão equivalente a 0,959.

7E→(6D→2A): A análise documental (currículos, ementas de disciplinas de cursos de pedagogia, arquivos de plataformas digitais, por exemplo) implica na impossibilidade de estabelecer associações, entre as interfaces da relação ao saber matemático dos futuros professores, que encontram-se em formação para atuar nos anos iniciais do ensino fundamental. Tal implicação se justifica porque o contexto destas investigações é a formação inicial dos pedagogos. Mais precisamente, os pesquisadores buscam perceber como as investigações, por eles realizadas, podem contribuir para a qualidade ou a melhoria dos cursos de formação de professores. O índice de coesão, nesta implicação corresponde a 0,929.

4D→(5D→CC): A pesquisa-ação representa a tipicalidade deste conjunto cujo índice de coesão é de 0,922 ou 92,2%. Todavia os autores que produziram este tipo de pesquisa não apresentaram as premissas ou as ações relativas ao(s) método(s) utilizado(s) na(s) análise(s) do(s) instrumentos que possibilitaram a construção dos dados. Tal aspecto é concernente às produções difundidas em eventos (ENEM, CIAEM e SIPEM), na modalidade das comunicações científicas.

5C→4B: A articulação de dois ou mais métodos de análise representa um índice de coesão de 0,82. Nas pesquisas analisadas predominam a análise do conteúdo aliada a análise do discurso. Tal variável implica que é característica comum às produções caracterizadas como estudos de caso.

7F→3B: Este agrupamento representa um percentual de 82% de coesão. A análise de como os professores avaliam ou interveem sobre as estratégias de resolução de problemas, apresentadas por seus alunos em testes ou avaliações, representa a tipicalidade deste jogo de variáveis. Tal elemento ganha significação associado ao ser associado as investigações acerca da metodologia dos professores que ensinam matemática nos anos iniciais. Neste caso, as pesquisas buscam esclarecer aspectos relativos às ações didáticas e a capacidade do professor de (re)direcioná-las para favorecer as aprendizagens acerca dos objetos matemáticos.

5A→6C: Com índice de coesão equivalente a 0,694 temos esta implicação cujo significado denota que ao adotar método a análise do discurso, os autores das publicações que nós analisamos acabam revelando as características da interface didática da relação ao saber do professor. Referimo-nos, como já foi mencionado anteriormente, às relações pessoais e institucionais que o professor estabelece com o saber que visa ensinar. Embora, não seja este o propósito dos pesquisadores, suas pesquisas revelam a influência do currículo, das políticas de ensino, do livro didático de

matemática e dos saberes experienciais na didática dos professores dos anos iniciais do ensino fundamental.

P→6B: As publicações em periódicos (tipicalidade do conjunto) implicam em discussões que trazem como plano de fundo a interface sociológica da relação ao saber do professor. Este profissional não está isolado, ele recebe a influência da cultura (escolar, profissional, formativa, etc.) que é fruto das próprias vivências; bem como, de seus pares (colegas de trabalhos, estudantes, gestores, formadores, por exemplo). Neste caso, o índice de coesão representa 67,1%.

9. O panorama das produções científicas acerca do professor que atua nos niveis iniciais do ensino fundamental

Ao realizarmos a revisão das publicações científicas divulgadas em eventos nacionais e internacionais sobre Educação Matemática; bem como, em periódicos categorizados como Qualis A e B, buscávamos identificar os focos de interesse, os processos metodológicos e os resultados das pesquisas desenvolvidas por pesquisadores brasileiros em relação aos professores que ensinam matemática nos anos iniciais do ensino fundamental. Dessa forma, estaríamos buscando subsídios que justificassem a autenticidade e a pertinência do projeto de tese (*Relação ao saber matemático de professores brasileiros e franceses que atuam no ensino fundamental*), que estamos desenvolvendo no PPGEC-UFRPE em cotutela com a Université Lumière – Lyon 2. Mas também, criando links entre o referido projeto e o conhecimento científico acerca dos professores que atuam nesta modalidade de ensino. Nesta pesquisa as análises estatísticas, com o suporte do software CHIC 6.0, foram realizadas com um banco de dados formado por 57 artigos científicos. Sendo que 48 destes representam comunicações científicas (ENEM, CIAEM e SIPEM) e 9 são artigos publicados em periódicos (BOLEMA, Ciência & Educação, ZETETIKÉ e Educação em Revista). A ASI revelou que 51% das pesquisas (29 publicações), foram produzidas por grupos de pesquisa ou pesquisadores vinculados às universidades (públicas ou privadas) e aos institutos privados de formação de professores, que estão concentrados na Região Sudeste. As pesquisas realizadas por instituições brasileiras das Regiões Nordeste e Sul aparecem no gráfico, ilustrado na Figura 5, entre as que mais contribuem para este nicho de pesquisa, com percentuais de 18% (10 publicações) e 14% (8 publicações), respectivamente. No que se refere ao contexto das investigações anteriormente mencionadas, a *formação continuada* foi a opção privilegiada em 40 das 57 publicações que analisamos. Isto corresponde a 70% das investigações realizadas com professores dos anos iniciais do ensino fundamental. Em se tratando do objeto da investigação, principal foco de interesse dos pesquisadores, 17 das 57 publicações (29%), focalizam o *conhecimento matemático* dos professores que ensinam nos anos iniciais. Enquanto, 13 das 57 pesquisas (22%) investigaram aspectos relativos à *metodologia de ensino da matemática* (resolução de problemas, modelagem, por exemplo).

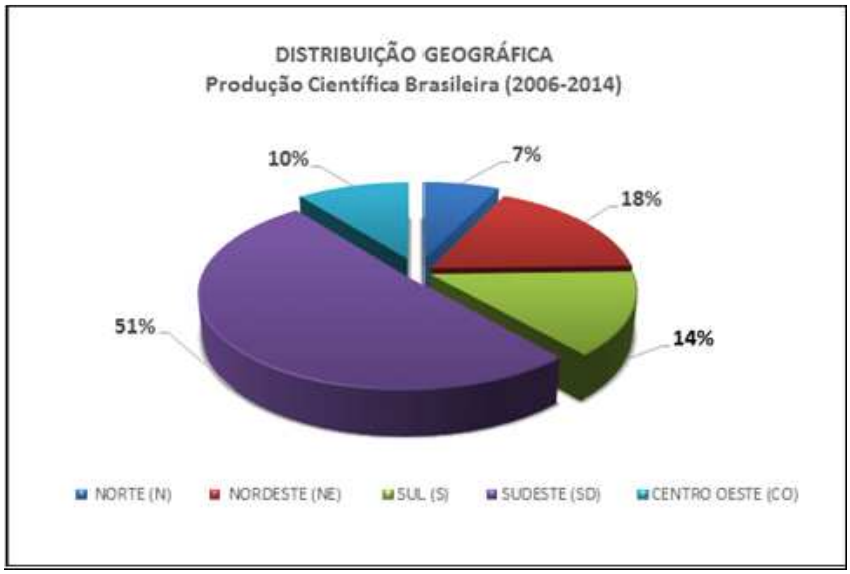


Figura 5: Distribuição percentual das publicações em eventos e periódicos

Por outro lado, 11 das 57 pesquisas (19%) o interesse estava voltado para os contributos da matriz curricular ou ainda, dos conteúdos das disciplinas acadêmicas dos cursos de formação de professores, essencialmente dos cursos de licenciatura em pedagogia. Neste caso, ao indicar os desafios e necessidades, concernentes ao processo formativo pelo qual passam os professores ao longo do processo de profissionalização, os pesquisadores auxiliam na elucidação dos fenômenos relativos à formação matemática dos professores. E, por conseguinte, fornecem subsídios para que outros pesquisadores analisem os possíveis efeitos dessa formação, quanto à obstaculização do ensino e/ou à mediação das aprendizagens dos objetos matemáticos.

No que diz respeito à tipologia da pesquisa, os dados mostram que 18 das 57 publicações analisadas (31%), podem ser categorizadas como *estudo de caso*. Em outras 14 publicações (24%) encontramos as diretrizes da pesquisa-participante intervencionista. Neste último caso, ressaltamos que a intervenção (oficinas pedagógicas, encontros formativos em serviço, cursos de extensão, etc.) é integralmente organizada, promovida e mediada pelo pesquisador. Em relação aos métodos, adotados pelos autores das publicações para analisar os dados, destacamos que em 33 das 57 pesquisas (57%) houve associação das estratégias metodológicas. Ou seja, nestes exemplares, a *análise do conteúdo* está associada às técnicas de análise do discurso ou a elementos concernentes à *etnografia*. Todavia, em 16 publicações (28%) trazem seus resultados analisados exclusivamente por meio da análise do conteúdo. Ao investigarem as questões ou hipóteses acerca dos seus objetos de estudos, os autores das publicações analisadas neste artigo, eles acabam trazendo à tona em suas abordagens, aspectos que têm impacto direto nas formas e na formação da relação ao saber matemático do professor. Ou seja, em 33 das 57 publicações, as discussões que trazem o *viés didático da relação ao saber* como plano de fundo na argumentação aparecem em psicanalítico. Em outras 16 publicações, o *viés sociológico* aparece por meio de abordagens que ressaltam a influência das relações interpessoais e da cultura escolar, no modo como o

professor se relaciona com a matemática que ensina nos anos iniciais do ensino fundamental. Entretanto, o viés psicanalítico não aparece nos argumentos dos autores dos artigos que selecionamos e analisamos. Também não foi possível identificar qualquer aspecto característico de uma das interfaces da relação ao saber no conteúdo e na discussão dos resultados de 8 publicações (14%). A construção dos dados analisados pelos autores das 57 publicações tomou como referência o cruzamento das informações obtidas mediante a utilização de *múltiplos instrumentos de pesquisa.* Neste caso, técnicas de observação dirigida (diário de campo), associadas ao uso de questionário, entrevistas ou aos relatos de experiência/vivências do professor (em sala de aula ou em momentos de formação). Isto é recorrente em 25 publicações (43%). Em 17% dos casos (10 das 57 publicações), os resultados são apresentados com base na aplicação de um único *questionário*, predominantemente constituído por questões fechadas e de múltipla escolha.

10. Considerações finais

O panorama das pesquisas sobre os professores que atuam nos primeiros anos da escolaridade que apresentamos neste texto explicita o crescente interesse das pesquisas, realizadas no Brasil entre 2006 e 2014, sobre aspectos que permeiam a formação e a prática deste profissional. O retrato da realidade, observada a partir das publicações que selecionamos e analisamos, encontra-se justificada nos argumentos de D'ambrósio (2011, p.83) quando afirma que:

« [...] A educação enfrenta em geral grandes problemas. O que considero mais grave, e que afeta particularmente a educação matemática de hoje, é a maneira deficiente como se forma o professor. Há inúmeros pontos críticos na atuação do professor, que se prendem a deficiências na sua formação. Esses pontos são essencialmente concentrados em dois setores: falta de capacitação para conhecer o aluno e obsolescência dos conteúdos adquiridos. »

Entretanto, o volume de publicações difundidas em eventos e periódicos, acerca da formação ou dos saberes dos professores, que ensinam (ou ensinarão) matemática nos anos iniciais do ensino fundamental, ainda é incipiente se comparado ao quantitativo de estudos cujos participantes são licenciandos ou licenciados em matemática. Muito embora, este campo de investigação seja promissor, ainda não despertou o interesse dos estudiosos no que diz respeito à relação ao saber matemático dos professores que atuam nos anos iniciais do ensino fundamental. Assim sendo, a noção de relação ao saber matemático dos professores polivalentes se apresenta para nós como nicho de pesquisa. Para Charlot (2008) as pesquisas sobre a relação ao saber buscam compreender como o sujeito apreende o mundo e, com isso, como se constrói e transforma a si próprio: um sujeito indissociavelmente humano, social e singular. Para tanto, partimos do pressuposto que toda relação implica uma correspondência entre dois elementos. No caso da relação ao saber, referimo-nos à ligação peculiar que há entre um sujeito (humano) e o objeto do conhecimento (saber matemático). Essa ligação pressupõe um distanciamento que sofre múltiplas influências de caráter social, institucional e psíquico. Para Charlot (2000, p.81) A relação é que se particulariza, não é o objeto da relação que se torna particular: o desejo do mundo, do outro e de si mesmo é que se torna desejo de aprender e saber. E, não, o "desejo" que encontra um objeto novo, "o saber". Nesse

sentido, percebemos que a necessidade de satisfazer o desejo de ser (professor), de fazer (ensino), de mediar (aprendizagens) e de produzir (conhecimento), é o que mobiliza o professor para o ensino e promove o acionamento de suas praxeologias (didática e matemática). Porém, todos estes esforços para satisfazer o(s) desejo(s) do professor não estão imunes às influências psíquicas (sentimentos, comportamentos e atitudes conscientes e inconscientes), sociológicas (relações culturais, sociais e institucionais) e didáticas (a organização didática e pedagógica; os materiais e métodos de ensino; e, o legítimo poder conferido, que transformam o professor em instituição). E, diante da complexidade destas variáveis, nós nos propusemos a desenvolver uma tese sobre a relação ao saber matemático dos professores que atuam no início da escolaridade, tanto no Brasil como na França. Com este intuito, dedicamos nossos esforços para descobrir, por meio das análises das publicações em eventos e periódicos, o teor das investigações realizadas pelos pesquisadores brasileiros acerca dos professores que ensinam matemática nos anos iniciais do ensino fundamental. Nesse sentido, os dados submetidos a ASI realizada com o CHIC 6.0, possibilitou concluir que:

- As pesquisas sobre a relação ao saber dos professores dos anos iniciais do ensino fundamental não avançaram. Uma vez que, no intervalo compreendido entre 2006 e 2014, não identificamos artigos com este foco específico. Assim sendo, a construção de modelo teórico sobre este fenômeno, proposta em nossa tese de doutoramento suscita a reflexão e esclarecimento sobre a temática. Tal proposição garante a nossa produção a autenticidade e a originalidade que se preza em um contributo científico.

- Embora, as publicações analisadas, proponham reflexões críticas acerca da (re)significação dos saberes docentes, construídos pelos professores dos anos iniciais, ao longo da formação (inicial, continuada e em serviço); suscitem a implementação de novos modelos formativos; proponham a auto avaliação tanto por parte dos professores como dos formadores de professores; e fomentem a revisão dos papéis daqueles que formam e dos que são formados. É preciso pesquisar também, como e em que bases se institui a relação ao saber matemático destes profissionais. Pois, ao compreender melhor tal fenômeno, criaremos argumentos mais seguros, para propor o (re)direcionamento das ações educativas em sala de aula, inclusive em relação ao ensino e a aprendizagem da matemática.

- Há uma preocupação latente (quase coletiva), por parte dos autores nos artigos analisados, quanto à identificação e descrição dos aspectos relativos aos saberes apreendidos (ao longo da profissionalização), requeridos (no exercício da docência) e acionados (presentes na prática) pelos professores que ensinam (ou ensinarão) matemática nos Anos Iniciais do Ensino Fundamental. Tal motivação é compreensível e pertinente. Todavia, mais do que buscar conhecer/compreender sobre o que os professores sabem (saberes a serem ensinados), e, como eles fazem, (conhecimento experiencial oriundos da prática ou resultantes da formação) é essencial nos preocuparmos também quanto as formas e a formação da relação que estabelecem com o objeto de ensino (saberes

matemáticos). Dessa forma, podemos contribuir e discutir acerca da (re)estruturação da formação matemática dos profissionais.

REFERÊNCIAS

[1] Acioly-Régnier, N. M., Régnier, J. C. Transparence pour comprendre, transparence pour agir : Questions autour d'un artefact informatique dans la formation à l'A.S.I. *Quaderni di Ricerca in Didattica (Mathematics*), n°20 suppl 1, 2010.

[2] Bardin, L. (2011), *Análise de conteúdo*. 5ª. Edição. Lisboa: Edições 70.

[3] Beillerot, J. Blanchard-Laville, C. ; Mosconi, N. (1989), *Savoir et rapport au savoir. Élaborations théoriques et cliniques*. Paris, Bégédis: Éditions Universitaires,

[4] Câmara dos Santos, M. (1997), *A relação do conhecimento do professor de matemática em situação didática: uma abordagem pela análise do seu discurso*. In: Anais doXIII Encontro de Pesquisa Educacional do Nordeste. Rio Grande do Norte: UFRN. p.101 – 109.

[5] Charlot, B. (2008), *Relação com o saber, formação de professores e globalização. Questões para a educação hoje*. Dados eletrônicos. Porto Alegre, RS: Artmed.

[6] Couturier, R., Bodin, A., Gras, R. (2004). *A classificação hierárquica implicativa e coesiva*. *Manual Curso CHIC versão*, 2.

[7] Couturier, R, Gras, R. (2013) *Spécificités de l'Analyse Statistique Implicative par rapport à d'autres mesures de qualité de règles d'association*. Educação Matemática Pesquisa, São Paulo, v.15, n.2, pp.249-291.

[8] D'ambrósio, U. (2011). *Educação Matemática: da teoria à prática*. São Paulo: Editora Papirus.

[9] Gras, R. & Ag Almouloud, S. (2012), *A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais. Revista Educação Matemática Pesquisa*, *4*(2-2002).

[10] Imbernón, F. (2009), *Formação Permanente do professorado.* Novas tendências. São Paulo, SP: Cortez.

[11] Imbernón, F. (2010), *Formação continuada de professores*. Porto Alegre: Artmed.

[12] Nóvoa, A. (1995) Os professores e sua formação. Porto: Porto Editora.

[13] Tardif, M. (2000), Saberes profissionais dos professores e conhecimentos universitários: Elementos para uma epistemologia da prática em relação à formação para o magistério. *Revista Brasileira de Educação*, v. 5, p. 5–24.

[14] Tardif, M. (2011), Saberes docentes e formação profissional. Petrópolis: Editora Vozes.

[15] Anais do Encontro Nacional de Educação Matemática – ENEM. www.sbembrasil.org.br/sbembrasil/index.php/anais/enem - Último acesso em: 10 de abril de 2015.

[16] Anais do Simpósio Internacional de Educação Matemática www.sbembrasil.org.br/sbembrasil/index.php/anais/sipem- Último acesso em: 10 de abril de 2015.

[17] Revista Zetetiké. www.fae.unicamp.br/revista/index.php/zetetike - Último acesso em: 10 de abril de 2015.

[18] Boletim de Educação Matemática (BOLEMA). www2.rc.unesp.br/bolema - Último acesso em: 10 de abril de 2015.
[19] Revista Ciência & Educação. www2.fc.unesp.br/cienciaeeducacao/ - Último acesso em: 10 de abril de 2015.

ANEXOS

ANEXO A: Tratamentos – CHIC 6.0

n col : 28, n fil : 57			
Variável	**Ocorrência**	**Média**	**Desvio padrão**
CC	48.00	0.84	0.36
P	9.00	0.16	0.36
V2A	17.00	0.30	0.46
V2B	40.00	0.70	0.46
V3A	24.00	0.42	0.49
V3B	9.00	0.16	0.36
V3C	10.00	0.18	0.38
V3D	10.00	0.18	0.38
V3E	4.00	0.07	0.26
V4A	6.00	0.11	0.31
V4B	18.00	0.32	0.46
V4C	4.00	0.07	0.26
V4D	18.00	0.32	0.46
V4E	11.00	0.19	0.39
V5A	7.00	0.12	0.33
V5B	18.00	0.32	0.46
V5C	14.00	0.25	0.43
V5D	16.00	0.28	0.45
V6A	0.00	0.00	0.00
V6B	16.00	0.28	0.45
V6C	33.00	0.58	0.49
V6D	8.00	0.14	0.35
V7A	10.00	0.18	0.38
V7B	5.00	0.09	0.28
V7C	9.00	0.16	0.36
V7D	25.00	0.44	0.50
V7E	4.00	0.07	0.26
V7F	4.00	0.07	0.26
2006 s	8.00	0.14	0.35
2007 s	4.00	0.07	0.26
2009 s	8.00	0.14	0.35
2010 s	5.00	0.09	0.28
2011 s	9.00	0.16	0.36
2012 s	9.00	0.16	0.36
2013 s	12.00	0.21	0.41
2014 s	2.00	0.04	0.18
N s	4.00	0.07	0.26
NE s	10.00	0.18	0.38
S s	8.00	0.14	0.35
SD s	29.00	0.51	0.50
CO s	6.00	0.11	0.31

ANEXO B: Planilha Geral – Conjunto de dados (1A)

	ANO DA PUBLICAÇÃO								REGIÃO DO BRASIL					TIPOLOGIA		CONTEXTO		OBJETO DA INVESTIGAÇÃO				
ORIGEM	**2006 s**	**2007 s**	**2009 s**	**2010 s**	**2011 s**	**2012 s**	**2013 s**	**2014 s**	**N s**	**NE s**	**S s**	**SD s**	**CO s**	**CC**	**P**	**V2A**	**V2B**	**V3A**	**V3B**	**V3C**	**V3D**	**V3E**
IX ENEM1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0
IX ENEM2	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0
IX ENEM3	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0
IX ENEM4	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0
X ENEM1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0
X ENEM2	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0
X ENEM3	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0
X ENEM4	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
XI ENEM1	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
XI ENEM2	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0
XI ENEM3	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
XI ENEM4	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0
XI ENEM5	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0
XI ENEM6	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1
XI ENEM7	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0
XI ENEM8	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
XI ENEM9	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0
XI ENEM10	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0
XI ENEM11	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1
XIII CIAEM1	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0
XIII CIAEM2	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0
XIII CIAEM3	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0

XIII CIAEM4	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
XIII CIAEM5	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0
XIII CIAEM6	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
XIII CIAEM7	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
XIII CIAEM8	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0
III SIPEM1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0
III SIPEM2	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0
III SIPEM3	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0
III SIPEM4	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0
III SIPEM5	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0
III SIPEM6	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0
III SIPEM7	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0
III SIPEM8	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0
IV SIPEM1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0
IV SIPEM2	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
IV SIPEM3	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1
IV SIPEM4	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
IV SIPEM5	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
IV SIPEM6	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1
IV SIPEM7	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0
IV SIPEM8	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
V SIPEM1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0
V SIPEM2	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
V SIPEM3	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0
V SIPEM4	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0
V SIPEM5	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0
BOLEMA1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0

BOLEMA2	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0
BOLEMA3	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0
BOLEMA4	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	0	0	0
BOLEMA5	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0
C&E	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	0
ER	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	0	0	0
ZETETIKÉ1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0
ZETETIKÉ2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0

ANEXO C: Planilha Geral – Conjunto de dados (1B)

	TIPIFICAÇÃO					MÉTODO				ABORDAGEM				INSTRUMENTOS					
ORIGEM	V4A	V4B	V4C	V4D	V4E	V5A	V5B	V5C	V5D	V6A	V6B	V6C	V6D	V7A	V7B	V7C	V7D	V7E	V7F
IX ENEM1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0
IX ENEM2	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1
IX ENEM3	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1
IX ENEM4	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
X ENEM1	0	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0
X ENEM2	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
X ENEM3	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
X ENEM4	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0
XI ENEM1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
XI ENEM2	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
XI ENEM3	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	0
XI ENEM4	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
XI ENEM5	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0
XI ENEM6	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0
XI ENEM7	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0
XI ENEM8	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0
XI ENEM9	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0
XI ENEM10	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0
XI ENEM11	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
XIII CIAEM1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0
XIII CIAEM2	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
XIII CIAEM3	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
XIII CIAEM4	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
XIII CIAEM5	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0

XIII CIAEM6	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0
XIII CIAEM7	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
XIII CIAEM8	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
III SIPEM1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
III SIPEM2	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
III SIPEM3	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0
III SIPEM4	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
III SIPEM5	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
III SIPEM6	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
III SIPEM7	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
III SIPEM8	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0
IV SIPEM1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
IV SIPEM2	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0
IV SIPEM3	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0
IV SIPEM4	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
IV SIPEM5	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0
IV SIPEM6	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
IV SIPEM7	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0
IV SIPEM8	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
V SIPEM1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
V SIPEM2	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0
V SIPEM3	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
V SIPEM4	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0
V SIPEM5	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0
BOLEMA1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0
BOLEMA2	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1

BOLEMA3	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
BOLEMA4	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0
BOLEMA5	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0
C&E	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0
ER	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0
ZETETIKÉ1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
ZETETIKÉ2	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0

THE INFLUENCE OF THE NATURE OF GEOMETRICAL FIGURES ON GEOMETRIC PROOFS AND THE ROLE OF GEOMETRICAL FIGURE APPREHENSION

Paraskevi MICHAEL – CHRYSANTHOU[1] and Athanasios GAGATSIS[2]

INFLUENCE DE LA NATURE DES FIGURES GEOMETRIQUES SUR LES PREUVES GEOMETRIQUES ET LE ROLE DE L'APPROCHE DE LA FIGURE GEOMETRIQUE

ABSTRACT

The purpose of this paper is to examine the way the nature of geometrical figures influences the students' solution of geometrical proof tasks and trace other factors that may affect the students' behavior when dealing with figures that can be either consistent or inconsistent to the given properties of the problem. This examination was based on two theoretical frameworks that deal with the nature (Mesquita, 1998) and the apprehension of geometrical figures (Duval, 1995). Based on these frameworks two groups of geometrical tasks were administered to 706 secondary school students. The results revealed that perception blocks the way the students look at figures and does not allow the use of the given properties. On the other hand, the operative apprehension can enable students overcome the obstacle of the inconsistency between the given properties and the given figure. The heuristic visualization of figures allows students to draw extra lines or draw a figure representing correctly the given properties and succeed the solution of the tasks.

Keywords : geometrical figures, nature, object, image, proofs.

1 Introduction

Geometry involves the mixing of two registers: the figure register in order to 'see' and the natural language register in order to 'explain' (Duval, 2002). According to Hanna (2000) diagrams and other visual aids have been used as facilitators of understanding and as heuristic accompaniments to proof, inspiring both the theorem to be proved and approaches to the proof itself. Thus, the learners have to work in another register than language and to come back to linguistic expression. The way of looking at any figure constructed with specific tools is a crucial cognitive factor in solving problems and in reasoning and proving in geometry. Figures are the blind spot in mathematics education as much for theories as for teaching. There is an equivocal use of the verb "to see" regarding figures in geometry, because there are two levels of cognitive functioning or recognition, which are not distinguished. Therefore, a strong incongruence often occurs between what is seen and

[1] University of Cyprus, Department of Education, P.O.Box 20537, 1678, Nicosia, Cyprus, pmicha@ucy.ac.cy

[2]University of Cyprus, University House "Anastasios G. Leventis", 1 Panepistimiou Avenue, Aglantzia, Nicosi, P.O. Box 20537, 1678 Nicosia, Cyprus, gagatsis@ucy.ac.cy

what is named or stated in the utterance of the problem (Duval, 2014). Consequently, a wide gap is created in this case, giving rise to recurrent and very often insuperable difficulties for most students. There is obviously a need to identify the cognitive processes that students mobilize during the resolution of geometrical tasks and examine whether these processes lead to a proper way of "looking" at geometrical figures.

It is well known to mathematics teachers that many students use geometrical figures as images (Mesquita, 1998), and not as a mathematical object that depends on definitions, properties, theorems etc. According to Duval (2011), a geometrical figure includes two levels of recognition: the figure and the object recognition. The first level corresponds to the natural perceptive way as for any visual representation of material objects or spatial organization (images, diagrams etc.) and the object recognition to the mathematical way for reasoning, defining, and problem solving or proving (Duval, 2011). Thus, the difficulties for most students are created due to a cognitive gap between these two opposite ways of looking at figures and recognizing what they stand for. Therefore, teaching should be organized to make students aware of the ways of how to see and think in a relevant way in elementary geometry.

For this educational goal, research should focus on the cognitive conditions that allow students to learn to seek and find out a solution of any geometry problem by themselves. Previous researcher studies examined factors that hamper the heuristic visualization of figures (Michael – Chrysanthou & Gagatsis, 2015; 2013) and the influence of the nature of geometrical figures in problem solving (Xistouri, Nicolaou, Koukkoufis, & Gagatsis, 2005, Gagatsis, Michael, Deliyianni, Monoyiou & Kuzniak (2011). However, there is a lack of studies examining how the different types of figural apprehension have an impact on the way the nature of geometrical figures influences the students' solutions. Based on the above, the aim of this research is to analyze the way students in secondary education look at a geometrical figure, according to its nature (either as an object or just as an image) in relation to the type of figural apprehension students' mobilize for the solution of geometrical tasks. For this investigation a number of geometric tasks are used on the basis of two theoretical frameworks, which are described further on.

2 Theoretical Framework

This study is based on a synthesis of two theoretical frameworks in order to analyze and interpret students' solution approaches in geometrical tasks. The first one concerns the distinction between four kinds of figure apprehension proposed by Duval (1995, 2005) within the register of geometrical visualization: perceptual apprehension, sequential apprehension, operative apprehension and discursive apprehension. In this study the focus is on the operative apprehension, which constitutes the heuristic reconfiguration (mereologic modification) of the recognized shapes into other shapes. Duval (1999) stresses the importance of reconfiguration of geometrical figures, as it goes against the perceptual recognition of figures.

The second theoretical framework focuses on Mesquita's discrimination between different types of geometrical figures or configurations based on their nature, according to the possibility of inferring or not geometrical relationships from it. According to Mesquita (1998), an external representation has the nature of an *object* when it is possible to infer geometrical relationships from the construction of the figure that may be used in geometrical reasoning and proof and when the visual perception of the figure is consistent with the verbal statements of the problem. On the contrary, when the external representation has the nature of an *illustration*, it is impossible to directly extract a geometrical relationship from the construction of the figure, the figure seems to 'mislead' and the visual perception of the figure is in contradiction with the verbal statements. The results of Gagatsis, Michael, Deliyianni, Monoyiou and Kuzniak (2011) show the strong impact of the nature of the geometrical figure on the students' understanding of proof.

Based on the synthesis of these two theoretical frameworks, the following research questions are addressed:

- How are students' responses to geometrical tasks differentiated when a geometrical figure is represented as an image or as an object?
- How does the type of geometrical figure apprehension students use contribute to an appropriate processing of geometrical figures as objects or as images in proofs?

3 Methodology

The participants were 706 students, aged 15 to 17, of lower (Grade 9) and upper (Grade 10, Grade 11) urban and rural secondary schools, in Cyprus. In particular, the participants were 312 students from Grade 9, 304 students from Grade 10 and 140 students from Grade 11. It shall be mentioned that the students in Grade 11 followed a scientific orientation, thus they were attending more hours of mathematics classes, with more advanced mathematical content. Based on our two main theoretical choices, two groups of tasks were used, which are presented in the next sections (Figures 1 and 2). Each group of tasks was used for answering a specific research question, as shown below.

4 Results

4.1 How are students' responses to geometrical tasks differentiated when a geometrical figure is represented as an image or as an object?

In relation to the first research question stated above, tasks that examined the impact of the nature of the geometrical figure (Mesquita, 1998) in students' geometrical proof – problem solving ability were used. Based on Mesquita's distinction of the nature of geometrical figures, two pairs of corresponding tasks were developed (figure 1). In each pair the first task included a figure which was consistent with the given properties in the wording of the problem, thus having the nature of an 'object' (OB1and OB2). The second

task included a figure which was inconsistent to the given properties, thus having the nature of an 'image' (IM1 and IM2). Students solved first the tasks OB1 and OB2 and they were given the tasks IM1 and IM2 at a different time.

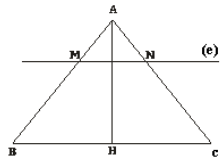
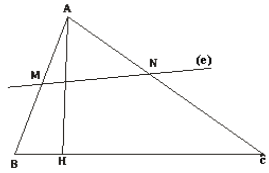
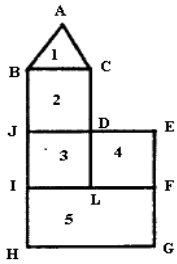
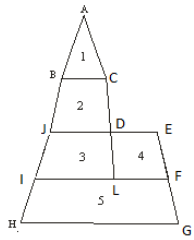
In the triangle ABC: AB and AC are equal, line (e) is parallel to BC and AH is perpendicular to BC. Choose the right answer, based on the data provided. The length of NH: a) is equal to MH b) is bigger than MH c) is smaller than MH d) it cannot be determined Explain your answer.	 (OB1)	 (IM1)
In the given figure: • figure 1 is an equilateral triangle • figure 2 is a rectangle • figures 3 and 4 are squares • the figure that consists of figures 3, 4 and 5 is a square. Show that sides AC, LF and FG are equal.	 (OB2)	 (IM2)

Figure 1 – Pairs of tasks including geometric figures or configurations as objects and as images respectively

The codification of students' answers in these tasks is shown in Table 1.

Pair 1	Pair 2
OB1: correct solution	OB2: correct solution
OB1j: correct justification of answer OB1drl: draw extra lines on the figure	OB2j: correct justification of answer OB2drl: draw extra lines on the figure
IM1: correct solution	IM2: correct solution
IM1j: correct justification of answer	IM2j: correct justification of answer
IM1dr: draw the correct figure	IM2dr: draw the correct figure
IM1drl: draw extra lines on the figure	IM2w: verbal expression about the inconsistency of figure to the wording of problem
IM1w: verbal expression about the inconsistency of figure to the wording of problem	

Table 1 – Codification of students' answers in the tasks including a figure as an object or as an image

In Pair 1 the correct answer was choice a (OB1). Students' answers explaining correctly the procedure they followed for solving the problem were codified as OB1j. In Pair 2 students were expected to make an inference, based on the properties of the geometrical figures and prove the equity of the three sides. A correct solution was considered when

students proved the equity of all three sides. Students' answers showing explicitly the transitivity of equality of different sides based on the properties of the geometrical figures were were codified as OB2j. Students' extra lines on the figure and expressions about the inconsistency of figure to the wording of problem were also taken into account.

To address the research question about the influence of the representation of a geometrical figure as an image on students' responses in geometrical tasks we will describe the students' solutions in the two pairs of tasks described above. Table 2 indicates the percentages of the different categories of students' answers.

		CORRECT ANSWER		WRONG ANSWER		NO ANSWER	
		OBJECT	IMAGE	OBJECT	IMAGE	OBJECT	IMAGE
PAIR 1	G9	57,1	15,7	11,2	44,6	31,7	39,7
	G10	61,2	15,1	7,2	52,0	31,6	32,9
	G11B	82,9	39,3	7,1	45,7	10,0	15,0
PAIR 2	G9	21,2	11,2	21,2	18,6	57,7	70,2
	G10	26,0	12,8	30,9	23,7	43,1	63,5
	G11B	57,1	40,0	20,7	30,7	22,1	29,3

Table 2 – Students' answers in the tasks examining the nature of figures by grade

In each pair of tasks, there are more correct answers in the case the figure has the nature of an object than in the tasks in which the figure has the nature of an image. These differences are greater in the first pair of tasks, compared to the second pair. This is the case for all the groups of students. The students' incorrect answers in the first task including a figure as an object are less than in the corresponding tasks including the figure as an image. This is observed in all groups of students. On the other hand, this is not always the case for the second pair of tasks, as in grade 9 and grade 10 the wrong answers when the figure has the nature of an image are less than in the task including a figure as an object. In the case the students do not give an answer, there are more students reacting in this way in the case the figure has the nature of an image, than in the case the figure has the nature of an object. This suggests that figures having the nature of an image increase the students' resistance for trying to solve the tasks.

4.2 How does the type of geometrical figure apprehension students use contribute to an appropriate processing of geometrical figures as objects or as images in proofs?

Regarding the second research question, the relation between the students' figural apprehension and the influence of the nature of the figure were examined. The purpose was to trace which way of looking at geometrical figures helps the students overcome the obstacles created to students when figures have the nature of an image: the perceptual way

or the heuristic visualization of figures? For succeeding this, the same students were administered a second group of tasks that examined the operative apprehension of geometrical figures (Figure 2). Actually, these tasks test students' ability to modify a geometrical figure, as they ask for a reconfiguration of the given geometrical figure in order to be solved.

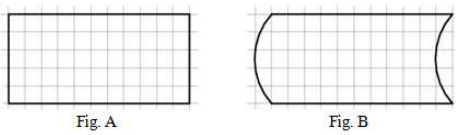
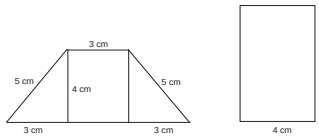
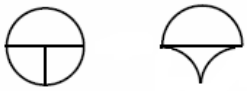
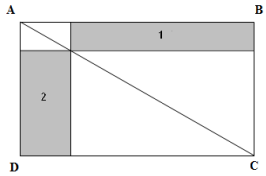
(OP1) Underline the right sentence and explain your answer. a) Fig. A has an equal area to Fig. B b) Fig. A has a smaller area than Fig. B c) Fig. A has a bigger area than Fig. B  Fig. A Fig. B	(OP2) The trapezium and the rectangle have equal areas. Find the length of the missing side of the rectangle and explain your answer. 
(OP3) Look at the figures and circle the correct answer and explain your choice. a) Figure A has a bigger perimeter than figure B. b) Figure A has an equal perimeter to figure B. c) Figure A has a smaller perimeter than figure B. 	(OP4) Figure ABCD is a rectangle. Look at the shadowed rectangles 1 and 2 and choose the correct answer. Then justify your choice.  a. Rectangle 1 has a bigger area than rectangle 2. b. Rectangle 1 has an equal area to rectangle 2. c. Rectangle 1 has a smaller area than rectangle

Figure 2 – Tasks examining the operative apprehension of geometrical figures

Students' answers were grouped into two categories based on the type of geometrical figure apprehension that could be inferred from their written inscriptions in their solution procedures. These categories are:

- Answers from the operative apprehension, using the mereologic modification (coded as OPme, followed by the number of the task). For example in the task OP1 students were performing a reconfiguration of Figure B, showing that when the front part of the figure is cut and moved to the back part of the figure, a rectangle is formed, similar to Figure A.

- Answers from the perceptual apprehension (coded as OPpe, followed by the number of the task). For example in the task OP1, students used counting to find the answer or tried to calculate the perimeter by estimating the length of each sector.

In fact, the variables used in the following data analysis represent the students' correct answers that occur from each type of approach.

For examining the relations between the students' behavior in the solution of the tasks of the two groups, the implicative analysis was performed, using the software CHIC (Classification Hiérarchique, Implicative et Cohésitive) (Bodin, Coutourier, & Gras, 2000). The implicative statistical analysis (Gras, Régnier, Marinica & Guillet, 2013) aims at giving a statistical meaning to expressions like: "if we observe variable A in a subject, then in general we observe variable B in the same subject". Thus, the underlying principle of the implicative analysis is based on the quasi-implication: "if A is true, then B is more or less true". An implicative diagram represents graphically the network of the quasi-implicative relations among the variables of the set V. In this study the implicative diagrams contain implicative relations, indicating whether success at a specific task implies success at another task related to the former.

The first implicative diagram (Figure 3) represents the relations between the students' answers in the operative apprehension tasks and their success in the two pairs of tasks examining the nature of the figure. The relations reveal that the students' solutions in the operative apprehension tasks that occur involve the mereologic modification of the given figures (OP3me, OP5me and OP2me) form implicative relations with the solution of the two tasks in which the given figure has the nature of an image (IM1 and IM2). On the other hand, at a second part of this implicative diagram, implications are found mainly between the perceptual answers in the operative apprehension tasks and the solution of the tasks that include figures having the nature of an object (OP2pe and OB2 / OP3pe and OB1). However, the solution of the tasks including figures as objects form also an implicative relation with the use of the mereologic modification in an operative apprehension task (OP1me and OB1). In addition, we observe that there is a consistency in the students' success in the tasks examining the influence of the nature of figures. Specifically, when the students are able to solve correctly the second task including a figure as an image (IM2), they also solve correctly the first task including a figure as an image (IM2). This is also the case for the two tasks including a figure as an object (OB2 and OB1).

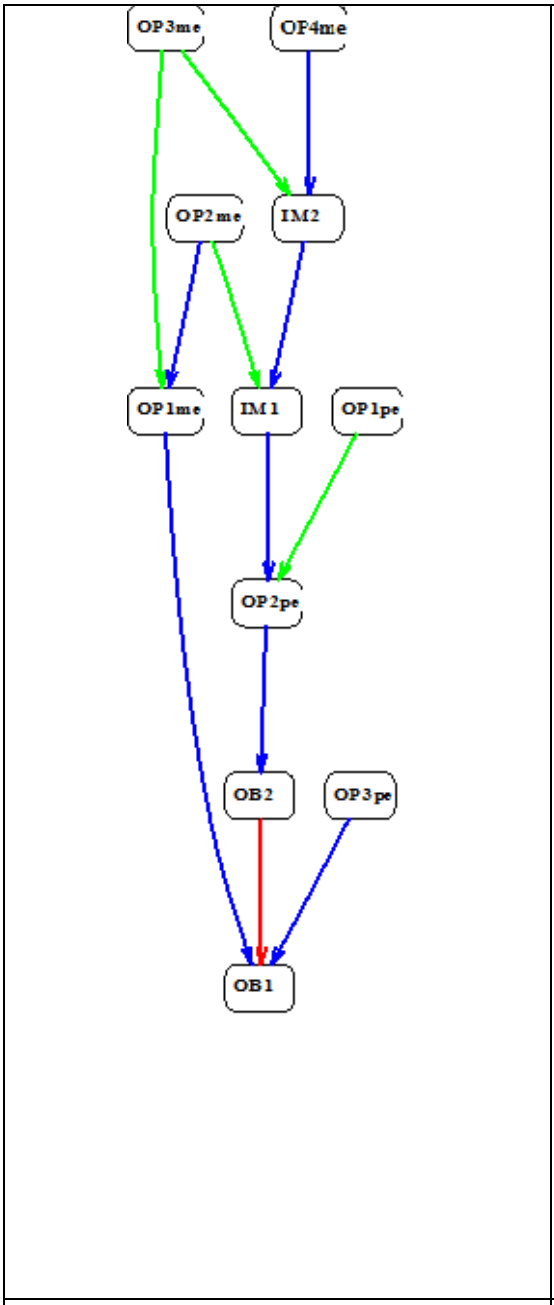


Figure 3 – Implications between the students' answers in the operative apprehension tasks and in the tasks examining the nature of figures.

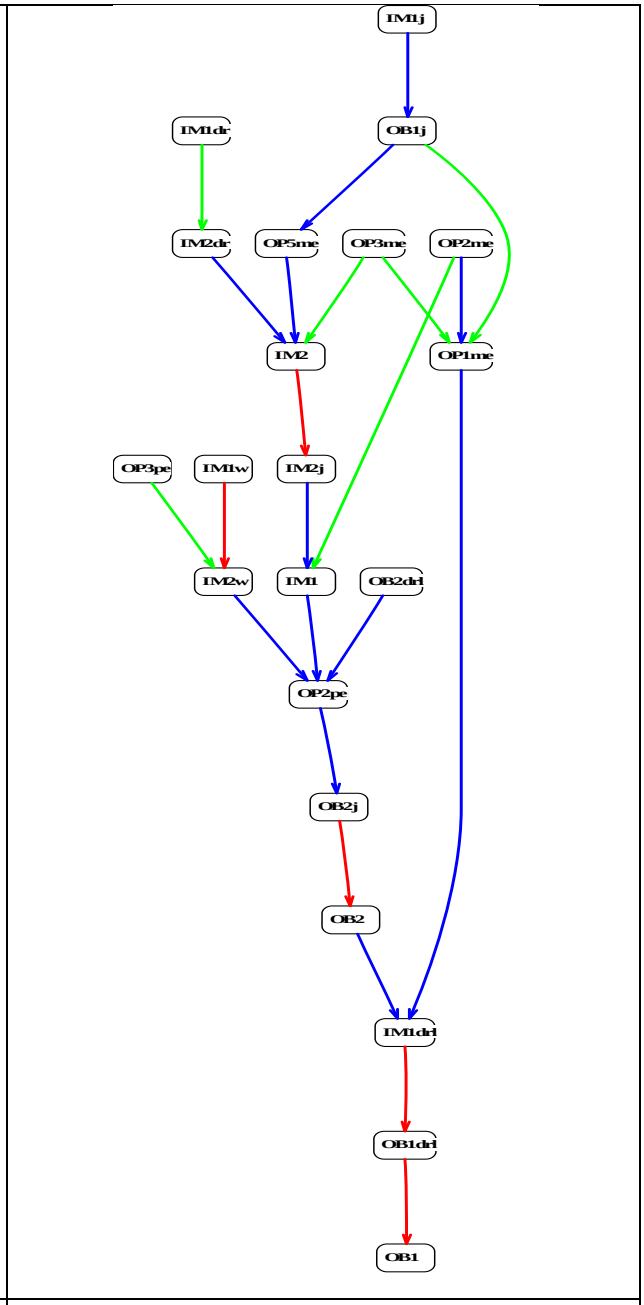


Figure 4 – Implications between the students' strategies in all the tasks.

Further on, the relations between the students' approaches in the operative apprehension tasks and their responses in solving the two pairs of tasks about the role of the nature of figures were examined. Based on a global look at the second implicative diagram, it is observed that while all four mereologic apprehension answers are involved, there are only

two perceptual apprehension answers that are involved in the total set of implicative relations. Based on the first relation appearing in figure 4, we can observe that students' behaviour in providing a justification for their answer is coherent for the first pair of tasks (OB1J and IM1j). It seems that when the students are able to provide a justification for the task including a figure as an image, they are also able to justify their answer in the corresponding task with the figure as an object. These variables are related to two variables representing a solution based on the mereologic modification of figures (OP1me and OP5me), indicating that the operative apprehension intervenes in the process of justifying an answer in the case the given figure is inconsistent to the given properties.

Furthermore, the students' tendency to draw a correct figure in the tasks the figure has the nature of an image can be also characterised by consistency (IM1dr and IM2dr). And when the students are able to draw a figure consistent to the given properties they are enhanced for succeeding the solution of such a task (IM2dr and IM2) and also for justifying their answer (IM2j). Similarly, using the mereologic modification of geometrical figures appears to enhance students' success in solving tasks with figures as images and also the justification of their answer (OP5me, IM2 and IM2j; OP3me, IM2 and IM2j; OP2me and IM1). Thus, both the drawing of appropriate figures and the fluency in applying the mereologic modification to geometrical figures contribute to students' achievement in geometrical tasks which include figures as images. Answers based on the mereologic modification (OP1me) seem to have a role also in drawing extra lines either on the figure as an object (OB1dr1) or as an image (IM1dr1). However, an answer through the perceptual apprehension (OP2pe) is also related to drawing extra lines on the figure having the nature of an image (OB2drl).

Moreover, perceptual answer in the operative apprehension tasks (OP3pe) is related to the recognition that a given figure is inconsistent to the wording of the problem, showing the role of perception in recognizing the incoherence between the figure and the given properties. Another relationship of the perceptual approach this time with the justification and the correctness of an answer in a geometrical task having the figure as an object indicates the important contribution perception had in the solution of the particular task.

5 Discussion

Based on the results from the students' answers in the tasks of the research, the discussion will be formed according to the two main questions that were posed about the way students respond when they deal with figures having the nature either of an image or of an object and about the contribution of the type of geometrical figure apprehension students adopt on these approaches.

The results regarding the students' answers in the tasks examining the nature of the figure suggest that the nature of the figure influences the students in the solution of the tasks. In the case the figure has the nature of an object and it is consistent with the properties given in the task, the students' proving and performance are at a higher level than it the tasks the figure has the nature of an image. An explanation for this finding is that, when the figure is inconsistent to the properties given in the wording of the problem,

students remain restricted within a perceptual way of looking at the figure and use incorrect treatments based on measures and proportionality (Mesquita, 1998). However, another important question we should reflect on further on is about the nature of the figure the students draw based on the utterance of the given problem. Will this figure have the nature of an object, an image or a mixed status, by image allowing the illustration of the properties of the figure?

As shown also from the implicative diagram, when a figure is represented as an image good perception of figures is basic for recognizing its incoherence to the given properties. But despite the fact that the students are able to recognize this inconsistency, they are not able to proceed to the solution of the tasks using only the given properties from the utterance of the problem. As a result, perception overrides the use of the given properties. This reveals the compartmentalization (Gagatsis, 2011) of the students' cognitive process between a mathematical and a perceptual way of looking at figures, as they focus on the given figure without trying to verify that the information they get from it are coherent with the properties given verbally. Thereafter, for the students that look at the geometrical figure through perception, the figure is mostly seen as a tangible object. And when geometry is reduced to perceptual apprehension there is no real progression for the students.

Consequently, the mobilization and functioning of students' discursive apprehension and proving abilities can be inhibited or influenced by the intervention of mainly perceptual apprehension. This is another indication which enhances the fact that in the cases the students think that the solution was found through perceptual recognition of relations between the different parts of the figure, no further need for reasoning using their geometrical knowledge is needed. This students' conception was identified also in the study of Sharon, McCrone and Martin (2004), adding that this conception may have implications for the students' motivation to construct proofs. If examples constitute a proof or visually obvious relationships may be used as facts in a proof, why is it necessary to write formal, deductive arguments? In addition, the students' difficulties in moving away from perceived features of a figure can mislead the students as to the mathematical properties and objects represented by a drawing and can obstruct appreciation of the need for the discovery of proofs (Duval, 1995).

Based on the results of the first implicative diagram, the operative apprehension is mainly related to the solution of the tasks in which the figure has the nature of an image, whereas the perceptual apprehension of figures is mostly related to the solution of the tasks in which the figure has the nature of an object. Furthermore, the students' constant behavior in the solution of each type of task according to the nature of the figure it includes was revealed. We could thus claim that the students' cognitive process, either guided by perception or by a heuristic visualization of the figures determines the coherence of students' behavior when dealing with geometrical figures of different nature. Consequently, the perceptual and the operative apprehension of figures appear to be regulatory factors for the students' way of looking at figures, either as objects or as images respectively.

The operative thinking enhances the object oriented geometrical thinking. Students that have developed the operative apprehension of geometrical figures do not remain restrained in the nature of the figure as an image, but are able to overcome this perceptual obstacle of

a figure (either given or constructed) that is inconsistent to the utterance of the problem. In fact, the relations in the second implicative diagram show that the use of the operative apprehension of geometrical figures (reconfiguration abilities) together with the drawing of the correct figure in the tasks which the given figure has the nature of an image contribute to the solution of tasks and the process of justifying an answer in the case the given figure is inconsistent to the given properties.

6 Teaching implications

The results of this study reveal the obstacles created to students because of the nature of geometrical figures and thus have direct implications for teaching. In fact, this study highlights that solving geometrical tasks requires interactions between the perception and the heuristic visualization of figures. Thus, the basic problem for the teaching of geometry at lower and upper secondary schools is how to get the pupils to manage the coordination between these different types of figural apprehension. Therefore, reflection on a new approach for introducing geometry in primary and secondary education is necessary; the principle of which would be the awareness of the different ways of looking at figures. Thereafter, the focus of teaching geometry should be firstly turned on the differentiation among the visualization processes and the reasoning processes, because the coordination of these kinds of processes can really occur only after this focus on their differentiation.

Another crucial point to take into account as regards the teaching of geometry is promoting the involvement of the operative apprehension, through the use of the mereologic modification of figures for solving geometrical problems. Systematic work from the early years, even from pre-school education, in tasks of composition and decomposition of geometrical figures in a mereologic way is necessary, as it can lead to the heuristic visualization of figures by students. In fact, new research efforts have recently began in Cyprus towards this direction (Evangelou, Elia & Gagatsis, 2015; Petridou, Elia & Gagatsis, 2015). In this way the students' geometrical thinking can be developed from the first years of education and students may become more flexible in the way they look at geometrical figures.

References

[1] Bodin, A., Couturier, R., & Gras, R. (2000). *CHIC : Classification Hiérarchique Implicative et Cohésive-Version sous Windows – CHIC 1.2.* Rennes : Association pour la Recherche en Didactique des Mathématiques.

[2] Duval, R. (2014). The first crucial point in geometry learning: visualization. *Mediterranean Journal for Research in Mathematics education, 13,* 1-28.

[3] Duval, R. (2011). Why figures cannot help students to see and understand in geometry? Analysis of the role and the cognitive functioning of visualization.

Abstract Booklet of the Symposium Mathematics Education Research at the University of Cyprus and Tel Aviv University (pp. 22-23). Nicosia: Cyprus.

[4] Duval R. (2005). Les conditions cognitives de l'apprentissage de la géométrie : développement de la visualisation, différenciation des raisonnements et coordination de leurs fonctionnements. *Annales de Didactique et de Sciences Cognitives, 10,* 5-53.

[5] Duval, R. (2000). Proof Understanding in Mathematics: What ways for students? In *Proceedings of 2002 international conference on mathematics: Understanding proving and proving to understand* (pp. 61-77).

[6] Duval, R. (1995). Geometrical Pictures: Kinds of representation and specific processes. In R. Sutherland & J. Mason (eds.), *Exploiting mental imagery with computers in mathematical education* (pp. 142- 157). Berlin, Springer.

[7] Evangelou, K., Elia, I. & Gagatsis, A. (2015). The contribution of gestures in the acquisition of geometric concepts in early childhood. *EAPRIL Conference Proceedings 2014*, Issue 1 (pp. 209-222). ISSN 2406-4653.

[8] Gagatsis, A. (2011). Compartmentalization in Learning. In: N.M. Seel (Ed.), *Encyclopedia of the Sciences of Learning,* pp.665-668. London: Springer.

[9] Gagatsis, A., Michael, P., Deliyianni, E., Monoyiou, A., & Kuzniak, A. (2011). Secondary students' behavior in proof tasks: understanding and the influence of the geometrical figure. *Proceedings of the 7th Conference of the European Society for Research in Mathematics Education: Working Group 4* (p.p.608-617). Rzeszów, Poland.

[10] Gras R., Régnier J.C., Marinica, C., & Guillet F. (Eds) (2013). *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.* Toulouse: Cépaduès Editions.

[11] Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics, 44* (1), 5-23.

[12] Michael – Chrysanthou, P., Gagatsis, A. (2015). Ambiguity in the way of looking at a geometrical figure. *Revista Latinoamericana de Investigación en Matemática Educativa – Relime, 17* (4-I), 191-210.

[13] Michael – Chrysanthou, P., Gagatsis, A. (2013). Geometrical figures in task solving: an obstacle or a heuristic tool? *Acta Didactica Universitatis Comenianae – Mathematics, 13,* 17-32.

[14] Mesquita, A. L. (1998). On conceptual obstacles linked with external representation in geometry. *Journal of Mathematical Behavior, 17* (2), 183-195.

[15] Petridou A., Elia I., & Gagatsis A. (2015). Preschool geometrical teaching practices and geometrical thinking development: a case study. *EAPRIL Conference Proceedings 2014,* Issue 1 (pp. 223-238). ISSN 2406-4653.

[16] Sharon M., McCrone, S. & Martin. T.M. (2004). Assessing high school students' understanding of geometric proof. *Canadian Journal of Science, Mathematics and Technology Education, 4* (2), 223-242.

[17] Xistouri, X., Nicolaou, A., Koukkoufis, A., & Gagatsis, A. (2005). The nature of external representation in geometry and the solution of mathematical problems. In A. Gagatsis, F. Spagnolo, Gr. Makrides, & V. Farmaki (Eds.), *Proceedings of the 4th Mediterranean Conference on Mathematics Education:* (Vol. I,pp.163-174). Palermo: University of Palermo, Cyprus Mathematical Society.

Notes

1. This paper is in the context of the research project "Ability to use multiple representations in Functions and Geometry: The Transition from Middle to High school" of the Research Promotion Foundation of Cyprus [HUMANITIES/0308(BE)/03] – Coordinator: Athanasios Gagatsis/ Researcher: Paraskevi Michael-Chrysanthou.
2. The tasks analyzed in this paper have been used by Paraskevi Michael-Chrysanthou in her research in the context of her PhD dissertation [Michael, P. (2013). *Geometrical Figure Apprehension: Cognitive Processes and Structure.* Unpublished thesis.]

L'APPROCHE SOCIO-CONSTRUCTIVISTE DANS LES SITUATIONS D'ENSEIGNEMENT-APPRENTISSAGE DE LA BIOLOGIE EN TUNISIE
LE CAS DE LA REPRODUCTION HUMAINE

Fadhila FARHANE HORRIGUE[1]

THE SOCIO - CONSTRUCTIVIST APPROACH IN SITUATIONS OF TEACHING - LEARNING IN BIOLOGY IN TUNISIA. THE CASE OF HUMAIN REPRODUCTION

RÉSUMÉ

Dans l'acte enseignement-apprentissage centré sur le développement des compétences de l'apprenant quant aux savoirs, savoir-faire et savoir-être, l'implication de celui-ci parait primordiale. L'apprenant doit apprécier la valeur et l'intérêt de ce qu'il apprend, prendre conscience de ses difficultés et de ses atouts, réussir à construire des savoirs en collaboration avec l'enseignant et notamment avec ses pairs. En effet, la didactique des sciences part du postulat constructiviste selon lequel l'apprenant est l'acteur principal de ses apprentissages. En tant que pôle central de la situation éducative, c'est à lui qu'incombe la construction de ses connaissances en coopération avec ses pairs. Les démarches de co-construction sont liées à une activité raisonnée du sujet apprenant, activité stimulée par une confrontation avec ses pairs créant ainsi un « conflit sociocognitif » qui est un élément fondamental de la théorie de l'apprentissage et qui conduit au développement cognitif. Une telle confrontation entre pairs pourrait bien ébranler des préconceptions et rompre avec un « déjà-là conceptuel ». Pourrait-on moyenner une telle approche didactico-pédagogique dans l'apprentissage de la reproduction humaine dans une société arabo-musulmane où parler de sexualité pourrait-être une attaque à la pudeur? Faut-il enseigner une science abstraite dissociée de tout contexte culturel?

Mots-clés : *apprenant, co-construction, apprentissage, conflit sociocognitif, reproduction humaine.*

ABSTRACT

In the teaching-learning act focused on the development of learner skills referring to the knowledge, skills and attitudes, his involvement seems paramount. The learner must appreciate the value and interest of what he learns, become aware of his difficulties and strengths, manage to construct knowledge in collaboration with the teacher and especially with peers. Indeed, didactics science is based on the constructivist assumption that the learner is the main actor of his learning. As the central hub of the educational situation, he must build his knowledge in cooperation with peers. The co-construction processes are related to a rational activity of the learner, activity stimulated by the confrontation with peers creating then a "socio-cognitive conflict" which is fundamental to the theory of learning and leads to cognitive development. Such confrontation between peers might undermine preconceptions and break with a "conceptual already there." Could we average such didactic-pedagogic approach in learning of human reproduction in an arab-muslim society in where talking about sexuality could be considered as an attack against decency? Should we teach an abstract science independent of any cultural context?

Keywords: *learner, co-construction, learning, socio-cognitive conflict, human reproduction.*

[1] Université de Lyon, ED 485 EPIC, UMR 5191 ICAR, fadhilamsh@yahoo.fr

1 Introduction

Aborder la question de l'enseignement scientifique d'une façon problématique, s'interroger et débattre sur l'apprentissage des sciences, se soucier des moyens qui permettent de rendre la culture scientifique plus adaptée à l'école, plus appréhendable et plus assimilable par la population scolaire, sont des préoccupations qui, depuis quelques décennies, ont fait l'objet de plusieurs recherches notamment en Sciences de l'éducation.

Si l'on reconnait l'intérêt de l'enseignement des sciences en général et de l'enseignement de la reproduction humaine en particulier, il n'en est pas de même pour une autre question importante qui s'est longtemps positionnée au cœur du débat sur l'école dans plusieurs pays et notamment les pays de cultures non occidentales, en l'occurrence la Tunisie, et qui continue à animer les débats : Quel enseignement scientifique pour quelle société et quelle pédagogie pour quel enseignement ? Tant de questionnements prennent leurs racines dans les méthodes et procédés utilisés dans la réalisation des apprentissages scolaires en rapport avec l'appartenance des apprenants, tout comme les enseignants, à une société arabo-musulmane.

L'intérêt d'une approche co-constructiviste dans les apprentissages scolaires est né au moment où on enseigne à des apprenants des concepts se rapportant à un thème dont nul n'a le droit d'aborder hors cadre scolaire. On le vit, on en parle mais jamais en public. De tels faits génèrent chez les apprenants une réticence manifeste qui s'exprime par un manque de participation lors de la construction des savoirs en faveur d'un paradigme éducatif simplement transmissif. Ainsi une problématique s'installe au moins à deux niveaux, lesquels sont en interaction certaine : d'un côté, le niveau didactico-pédagogique et de l'autre, le niveau socio-culturel.

Faut-il enseigner une science abstraite dissociée de tout contexte culturel ou, au contraire, élaborer des curricula scientifiques qui prennent en considération le système de valeurs de la société de référence des apprenants ? Faut-il enseigner ce que l'on appelle « éducation à la sexualité » telle quelle, en tant que « produit culturel né dans les pays de culture occidentale », ou encore faut-il l'adapter aux spécificités culturelles du pays concerné ? Les savoirs, les croyances religieuses et les systèmes de valeurs traditionnels ne sont-ils pas des obstacles à l'accès aux sciences « pratiques » concernant la vie et la santé dont nul n'est censé ignorer ?

2 La reproduction humaine, discipline d'apprentissage et discipline d'enseignement : L'éducation sexuelle dans les réformes de l'éducation en Tunisie

L'enseignement scientifique n'était pas assuré dans les écoles avant l'indépendance (1956). Mais c'était grâce à la réforme de 1958 qu'il fut introduit dans le programme d'enseignement. La réforme de 1958 s'est engagée à apprêter l'enfant tunisien à son rôle de « citoyen, d'homme et de former des cadres nécessaires aux développements de l'activité nationale » (JORT, 1958). Ce rôle de citoyen attribué à l'élève a pour but de favoriser le développement et l'épanouissement de la culture nationale. Ainsi, les finalités de cette réforme mettaient l'accent plutôt sur le développement de la Tunisie

que sur le profil du citoyen. Pour ce qui est en relation avec la personnalité de l'élève, les composantes en relation avec la sexualité sont absentes des finalités fixées par cette réforme.

L'introduction et la généralisation de l'enseignement de l'éveil scientifique[2] dans les écoles primaires se fit à partir de 1970[3]. Cependant, au cycle primaire, on ne faisait nulle référence à la reproduction humaine. Au collège, l'enseignement de la reproduction humaine se limite à quelques notions simples et superficielles sans pour autant insister sur des compétences à acquérir. Ainsi, au cœur du changement social que connaissait la Tunisie pendant cette période, l'abord de la reproduction humaine dans les programmes de sciences naturelles ne commençait qu'en troisième année secondaire où l'âge des élèves varie entre 15 et 17 ans ; âge très proche de l'âge moyen du mariage, en Tunisie à l'époque. Par conséquent, les jeunes abordaient une vie sexuelle sans avoir encore eu l'occasion de connaitre leurs corps, les changements physiologiques qui s'y produisaient, affrontant ainsi des problèmes qu'ils ne sauraient gérer vu le manque de formations et d'informations adéquates.

En 1986, un nouveau programme des Sciences de la Vie et de la Terre (SVT) a été mis en place avec introduction de l'étude des caractères sexuels secondaires à partir de la septième année de l'école de base (première année de collège où l'âge des élèves varie de 11 à 13 ans). Mais cette partie a été très vite retirée du programme. La reproduction humaine réapparait dans le cursus de l'élève à partir de la neuvième année de l'enseignement de base (dernière année de collège). A ce niveau de scolarité, l'âge moyen des apprenants varie entre 15 et 17ans. De ce fait, les adolescents font recours à des sources d'informations diverses et non toutes fiables, étant donné que la découverte de la sexualité se fait actuellement à un âge de plus en plus précoce chez les jeunes.

Dans ces programmes, on souligne un enrichissement relatif par rapport à ceux de 1970 avec un intérêt accordé à l'hygiène, à la culture sanitaire, au bien-être de l'individu et à la prévention de certaines maladies. Par ailleurs, le programme aborde l'hygiène sociale à travers la prévention des maladies sexuellement transmissibles (MST) mais, où seule la blennorragie (signes, agent, traitement et prévention) est traitée. En outre et en réponse aux actions de l'ONFP[4] sur le plan familial et social, a été introduite la notion de cycle sexuel chez la femme avec le mécanisme des menstruations, l'activité des ovaires, la grossesse, la régulation des naissances et la stérilité féminine. Limiter ainsi les informations à propos de cette thématique, revient à priver l'élève des connaissances nécessaires relatives à son éducation reproductive et sexuelle.

Notons que dans la conception du programme de SVT, la partie relative à la contraception se limite aux moyens contraceptifs féminins, comme les pilules contraceptives, le diaphragme le stérilet et la ligature des trompes ce qui relève de l'attribution sociale, à la femme, de la charge de la contraception.

---

[2] Discipline enseignée en primaire dans le but d'une initiation aux sciences expérimentales
[3] Selon le rapport sur le mouvement éducatif tunisien, 1988
[4] Office National de la Famille et de la Population

3 Intérêt didactique de l'implication de l'apprenant dans ses apprentissages

Dans l'acte enseignement-apprentissage centré sur le développement des compétences de l'apprenant sur les plans savoir, savoir-faire et savoir-être, l'implication de celui-ci s'avère primordiale. Ce dernier doit apprécier la valeur et l'intérêt de ce qu'il apprend, prendre conscience de ses difficultés et de ses atouts, savoir se relire, s'autocorriger, s'autoévaluer le mieux possible. L'objectif final est de l'amener à réfléchir sur ses démarches intellectuelles afin de développer ses capacités. Comment l'enseignant pourrait-il faciliter l'instauration de telles conduites ?

Par un discours psychopédagogique promouvant un climat de confiance et de partage, l'enseignant suscite une attitude participative et un esprit coopératif chez l'apprenant. L'enseignant agira en classe comme formateur et guide (Hoffman-Gosset, 2006), comme accompagnateur (Chesnais, 1998), comme animateur (Liquète, Maury, 2007) comme facilitateur (Rogers, 1996) et non comme détenteur de savoirs. Il accompagne l'apprenant dans son processus d'apprentissage et l'aide à progresser. Une telle attitude de l'enseignant pourrait bien conduire l'accompagné vers une autonomie dans la mesure où il l'encourage, l'épaule, l'écoute, le soutient et le guide afin qu'il s'observe davantage, prenne du recul par rapport à ce qu'il vit, analyse mieux les causes de ses difficultés et trouve un moyen pour les surmonter.

Les préoccupations de la didactique des sciences en matière d'enseignement-apprentissage peuvent, dès lors, être ramenées à deux questions radicales intimement liées qu'on peut formuler de la manière suivante :

- Comment des apprenants s'approprient-ils le savoir ?

- Quels sont les moyens et conditions didactiques à rassembler afin d'assurer et de garantir une meilleure appropriation de ce savoir ?

Toutefois, appréhendées dans une perspective didactique, les deux interrogations s'imbriquent tellement qu'il n'est point judicieux de chercher à les distinguer.

La première question se rapporte au processus d'apprentissage, la deuxième renvoie au processus d'enseignement. Ces deux processus sont si solidaires et si synergiques, en contexte scolaire, qu'il est presque impossible de parler de l'un sans discuter de l'autre.

La didactique des sciences part du postulat constructiviste selon lequel l'élève est l'acteur principal de ses apprentissages. En tant que pôle central de la situation éducative, c'est à lui, et à lui seul qu'incombe la construction des connaissances. Il convient donc d'aller plus loin dans la réflexion et essayer de comprendre comment s'effectue réellement cette auto/co-construction des savoirs chez l'apprenant. C'est là tout l'enjeu de l'acte d'apprendre.

S'il n'y a pas de problème, il n'y a pas de difficultés et donc pas d'apprentissages. Tant que l'on n'a aucun problème avec une situation donnée, on se satisfait de ce que l'on est, de ce que l'on a et de ce que l'on fait. Ainsi, un apprenant maintien sa représentation quant à un objet tant et aussi longtemps qu'il n'a pas de problèmes avec celle-ci. En revanche, face à une situation problème, tantôt aimanté par ses représentations, tantôt refoulé par les résultats, l'apprenant se trouve dans l'obligation de défier le problème, le sien, afin d'apporter une solution et ce, généralement en

collaboration avec ses pairs au sein d'un conflit sociocognitif. De ce dernier émergent deux résultats, à savoir une solution au problème d'une part, et une reconstruction cognitive d'autre part.

4 Contexte socioculturel tunisien et implication de l'apprenant dans le champ de l'enseignement apprentissage de la reproduction humaine

Penser l'enseignement et l'apprentissage dans une perspective anthropologique et culturelle parait une voie inéluctable si l'on veut envisager une vraie pédagogie de l'apprenant, c'est-à-dire une pédagogie de la différence qui tient compte, non seulement des spécificités psychologiques et historiques de l'élève tunisien, mais aussi des spécificités culturelles de sa société d'appartenance[5] , comme l'exprime fort bien la formule de Bernard Charlot : « Le sujet est indissociablement humain, social et singulier. Le sujet est engagé dans une histoire, dans laquelle il est à la fois porteur de désir et confronté à du "déjà-là" (du patrimoine humain dont il doit s'approprier une part). Le sujet interprète le monde, fait sens du monde, des autres et de lui-même». (Charlot, 2003, p.14)

Mais aussi incroyable que cela puisse paraître, la plupart des oppositions à l'éducation à la sexualité sont basées sur la supposition que l'information est nuisible. Certains voient que parler de sexe aux enfants, c'est les inciter à essayer ; leur parler de MST revient à les encourager pour sortir et en attraper. Traiter de l'éducation à la sexualité est synonyme de s'attaquer à la pudeur et inculquer l'immoralité alors que les recherches dans ce domaine révèlent, qu'au contraire, c'est l'ignorance et la curiosité non résolue qui sont nuisibles et non pas l'information. La sexualité est bien plus qu'un comportement, elle est un mode de pensée qui ne peut se réduire à une génitalité.

Dans le cadre de l'enseignement de la reproduction humaine aux apprenants adolescents au sein des lycées secondaires tunisiens, l'approche anthropologique et culturelle est particulièrement intéressante pour comprendre certains aspects liés à l'enseignement de la reproduction humaine. Dans ce contexte particulier, Ahmed Chabchoub énonce : « Dans une société comme la nôtre, à peine sortie du modèle traditionnel, une ouverture anthropologique nous permettrait de contextualiser les connaissances scientifiques véhiculées par l'école "longtemps tenues pour universelles" et de poser des questions essentielles du genre : que font les élèves tunisiens – qui appartiennent à une culture "prémoderne" – du savoir scientifique "moderne" que leur communique l'école ? » (Chabchoub, 2000, p.37)

Cette réflexion témoigne d'une prise de conscience de l'importance éducative des facteurs liés à la culture et à la civilisation de références des élèves. Une autre raison rend légitime le regard anthropologique et culturel quant à l'enseignement de la reproduction humaine, c'est le chevauchement conceptuel entre le culturel et le religieux.

---

[5] Soulignons toutefois avec Charlot que la différence culturelle n'est pas et ne peut être réduite à « une confrontation entre systèmes symboliques, projetée ensuite de façon mécanique ("intériorisée") dans les sujets, [mais doit être appréhendée] dans la forme qu'elle prend chez un sujet à travers ce qu'il en fait »

Un tel chevauchement s'oppose-t-il à une approche constructiviste de l'apprentissage de la reproduction humaine dans une société arabo-musulmane, en l'occurrence tunisienne ? Un changement de paradigme est-il envisageable ?

5 Problématique : Passer du paradigme de l'enseignement au paradigme de l'apprentissage ?

Le paradigme d'apprentissage s'appuie sur une conception constructiviste qui cherche à comprendre comment la connaissance se construit chez un sujet. Il s'éloigne par conséquent du paradigme de l'enseignement qui prend plutôt sa source dans les théories cognitivistes et béhavioristes. De telles théories favorisent l'adoption de pratiques pédagogiques centrées sur l'identification de comportements ou stratégies d'enseignement qui favorisent l'apprentissage. En revanche, le constructivisme est essentiellement centré sur l'apprenant et sur la manière dont il construit son savoir éventuellement avec l'aide de ses pairs. Il repose sur le postulat voulant qu'il n'y ait de connaissance que construite par l'apprenant lui-même, c'est-à-dire par son activité cognitive.

En dehors de cette activité, menée par l'apprenant lui-même, ne subsisterait que la trace de la connaissance construite par autrui qui ne saurait devenir sienne que s'il fait l'effort de la reconstruire et la reconstituer. Dans ce sens, la connaissance n'est pas transmissible, elle est constructible.

Dans le contexte tunisien et dans le champ de la reproduction humaine, un virage en faveur du paradigme de l'apprentissage est-il réalisable là où nature sonne comme une vérité populaire qui justifie souvent les contre-attitudes éducationnelles postulant que la sexualité... « Ça ne s'apprend pas, c'est naturel ! » ?

Cette logique de rupture, du passage d'un paradigme à l'autre, à savoir de celui de l'enseignement à celui de l'apprentissage, nécessite une profonde remise en question des pratiques des enseignants et surtout des conceptions des apprenants. C'est pourquoi il convient de se demander si cette injonction de passer du paradigme de l'enseignement à celui de l'apprentissage est bien fondée et que si les apprenants, tunisiens en l'occurrence, sont disposés à s'y mettre. Plusieurs questions se posent à savoir : Quelles attitudes développent les apprenants tunisiens par rapport aux connaissances scientifiques susceptibles de bousculer leurs convictions intimes et perturber leur identité » ? Leurs pratiques religieuses les empêcheront-elles de travailler en groupes mixtes afin de co-construire un savoir quant à la sexualité ? Parlent-ils de sexualité en famille pour en parler librement en classe ? Questions auxquelles nous tenterons d'apporter des éléments d'explication et de réponse grâce à une analyse de données fondée sur l'approche Analyse Statistique Implicative (ASI) développée par Régis Gras (1979) et ses collaborateurs (Gras et al. 1996, 2009, 2013) instrumentée par le logiciel

C.H.I.C[6] sur lequel nous avons travaillé sous les directives éclairées de notre directeur de thèse.

L'analyse statistique implicative est définie comme étant « [...] Un champ théorique centré sur le concept d'implication statistique ou plus précisément sur le concept de quasi-implication [...] L'étude de ce concept de quasi-implication en tant qu'objet mathématique, dans le champ des probabilités et de la statistique, a permis de construire des outils théoriques qui instrumentent une méthode d'analyse de données ». (Gras, Régnier, 2009, p.12)

6 Méthodologie

Notre étude se rapporte à l'apprentissage de la reproduction humaine en terminale sciences expérimentales moyennant une pratique pédagogique qui n'est ni véritablement traditionnelle ni pour autant moderne, une pédagogie baptisée « hybride » (Farhane, 2012). Il est clair que nous nous intéressons à une discipline particulière qui est « les sciences de la vie » dans laquelle est inscrite la physiologie.

6.1 Outil de construction de données

Les données mobilisées dans cet article ont été collectées en décembre 2014, par questionnaires anonymes afin que les sujets puissent répondre plus librement et avec le plus de sincérité possible.

L'expérimentation a porté sur 71 élèves appartenant à trois classes de terminales sciences expérimentales, soient 50 filles et 21 garçons dont les âges varient entre 19 et 21 ans.

6.2 Hypothèses

Hypothèse générale :

La religion ne réfute guère l'éducation à la sexualité et qu'une telle conception ne pourrait résulter que d'une « mauvaise » compréhension de la religion islamique, voire un fondamentalisme qui n'admet qu'une interprétation littérale du sacré Coran en s'opposant à toute lecture historique et scientifique. Ainsi, refuser de parler de sexualité n'est que de l'ordre de la construction culturelle.

Hypothèses de recherche :

Hypothèse 1 : Si l'on est pratiquant, on présume que parler de sexualité est un tabou familial;

Hypothèse 2 : C'est plutôt les filles qui se trouvent embarrassées lors des cours sur la fertilité chez la femme;

[6] Classification Hiérarchique Implicative et Cohésitive

Hypothèse 3 : Que l'on soit fille ou garçon, on est embarrassé pendant les cours se rapportant sur la fertilité chez l'homme;

6.3 Variables et résultats

Nous avons travaillé avec 15 variables (V01 à V15) se rapportant à 15 items. La réponse à chacun des items sera par « oui » ou « non ». Les réponses par « oui » seront codées « 1 » et celles par « non » le seront par « 0 ».

Les variables sont les suivantes :

V01 Sexe : Homme / Femme

V02 Vous êtes musulman pratiquant

V03 Votre mère est pratiquante

V04 Votre père est pratiquant

V05 Vous parlez de sexualité en famille

V06 Vous parlez de sexualité hors cadre familial

V07 Vous préférez que le chapitre « la fertilité chez la femme » soit enseigné par une femme

V08 Vous préférez que le chapitre « la fertilité chez l'homme » soit enseigné par un homme

V09 Vous êtes embarrassé(e) pendant les cours se rapportant à « la fertilité chez la femme »

V10 Vous êtes embarrassé(e) pendant les cours se rapportant à « la fertilité chez l'homme »

V11 Vous préférez être éduqué(e) à la sexualité à l'école avant l'âge de la puberté

V12 Vous préférez travailler entre pairs, en petits groupes mixtes, sur « la fertilité chez la femme »

V13 Vous préférez travailler entre pairs, en petits groupes mixtes, sur « la fertilité chez l'homme »

V14 L'éducation à la sexualité est un tabou familial

V15 L'éducation à la sexualité est un tabou sociétal

Les résultats du questionnaire figurent dans les tableaux 1 et 2 situés en annexe.

7 Interprétations des résultats et discussions

Nous exposerons les résultats obtenus grâce à la méthode Analyse Statistique Implicative (ASI) afin de voir si des liens existent entre les divers items préalablement explicités (§ 6.3).

Les variables Vi (i=1 à 15) seront traitées comme variables binaires. Grâce à la méthode ASI, méthode fine et riche, nous étudierons l'arbre de similarité, le graphe implicatif et l'arbre cohésitif.

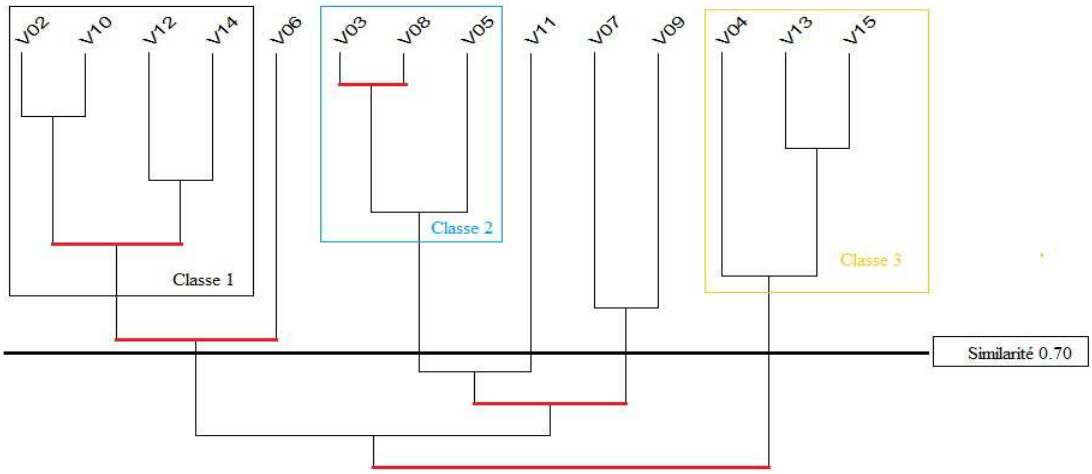


Figure 1 – Arbre des similarités

Comme le montre l'arbre des similarités, et en se limitant à un niveau de confiance en la similarité de 0.70, nous distinguons sept niveaux de classification qui sont les suivants:

Classification au niveau : 1 : (V03 V08) similarité : 0.983175

Classification au niveau : 2 : (V02 V10) similarité : 0.957933

Classification au niveau : 3 : (V13 V15) similarité : 0.872316

Classification au niveau : 4 : (V12 V14) similarité : 0.866096

Classification au niveau : 5 : ((V03 V08) V05) similarité : 0.752446

Classification au niveau : 6 : ((V02 V10) (V12 V14)) similarité : 0.744437

Classification au niveau : 7 : (V04 (V13 V15)) similarité : 0.725211

Les trois classes que l'on distingue sur l'arbre des similarités, toujours en se limitant à un niveau de similarité de 0.70 ainsi que leurs caractéristiques sont les suivantes :

Contribution à la classe 1 : V02, V10, V12, V14 (2, 4, 6)
Le groupe optimal (GO) est: Ind68, Ind19, Ind60, Ind53, Ind65, Ind66

Card GO = 6 p = 0.0845 1-p = 0.915
La variable Homme est typique à cette classe, elle contribue le plus avec un risque de : 0.00649
La variable Femme contribue le plus à cette classe avec un risque de : 0.932

De par ces données, nous voyons clairement que les apprenants de sexe masculin et qui sont pratiquants sont les plus embarrassés pendant les cours se rapportant à « la fertilité chez l'homme » ce qui ne plaide aucunement en faveur de notre troisième hypothèse. Mais, ces mêmes apprenants préfèrent travailler avec leurs pairs de sexe

féminin sur « la fertilité chez la femme ». Nous voyons nettement une envie de découvrir la sexualité de « l'autre » avec « l'autre » du moment que pour eux, l'éducation à la sexualité est bel et bien un tabou au sein de la famille et qu'il fallait bien en parler hors cadre familial.

Contribution à la classe 2 : V03, V08, V05 (1,5)
Le groupe optimal (GO) est: Ind04, Ind27, Ind64, Ind66, Ind60, Ind50, Ind18, Ind44, Ind55, Ind24, Ind52, Ind32, Ind31, Ind69, Ind15, Ind05

Card GO = 16 p = 0.225 1-p = 0.775
La variable Homme est typique à cette classe et y contribue avec un risque de : 0.176
La variable Femme est typique à cette classe avec un risque de : 0.591

Depuis ces données nous voyons que les apprenants de sexe féminin ayant des mères pratiquantes parlent plus de sexualité en famille plus que ceux de sexe masculin. Non embarrassées par la sexualité masculine, elles préfèrent que le cours sur « la fertilité chez l'homme » soit enseigné par un homme. D'autant plus, le fait que la mère soit musulmane pratiquante n'exclue pas le fait de parler de sexualité en famille.

Contribution à la classe 3 : V04, V13, V15 (3,7)
Le groupe optimal (GO) est: Ind67, Ind62, Ind08, Ind11, Ind16, Ind21, Ind01, Ind02, Ind03, Ind07, Ind22, Ind40, Ind43, Ind46, Ind51, Ind28, Ind30, Ind36, Ind39

Card GO = 19 p 0.268 1-p = 0.732
La variable Homme est typique à cette classe avec un risque de : 0.853
La variable Femme est typique à cette classe, elle y contribue avec un risque de : 0.159

Ces résultats viennent consolider les précédents. En effet, s'agissant de filles avec des pères musulmans pratiquants, elles adoptent l'opportunité de travailler sur « la fertilité chez l'homme » en collaboration avec leurs pairs de sexes opposés, sachant qu'elles voient que « l'éducation à la sexualité » n'est pas un tabou familial mais plutôt sociétal ce qui nous permet d'infirmer notre troisième hypothèse. Ainsi, peut-on associer les variables selon des critères de vraisemblances de la classe 2 avec celles de la classe 3 pour dire que les filles issues de parents pratiquants ont plus tendance à parler de sexualité en famille et choisissent de travailler avec leurs pairs de sexe masculin sur les thèmes se rapportant à la sexualité. Toutefois, même si des barrières s'opposent à cette idée d'ouverture d'esprit, elles ne sont que de l'ordre du sociétal et non du familial.

Voyons ce que nous rapporte le graphe implicatif comme résultats, sachant qu'il nous renseigne quant aux tendances entre les différentes variables.

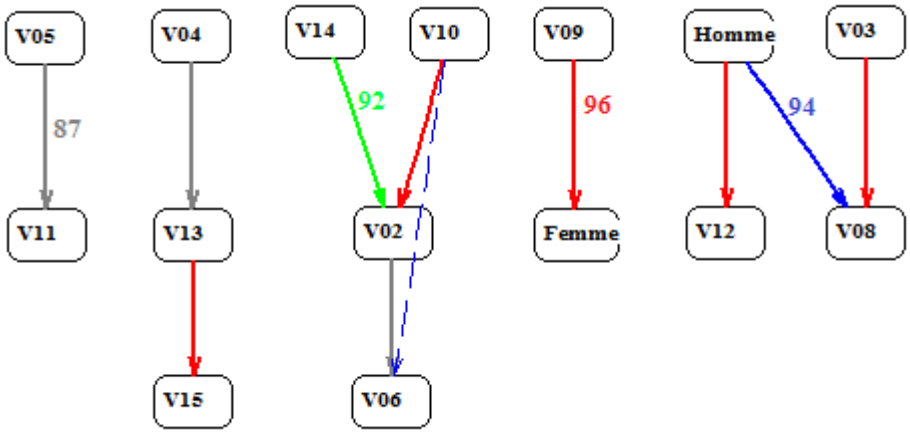


Figure 2 – Graphe implicatif

Signalons qu'il y a différents niveaux de confiance, allant de 0.5 à 1.

À un niveau de confiance de 0.96 (flèches rouges), une quasi-implication s'avère nette ; si l'on est embarrassé pendant les cours sur « la fertilité femelle » c'est que l'on est de sexe féminin. En revanche, c'est plutôt les hommes qui préfèrent travailler en petits groupes mixtes sur « la fertilité femelle », ce qui pourrait expliquer l'embarras des filles. Un tel constat nous amène à confirmer notre seconde hypothèse. Au même niveau de confiance, ceux qui choisissent travailler en groupes mixtes à propos de « la fertilité chez l'homme » voient que l'éducation à la sexualité n'est qu'un tabou sociétal. Également, les apprenants ayant des mères pratiquantes supputent le fait que « la fertilité chez l'homme » soit enseignée par un homme. En effet, l'indice de similarité entre les deux variables (V03 et V08) est de 0.98. Nous voyons bien une appétence à découvrir « l'autre à travers l'autre » même au sein de familles musulmanes pratiquantes.

En descendant à un niveau de confiance plus bas, voire 0.94, nous voyons que quand l'apprenant est de sexe masculin, il a tendance à préférer être renseigné sur « la fertilité chez l'homme » par un homme d'où une certaine réserve voir même une pudeur. Dans ce sens, un homme conçoit mal qu'une femme lui parle de sa sexualité mais il apprécie bien qu'elle lui parle de la sienne.

À un niveau de confiance encore plus bas mais assez fort, à 0.92, ceux qui supposent que l'éducation à la sexualité est un tabou familial sont essentiellement des musulmans pratiquants, ce qui nous conduit à confirmer notre première hypothèse. Ces sujets figurent parmi ceux qui ont toujours cru que la religion musulmane n'est jamais en faveur d'une éducation à la sexualité et ne l'encourage sous aucun prétexte. En revanche, l'Islam met l'emphase sur l'acquisition de la connaissance et, durant la vie du prophète Mohamed - qui considère que la sexualité est un acte de foi et que le plaisir y est un droit absolu - les musulmanes et musulmans n'étaient pas gênés pour le questionner, notamment à propos d'affaires privées notamment de la vie sexuelle, il y répondait de bon gré et en discutait ouvertement.

Les apprenants qui abordent la sexualité en famille préfèrent être éduqués à la sexualité à l'école avant l'âge de la puberté ; quasi implication vérifiée avec un niveau de confiance de 0.87.

Enfin, tentons d'analyser les données de l'arbre cohésitif qui nous renseigne sur le degré de cohésion entre les différentes variables préalablement fixées. Le degré de cohésion le plus fort correspond à 1 ; ainsi plus faible est le degré de cohésion, moins de cohésion entre les éléments il y en a.

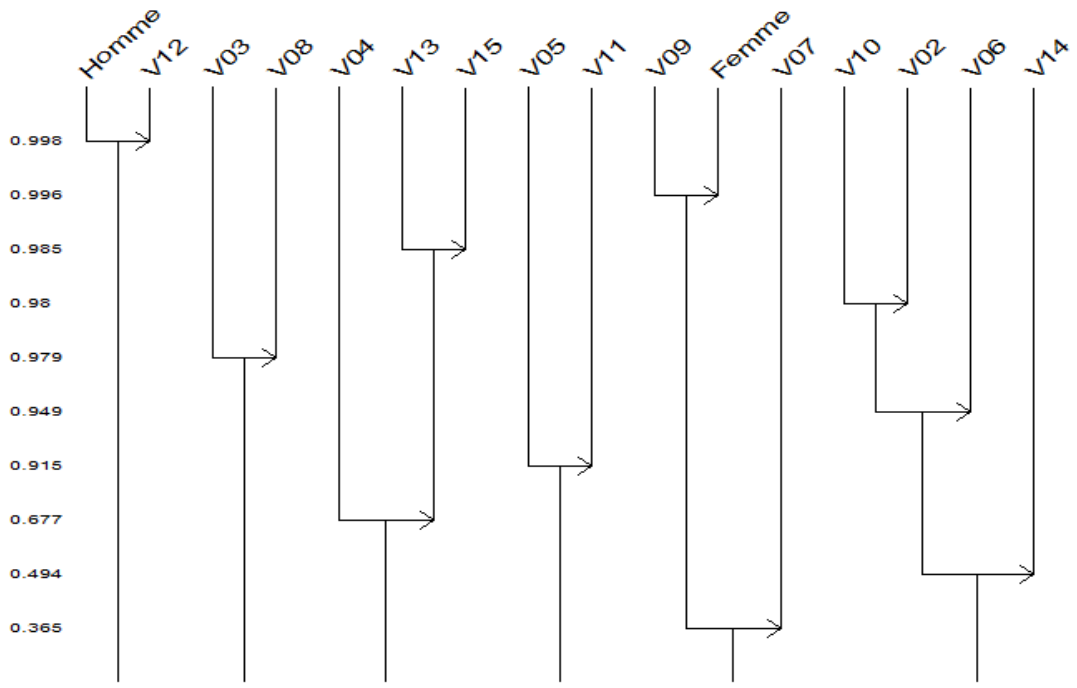


Figure 3 – Classification hiérarchique orientée (Arbre Cohésitif) – Variables « Homme » et « Femme » variables actives.

Nous voyons clairement la forte cohésion (0.998) entre la variable « Homme » et la variable V12 (Vous préférez travailler entre pairs, en petits groupes mixtes, sur la fertilité chez la femme). Cependant, les apprenants de sexe masculin, apprécient parler de la sexualité chez la femme avec leurs pairs des deux sexes, ce qui exclue chez eux, tout embarras d'aborder la sexualité de « l'autre » avec « l'autre ». En revanche, les filles s'avèrent nettement gênées pendant les cours se rapportant à la sexualité chez la femme, peut-être est-il par pudeur. En effet, nous décelons une forte cohésion (0.996) entre « Femme » et V09 (Vous êtes embarrassés pendant les cours sur la fertilité chez la femme) ce qui confirme encore davantage notre seconde hypothèse qui stipule que c'est plutôt les femmes qui se trouvent embarrassées les cours sur la fertilité femelle durant.

Un degré de cohésion de 0.98 est entre V10 (vous êtes embarrassés pendant les cours se rapportant à la fertilité chez l'homme) et V02 (vous êtes musulman pratiquant) est décelable ce qui vient infirmer encore une fois notre troisième hypothèse.

Exploitons les données apportées par l'arbre cohésitif où la variable « sexe » est considérée comme variable « supplémentaire », donc illustrative.

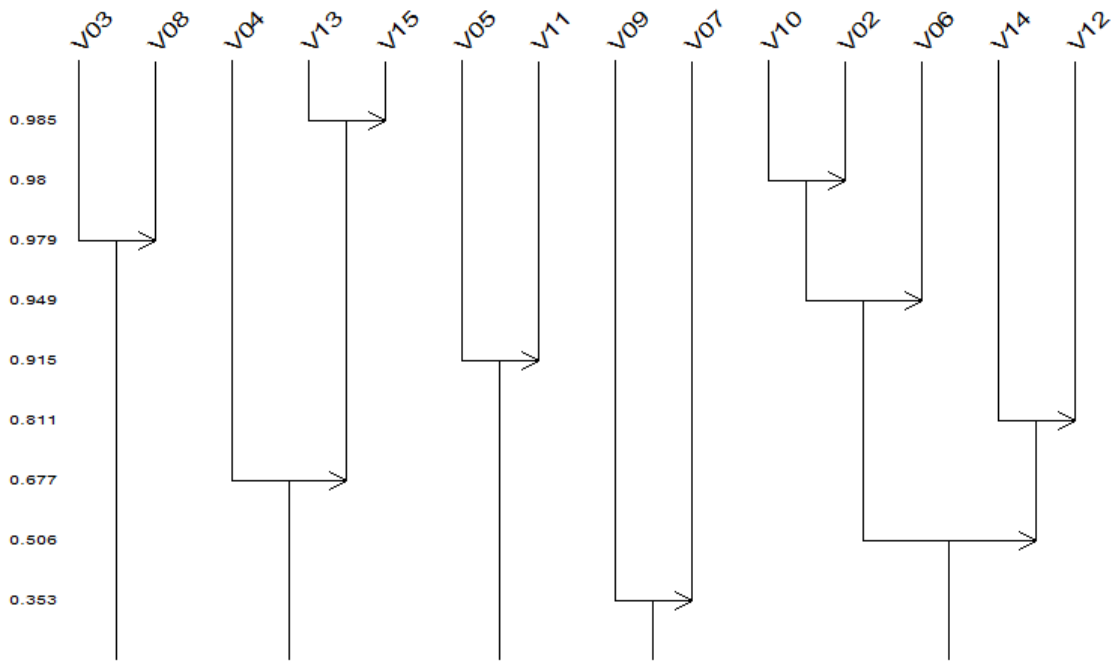


Figure 4 – Classification hiérarchique orientée (Arbre Cohésitif) – Variables « Homme » et « Femme » variables illustratives.

Nous voyons que la V13 (préférer travailler avec ses pairs en petits groupes mixtes sur la fertilité chez l'homme) est bien corrélée à V15 (l'éducation à la sexualité est un tabou sociétal) avec un degré de corrélation de 0.985. La variable typique à cette classe (V13, V15) est la variable « femme » avec un risque de 0.115 ce qui confirme que « de la sexualité... on parle, on en discute mais jamais hors cadre scolaire ». (1. §3)

Encore, quand on est musulman pratiquant (V02), on est bien embarrassé par les cours sur la « fertilité chez l'homme » (V10). Le degré de corrélation entre V02 et V10 est de 0.98. La variable typique à cette classe est la variable « homme » avec un risque de 0.0131. Ce résultat prouve que ce sont plutôt les garçons musulmans pratiquants qui se vérifient dérangés pendant les cours se rapportant à la sexualité chez l'homme ce qui invalide bien l'énoncé de la troisième hypothèse

À partir des résultats fournis par l'arbre des similarités (fig.1), le graphe implicatif (fig.2) et les arbres cohésitifs (fig. 3 et 4), récapitulons nos trois hypothèses.

Hypothèse première confirmée, quand on est musulman pratiquant, aborder la sexualité est un tabou familial.

Hypothèse seconde encore confirmée. En effet, c'est plutôt les filles qui se sentent embarrassées pendant les cours sur la fertilité femelle, contrairement aux garçons.

Hypothèse troisième infirmée, c'est essentiellement les garçons qui sont embarrassés durant les cours à propos de la fertilité chez l'homme.

8 Conclusion générale

Au travers de ces résultats, on voit distinctement que la religion ne réfute guère l'éducation à la sexualité et qu'une telle conception ne pourrait résulter que d'une

« mauvaise » compréhension de la religion musulmane, voire un fondamentalisme qui n'admet qu'une interprétation littérale du sacré Coran en s'opposant à toute lecture historique et scientifique. La sexualité est loin d'être une affaire « sale » et rien de religieux ne saurait justifier l'ignorance et la méconnaissance de certains musulmans sur ce sujet précis. Par conséquent, comme tout autre fait social, elle doit être étudiée et faire l'objet d'attention, de lectures et de réflexions. Mais pour les musulmans, en général, et les pratiquants en particulier, l'affaire semble être plus compliquée, ils portent en eux le lourd fardeau de la « honte », héritée, exportée et mal comprise de générations en générations bien que le Saint coran énonce que « La hayâa fiddin » autrement dit « nulle honte dans la religion ». Manifestement, la définition de la pudeur « religieuse » s'est perdue au fil du temps, au détriment des traditions et des coutumes locales. Le glissement des mots a fini par installer la peur, la gêne et le tabou au niveau de la société arabo-musulmane. Or, cette réflexion est contradictoire avec l'esprit de l'Islam souvent mis en accusation. En effet, celui-ci encourage vivement la sexualité dans le cadre du mariage, qui est de fait considéré comme une aumône mutuelle, une donation réciproque.

En effet, aujourd'hui au sein de la société tunisienne musulmane, « sexe » est un mot «sale» et «grossier». Des parents se sentent embarrassés de se voir discuter de l'éducation à la sexualité avec leurs enfants et laissent plutôt libre cours aux pairs, aux autres adultes et aux médias qui pourraient leur conférer de fausses informations qui pourraient bien compromettre leur « capital santé » par ignorance.

Des parents mettent en cause le fait même que les enfants aient besoin d'éducation à la sexualité, sous prétexte qu'on n'enseigne guère à un caneton à nager, suffirait de le mettre à l'eau et le laisser nager. Leur thèse est que : « après tout, pendant des milliers d'années, les hommes et les femmes ont eu des relations sexuelles sans éducation formelle et que l'éducation sexuelle commence après le mariage et par tâtonnement ». Il est impératif de rompre avec ces préconceptions à propos de la sexualité et faire comprendre à tous que parler de sexe n'est ni vulgaire ni une attaque à la pudeur. Sans doute faut-il hausser les voix pour changer de voie.

Nous devrions être sensibilisés quant à l'éducation à la sexualité dans la mesure où connaître son corps et son fonctionnement, éprouver sa psychologie est, nous semble-t-il, le meilleur moyen de « s'apprivoiser » et de canaliser la puissance de ses pulsions sexuelles. De plus, nous pensons fermement que le fait de combler ses carences - en termes d'information - permet de limiter les incompréhensions entre les deux sexes et donc, de désamorcer les conflits que la méconnaissance de l'autre pourrait engendrer. La principale raison pour laquelle les parents musulmans ne parlent pas suffisamment de sexualité avec leurs enfants est de l'ordre de leur éducation culturelle et non de leur formation religieuse. Ils sont souvent éduqués dans l'ignorance des questions se rapportant à la sexualité.

Références

[1] Abdelli, S. (2011), *La reproduction humaine et l'éducation à la sexualité en Tunisie*, Thèse de doctorat, Université Lyon 1.

[2] Chabchoub, A. (2000), *Rapports aux savoirs, Didactiques des sciences et Anthropologie,* in : Rapports aux savoirs et apprentissages scientifiques. Acte du 5ème colloque international de didactique et d'épistémologie des sciences, Tome I, GCP, Tunis, 37.

[3] Charlot, B. (2003), *La problématique du rapport au savoir,* Rapport au savoir et didactique, Fabert, 14.

[4] Chesnais, M., F. (1998), *Vers l'autonomie, L'accompagnement dans les apprentissages,* Hachette, 12.

[5] Farhane, F. (2012), *L'autonomie dans l'autocorrection/auto-évaluation dans l'apprentissage de la reproduction humaine en Tunisie au sein d'une pédagogie hybride entre non-sens et fructification,* Mémoire de master 2 recherche, Université Lyon 2.

[6] Gras, R., Régnier, J-C, Guillet, F. (Eds) (2009) Analyse statistique implicative. Une méthode d'analyse de données pour la recherche de causalités. RNTI-E- 16 Toulouse. Cépaduès Editions

[7] Hoffman-Gosset, M., A. (2006), *Apprendre l'autonomie, apprendre la socialisation,* Chronique sociale.

[8] JORT n°89, 7 novembre 1958, p.1056

[9] Liquète, V. et Y. Maury (2007), *Le travail autonome,* Colin, Paris.

[10] Rogers, C. A. (1996), *Liberté pour apprendre,* Dunod, Paris, 122.

Annexe

	H	F	V02	V03	V04	V05	V06	V07	V08	V09	V10	V11	V12	V13	V14	V15
I01	0	1	1	0	1	0	1	1	0	0	0	1	0	1	1	1
I02	0	1	1	0	1	0	0	1	0	0	0	1	0	1	0	1
I03	0	1	0	1	1	1	0	1	0	1	0	1	0	1	0	1
I04	0	1	1	1	0	0	1	0	1	1	1	0	1	0	0	0
I05	0	1	0	1	0	1	0	0	1	1	0	1	1	1	0	0
I06	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0	1
I07	0	1	1	0	1	1	0	1	0	0	0	1	0	1	0	1
I08	0	1	1	0	1	0	1	1	0	0	0	1	1	1	1	1
I09	0	1	0	0	1	1	0	1	1	1	0	1	0	0	0	1
I10	0	1	1	0	1	0	1	1	0	1	0	1	1	1	0	0
I11	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1
I12	0	1	0	0	0	0	1	0	1	0	1	1	1	1	1	1
I13	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0
I14	0	1	1	0	0	1	1	1	0	0	1	1	0	1	0	1
I15	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	1
I16	0	1	1	1	1	0	1	1	0	1	1	0	0	1	0	1
I17	0	1	1	0	0	0	1	1	0	1	0	1	0	1	0	1
I18	0	1	0	1	0	0	1	1	1	1	0	0	0	0	0	1
I19	0	1	1	0	0	0	1	1	0	1	1	1	1	1	1	0
I20	0	1	0	1	0	0	1	1	0	0	0	1	0	1	0	1
I21	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0	1
I22	0	1	1	0	1	0	1	1	0	1	0	1	0	1	1	1
I23	0	1	0	0	0	0	1	1	1	1	0	1	0	0	0	1
I24	0	1	0	1	0	0	0	1	1	0	1	1	0	0	1	1
I25	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0
I26	0	1	0	0	0	0	1	1	0	1	0	1	0	1	0	1
I27	0	1	0	1	0	0	0	1	1	1	0	1	0	0	1	1
I28	0	1	1	0	1	0	1	1	0	0	0	1	1	1	0	1
I29	0	1	1	0	0	0	0	1	0	1	0	1	0	1	1	1
I30	0	1	0	0	1	0	1	1	0	0	0	1	0	1	0	1
I31	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	1
I32	0	1	0	1	0	1	0	1	1	1	0	1	0	1	0	1
I33	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0
I34	0	1	1	0	0	0	1	1	0	0	0	1	0	1	0	1
I35	0	1	1	0	0	0	1	1	0	1	0	1	1	1	1	1

Tableau 1 – Résultats du questionnaire

	H	F	V02	V03	V04	V05	V06	V07	V08	V09	V10	V11	V12	V13	V14	V15
I36	0	1	1	0	1	1	1	0	0	1	1	1	1	1	0	1
I37	0	1	1	0	0	0	1	1	1	1	0	1	0	0	1	1
I38	0	1	1	0	0	0	0	0	0	1	0	1	1	1	0	1
I39	0	1	0	0	1	1	0	1	0	0	0	1	0	1	0	1
I40	0	1	1	0	1	1	1	0	0	1	0	1	1	1	0	1
I41	0	1	1	0	0	0	1	0	0	0	0	1	1	1	0	1
I42	0	1	1	0	0	1	1	0	0	0	0	1	1	0	0	0
I43	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	1
I44	0	1	0	1	0	0	0	1	1	1	0	1	0	0	0	1
I45	0	1	1	0	0	0	1	0	0	0	0	0	1	1	1	1
I46	0	1	0	0	1	0	1	0	0	0	0	1	1	1	0	1
I47	0	1	1	0	0	1	1	0	0	0	0	1	1	1	1	1
I48	0	1	1	0	1	0	0	1	0	1	0	1	0	0	0	0
I49	0	1	0	1	1	1	1	0	0	1	0	1	1	0	0	0
I50	0	1	1	1	0	0	1	1	1	1	0	0	0	1	0	1
I51	1	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1
I52	1	0	0	1	0	0	1	0	1	0	1	1	1	0	0	1
I53	1	0	1	0	0	0	1	1	1	0	1	1	1	0	1	1
I54	1	0	0	1	0	1	1	1	0	0	0	1	1	1	0	0
I55	1	0	0	1	0	0	1	1	1	0	0	1	1	1	0	0
I56	1	0	1	0	0	0	1	1	1	0	1	1	0	0	0	1
I57	1	0	0	1	1	0	1	1	0	0	0	0	1	0	0	0
I58	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0
I59	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0
I60	1	0	1	1	0	0	1	0	1	0	1	0	1	1	1	1
I61	1	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0
I62	1	0	0	0	1	0	0	0	1	0	0	1	1	1	0	1
I63	1	0	1	0	1	0	1	1	1	0	0	1	0	0	0	1
I64	1	0	1	1	0	0	1	1	1	0	0	0	1	1	0	1
I65	1	0	1	0	0	0	1	0	0	0	1	1	1	0	1	0
I66	1	0	1	1	1	0	1	1	1	1	1	0	1	0	1	0
I67	1	0	0	1	1	0	1	0	0	0	0	1	1	1	0	1
I68	1	0	1	0	0	0	1	1	1	0	1	1	1	0	1	0
I69	1	0	0	1	0	1	0	1	1	0	0	1	1	1	0	1
I70	1	0	1	0	0	0	1	1	1	0	1	1	1	0	0	0
I71	1	0	1	1	0	0	1	1	0	1	1	0	0	1	0	1

Tableau 2 – Résultats du questionnaire

ÉTUDE DES SIGNIFICATIONS DES FRACTIONS DONNÉES PAR DES ÉLÈVES DU CYCLE 3 DE L'ECOLE PRIMAIRE EN FRANCE.

Abdul Aziz ALAHMADATI[1]

STUDY OF FRACTIONS' MEANINGS GIVEN BY GRADE 3 PUPILS OF ELEMENTARY SCHOOL IN FRANCE.

RESUME

Le présent article s'intéresse à l'étude des conceptions des élèves de CM1 et CM2 à l'égard de la notion de fraction, en particulier l'étude des différentes significations données par ces élèves pour définir la fraction, pour donner un exemple sur l'utilisation des fractions au quotidien et l'illustration de la fraction 1/4. Diverses significations sont attribuées à la fraction : Partie d'un tout, Partie d'un ensemble, Opérateur, Mesure, Quotient, Nombre sur une droite numérique, Nombre et Probabilité. Pour servir de base à l'expression des variables du tableau de données, nous examinons l'ensemble plausible des conceptions à ces niveaux scolaires. Celles-ci, une fois recueillies, sont exploitées selon deux approches statistiques : la similarité des procédures associées aux conceptions et les relations implicatives entre ces procédures. Le graphe implicatif met en évidence la liaison non symétrique entre ces conceptions et la façon de les mettre en œuvre dans des situations concrètes.

Mots-clés : *Significations de la fraction, Analyse Statistique Implicative (A.S.I)*

ABSTRACT

This paper focuses on the study of French students' conceptions of CM1 and CM2 on the concept of fraction, in particular the study of the different meanings given by these students to the definition of fraction; to give an example on the use of fractions in everyday life, and to the illustration of the fraction 1/4.Various meanings are attributed to the fraction: Part of a whole, Part of a set, Operator, Measurement, Quotient, Number on a number line, Number and Probability. All plausible designings for this grade level are considered to be the basis for the expression of the data array variables. These, once collected, are used in two statistical approaches: the similarity of the procedures associated with designs and implicative relationships between these procedures. The implicative graph shows the non-symmetrical connection between these designings and the way to implement them in concrete situations.

Keywords: *Meanings of the fraction, Implicative statistical analysis (ISA)*

1 Introduction

Les fractions se retrouvent dans certaines activités de la vie quotidienne, certains métiers et certaines matières scolaires. Nous les utilisons dans nos activités courantes : partage (d'aliments par exemple), distances, comparaisons d'objets, heures, unités de mesure, etc.. L'apprentissage des fractions commence pleinement au troisième cycle de

[1] Université Lumière Lyon2, ED 485 EPIC, UMR 5191 ICAR, azezahmadty@yahoo.com

l'école primaire, surtout en CM1 et CM2. Le but d'enseigner cette notion à ce stade scolaire est d'amener les élèves à résoudre des problèmes variés portant sur les fractions. La compréhension des fractions est également importante chez les élèves pour poursuivre l'apprentissage d'autres notions comme les fractions algébriques.

2 Problématique générale de l'étude

Notre étude s'inscrit dans la recherche de doctorat menée à l'Université LYON 2 par A. A. Alahmadati sous la direction de J.-C. Régnier. Il s'agit de l'étude des conceptions des élèves de CM1 et de CM2 sur les différentes significations possibles attribuées à la fraction.

Les fractions sont parmi les concepts mathématiques les plus complexes rencontrés par les enfants dans les années du primaire (Charalambos et Pitta-Pantazi, 2005). De plus, dans l'introduction de son livre « Pourquoi ont-ils inventé les fractions », Nicolas Rouche (1998) dit : « Les fractions sont un des premiers et des principaux terrains où se développe le dégoût des mathématiques […] ».

Diverses raisons peuvent expliquer les difficultés liées à l'apprentissage des fractions : l'utilisation d'un symbole comportant deux nombres *numérateur* et *dénominateur*, le manque de diversité des représentations possibles des fractions, les connaissances antérieurs sur les nombres entiers qui interfèrent parfois avec l'apprentissage des fractions, les opérations sur les fractions qui impliquent beaucoup de règles que les élèves tendent à utiliser mécaniquement et la multiplicité des significations qui sont attribuables aux fractions. Ce dernier caractère plurivoque des fractions constitue une difficulté importante de leur apprentissage (Kieren, 1993; Brousseau, 2004). Par exemple, les activités qui portent sur les fractions se limitent souvent à la signification Partie d'un tout (Blouin, 2002 ; Adjiage et Pluvinage, 2000 ; Kieren, 1980). Ceci entraîne comme conséquence un répertoire limité de procédures chez les élèves au moment de résoudre un problème. Toutefois, plusieurs significations, attribuées à la fraction, sont possibles : Partie d'un tout, Partie d'un ensemble, Opérateur, Rapport, Quotient, Mesure, Nombre sur une ligne numérique, Nombre et Probabilité (Hétu et Desjardins, 1974 ; Kieren, 1980 ; Behr et al., 1983 ; Terrien, Dionne et Mura,1994 ; Rouche,1998 ; Adjiage et Pluvinage, 2000 ; Watanabe, 2002 ; Blouin, 2002) et l'utilisation de ces significations dans l'apprentissage des fractions peut faciliter la compréhension chez les élèves. Dans notre étude, nous avons décidé de les conserver toutes.

3 Le cadre théorique

Dans cette section, dans un premier temps, nous présentons la définition de la fraction et dans un second temps, nous décrivons les significations possibles de la fraction à partir de la recension des études précitées.

3.1 Définition de la fraction

La fraction est une notation pour représenter un nombre rationnel. Une fraction se compose donc de deux nombres entiers naturels, a et b, séparés par une barre

horizontale, alors a/b. Le chiffre du haut (a) s'appelle numérateur et le chiffre du bas (b) s'appelle dénominateur. Le dénominateur (b) est toujours différent du zéro (Corrieu, 1999). Il est nécessaire de situer les fractions par rapport aux nombres rationnels, il ne s'agit pas d'une fraction en tant qu'une notation, mais les nombres rationnels exprimés sous la forme de fractions. Selon de Champlain, Mathieu, Pathenaude et Tessier (1996), Le nombre rationnel est « un nombre obtenu à partir du quotient de a et b où a et b sont des entiers et b est différent de 0 ». Un même nombre rationnel peut s'exprimer sous la forme d'une fraction, d'un nombre décimal ou d'un pourcentage. Par exemple, les expressions 1/2; 0,5; 4/8; 0,50 et 5/10 représentent toutes le même nombre rationnel et chacune peut être utilisée pour le désigner. Un nombre rationnel est donc une classe de fractions équivalentes.

Pour cette recherche, nous ne nous intéresserons toutefois qu'au nombre rationnel représenté sous la forme d'une fraction.

3.2 Les significations possibles de la fraction

Les fractions relèvent d'une notion multiforme englobant plusieurs significations interdépendantes, c'est l'un des principaux facteurs de leur complexité (Charalambos et Pitta-Pantazi, 2005).

Les lectures faites sur les travaux de Hétu et Desjardins (1974), Kieren (1980), Behr et al. (1983), Terrien, Dionne et Mura (1994), Rouche (1998), Adjiage et Pluvinage (2000), Watanabe (2002) et Blouin (2002) nous ont permis de répertorier neuf significations accordées aux fractions. Ces significations possibles sont : Partie d'un tout, Partie d'un ensemble, Opérateur, Rapport, Quotient, Mesure, Nombre sur une ligne numérique, Nombre et Probabilité. Dans notre étude, nous avons décidé de les conserver toutes. Dans les paragraphes qui suivent, nous décrivons ces significations à partir de la recension des études précitées.

3.2.1 La fraction en tant que Partie d'un tout

La signification partie d'un tout de la fraction est définie comme une situation dans laquelle une quantité continue est partitionnée en parties de taille égale (Lamon, 1999). Selon Kieren (1980), Blouin (2002) et Watanabe (2002) la signification la plus commune est celle partie d'un tout. Le modèle partie d'un tout est considérée par Kieren (1980) comme la base des connaissances sur les fractions. Cette signification se base sur le principe de la division d'une quantité continue, en parties égales.

3.2.2 La fraction en tant que Partie d'un ensemble

Cette signification s'apparente beaucoup à la signification de partie d'un tout, le tout étant une quantité continue ou un seul objet. Plutôt qu'une grandeur continue soit divisée (un objet ou une région), une grandeur discrète, telle un ensemble, est divisée en parties égales et une part choisie. La fraction 5/18 peut représenter ici cinq objets de même nature sur un total de dix-huit.

3.2.3 La fraction en tant qu'Opérateur

Selon Blouin (2002), Rouche (1998) et Kieren (1980), la fraction peut être représentée comme un opérateur. Le cas « opérateur » désigne des situations où la fraction opère sur une quantité ou une mesure. Dans la signification opérateur, la fraction peut être considérée comme une fonction algébrique qui transforme des figures géométriques ou des ensembles d'objets (Behr et al, 1983). Autrement dit, lorsqu'une fraction opère sur un objet continu, il s'étend ou se rétracte de l'objet. Par exemple, pour une longueur de 6 et un opérateur de 2/3, le résultat serait $2 \times 6 \div 3$, ou 4. De même, quand une fraction opère sur un ensemble discret, il s'agit d'un multiplicateur ou d'un diviseur. Par exemple, si un ensemble contenait 12 objets a été opéré par 2/3, le résultat serait $12 \times 2 \div 3 = 24 \div 3 = 8$.

3.2.4 La fraction en tant que Rapport

Les fractions sont des nombres de la forme p/q, où p et q représentent des entiers qui sont sous la forme d'un rapport de nombres. Le rapport est une relation qui exprime la notion de quantité relative. Donc, il est plus correct de le considérer comme un indice comparatif plus tôt que comme un nombre. Lorsque deux rapports sont égaux, ils sont dits en proportion les uns aux autres.

Plusieurs auteurs comme Kieren (1980), Blouin (2002), Watanabe (2002) et Rouche (1998) évoquaient cette signification de la fraction. Celle-ci considère la fraction comme une comparaison entre deux quantités. Selon Rouche (1998), un rapport est une comparaison et il exprime une relation entre deux grandeurs ou entre deux ensembles finis d'objets.

3.2.5 La fraction en tant que Quotient

Cette signification est rattachée à la définition du nombre rationnel. Il s'agit de connaître que la barre d'une fraction est un symbole de division où le quotient (la fraction) est le résultat de la division dans laquelle le numérateur définit la quantité à être partagée et le dénominateur définit les partitions de la quantité. Ainsi, 1/3 est le résultat lorsque le numérateur 1 est divisé par le dénominateur 3, cet exemple fait référence à un contexte numérique. L'aspect « quotient » de la fraction se présente donc lorsque la fraction a/b est définie comme la division $a \div b$.

La signification de la fraction en tant que résultat d'une division est forte utile, grâce à cette interprétation, nous pouvons dire que 8/4 est équivalent à 2, tout comme 2/3 équivalent à 0.6. Dans ce contexte, la notation fractionnaire a/b est utilisée pour représenter le résultat de a (numérateur) divisé par b (dénominateur), c'est-à-dire, le résultat d'équations linéaires du type $b\,x = a$.

3.2.6 La fraction en tant que Mesure

Concevoir la fraction comme une mesure suppose l'existence d'une unité de mesure. L'unité de mesure peut être, entre autres, une mesure de longueur, d'aire, de volume, de temps ou d'argent. Ainsi, la fraction 3/4 serait le résultat de l'itération de la fraction unité 1/4, 3/4 serait donc 1/4 + 1/4 + 1/4. Selon Rouche (1998), les mesures expriment

des grandeurs, la fraction mesure est un rapport entre une grandeur mesurée et une autre grandeur de même espèce choisie comme unité de mesure.

3.2.7 La fraction en tant que Nombre sur une droite numérique

La fraction a effectivement droit à une place sur la droite graduée. Cette signification est liée à celle de mesure puisqu'elle fait appel à la signification des fractions comme mesure des longueurs.

La droite numérique est utile pour la comparaison des fractions, celle-ci, à partir des points, s'effectue en considérant le point le plus à droite comme représentant la fraction la plus grande. Cette connaissance provient de l'extension de la comparaison des nombres entiers, à l'aide d'une droite numérique à la comparaison des fractions. Le concept de la droite numérique peut renforcer la compréhension des enfants sur les concepts de l'ordre et sur l'équivalence des fractions.

3.2.8 La fraction en tant que Nombre

La signification de la fraction comme nombre désigne les fractions qui servent à effectuer des calculs où aucun autre contexte n'intervient. Les opérations arithmétiques, la comparaison de fractions et la notion de fractions équivalentes, semblent être des étapes importantes dans cette compréhension de la fraction comme nombre.

3.2.9 La fraction en tant que Probabilité

La probabilité est un contexte dans lequel nous utilisons les fractions. Il s'agit de saisir, d'évaluer ou de communiquer, des données relatives aux chances pour qu'un événement survienne. Cette catégorie est liée à celle de la partie d'un ensemble. En effet, si nous voulons trouver quelle est la probabilité de tirer une boule jaune dans une urne qui contient 10 boules jaunes et 13 boules vertes, nous construisons un rapport entre les boules jaunes et l'ensemble des boules (10/23). Cependant, comme la probabilité ne s'exprime qu'à l'aide d'une quantité se situant entre 0 et 1, nous la considérons comme une signification possible de la fraction, les fractions sont indispensables dans l'expression des probabilités. Enfin, il faut constater que Rouche (1998) mentionne que l'écriture fractionnaire est intéressante comme expression des probabilités et que les lois fondamentales des probabilités amènent à calculer avec des fractions.

4 Question et hypothèses de l'étude

Nous allons dans la section qui vient présenter la question et les deux hypothèses de notre étude.

4.1 Question de l'étude

Notre objectif est : connaître quelles significations les élèves de CM1 et de CM2 donnent à la fraction lorsqu'ils répondent aux trois questions simples que nous leur avons posées : définir la fraction, donner un exemple sur l'utilisation des fractions au

quotidien, illustrer une fraction donnée. Nous tenterons d'apporter des éléments d'explication et de réponse grâce à une analyse de données collectées à travers un questionnaire écrit soumis aux élèves de CM1 et de CM2.

4.2 Hypothèses de l'étude

Pour cette étude, nous avons formulé deux hypothèses:

Hypothèse 1 : Pour répondre aux trois questions qui leur ont été posées, la signification *partie d'un tout* est celle qui est la plus utilisée chez les élèves de CM1 et CM2.

Hypothèse 2 : Lorsque les élèves utilisent une des significations de la fraction pour définir la fraction, ils ont tendance à utiliser cette même signification pour donner un exemple sur l'utilisation des fractions au quotidien et/ou pour illustrer une fraction donnée. De plus, si les élèves ne donnent pas de réponse pour définir la fraction, ils ont tendance à ne pas donner de réponse ni pour donner un exemple sur l'utilisation des fractions au quotidien ni pour illustrer la faction 1/4.

5 Méthodologie de l'étude

Cette section vise à présenter d'abord l'outil retenu pour la construction des données. Puis, elle décrit les caractéristiques de notre échantillon d'élèves. Enfin, elle présente les variables retenues pour cette étude.

5.1 Outil de construction des données

Dans la recherche que nous présentons dans cette communication, nous ne posons que trois questions qui sont issues du questionnaire appliqué dans notre travail de doctorat. Ce questionnaire écrit comporte d'une part trois questions (ces questions prises pour notre étude) qui portent sur la définition de la fraction, donner un exemple sur l'utilisation des fractions au quotidien et illustrer une fraction, et d'autre part des questions concernent les différentes significations de la fraction. Ces trois questions vont nous servir, en effet, à répondre à la question posée dans cette étude, soit « Pour définir la fraction, pour donner un exemple sur l'utilisation des fractions au quotidien et pour illustrer une fraction : quelles significations les élèves de CM2 donnent-ils de la fraction ? ».

5.2 Les participants à l'étude

Les participants à notre étude sont des élèves. Nous décrivons maintenant les caractéristiques de notre échantillon d'élèves qui appartiennent à douze classes issues de neuf écoles primaires françaises, 7 classes de CM1 et 5 de CM2. Pour notre recherche, nous travaillerons avec des élèves de niveaux CM1 et CM2 qui représentent les deux derniers niveaux du primaire en France. En effet, 275 élèves ; 160 de CM1 et 115 de CM2 ont répondu aux trois questions suivantes :

Q1 : Pour toi, « Une fraction, qu'est-ce que c'est ?»

Q2 : Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi?

Q3 : Représente la fraction $\frac{1}{4}$?

5.3 Variables de l'étude

Nous avons travaillé avec 12 variables (V01 à V12), celles-ci seront utilisées pour répondre aux trois questions posées aux élèves. Nous avons défini neuf catégories possibles des fractions : Partie d'un tout, Partie d'un ensemble, Opérateur, Quotient, Mesure, Nombre sur une ligne numérique, Nombre, Rapport et Ecriture fractionnaire. Nous nous posons la question de notre étude, Quelles catégories des fractions sont utilisées par les élèves de CM1 et de CM2 pour répondre aux questions sur les trois idées : définir la fraction, donner un exemple sur l'utilisation des fractions au quotidien et illustrer la fraction $\frac{1}{4}$ Nous regarderons les liens qui existent entre ces trois idées concernant les différentes significations de la fraction définies précédemment. De plus, nous allons également, pour notre étude, prendre en compte la variable *Niveau Scolaire* (CM1 et CM2).

Voici le tableau qui présente les différentes catégories (significations) de la fraction avec les catégories *Non réponse et les niveaux scolaires CM1 et CM2*, celles-ci représentent les variables retenues pour l'étude :

Tableau 1- Types des réponses concernant (les variables de l'étude)

Q1 : Pour toi, « Une fraction, qu'est-ce que c'est ?» Q2 : Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi? Q3 : Représente la fraction 1/4 ?	
Types	**Modalité**
V01	CM1
V02	CM2
V03	Partie-tout
V04	Partie-ensemble
V05	Opérateur
V06	Quotient
V07	Mesure
V08	Nombre sur une ligne numérique
V09	Nombre
V10	Rapport
V11	Ecriture fractionnaire
V12	Non réponse

Nous présentons les notations utilisées pour l'analyse des données : apparaît d'abord la lettre Q qui énonce la question posée, celle-ci est suivie par un chiffre qui représente le numéro de cette question. Puis apparaît la lettre V qui énonce la variable utilisée, celle-ci est suivie également par des chiffres qui donnent le numéro de cette variable manifestée. Nous introduisons un trait pour séparer la question et la variable. Par exemple :

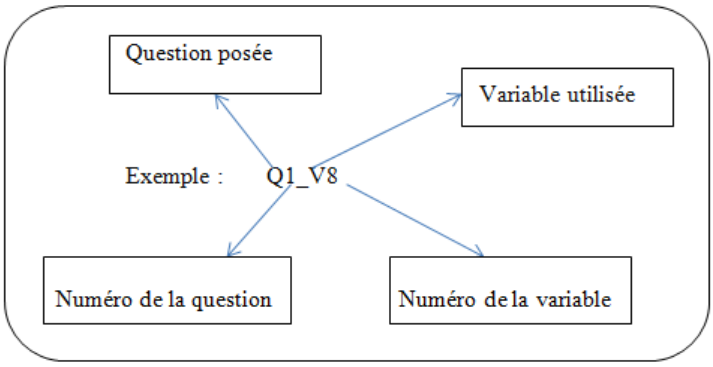


Figure 1 – La notation utilisée pour l'analyse des données

6 Analyse des réponses des élèves

Nous catégorisons les réponses fournies par les élèves aux trois questions par significations des fractions, nous les classons selon les définitions établies dans la section consacrée à la présentation des significations possibles de la fraction.

6.1 Analyse des réponses des élèves de CM1 et de CM2 à : *Pour toi, une fraction, qu'est-ce que c'est ?*

Le but de cette question est de savoir ce que l'élève comprend de la fraction et pour voir quelle signification l'élève utilise pour la définir. Les réponses que les élèves peuvent donner feront intervenir une des significations de la fraction : Partie-tout, Partie-ensemble, opérateur, rapport, quotient, mesure, nombre sur une droite numérique et nombre. De plus, les élèves peuvent considérer la fraction comme un symbole d'écriture du nombre rationnel qui contient un numérateur et un dénominateur.

L'analyse des réponses des élèves montre que la signification *Partie d'un tout* est celle qui est la plus utilisée chez les élèves de CM1 et de CM2 pour définir la fraction. En effet, 43,75% d'élèves de CM1 et 33,91% de CM2 l'utilisent pour définir la fraction. La signification *Partie d'un tout* consiste à une quantité continue (objet ou région) et à la subdiviser en parties égales afin de choisir certaines parties sur le tout. Les définitions données par les élèves des deux niveaux rejoignent cette définition. Par exemple : « C'est par exemple, un rond qu'on partage en 4 parties, et qui en a 2 parties qui sont coloriées on appelle deux quarts» (un élève du niveau CM1) ou encore « Si on prend un cercle et qu'on le partage en deux c'est un demi 1/2. Si on prend un carré et qu'on le partage en 4 et on en prend 3 c'est 3/4 » (un élève du niveau CM2). La signification *Nombre* est utilisée par 30,63% d'élèves de CM1 et par 19,13% de ceux de CM2 pour définir la fraction. Nous avons 13,75% (22/160 élèves) d'élèves de CM1 et 9,57% (11/115 élèves) d'élèves de CM2 qui n'ont pas répondu pour donner une définition à la

fraction. La signification *Rapport* n'est pas utilisée par les élèves des deux niveaux pour définir la fraction. Enfin, les autres significations de la fraction sont utilisées par 28,14% des élèves de CM1 et par 66,02% des élèves de CM2 pour définir la fraction.

6.2 Analyse des réponses des élèves de CM1 et de CM2 à : Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi?

Le but de la question est de savoir quelle signification vient à l'esprit de l'élève en premier. En effet, cette signification est probablement celle avec laquelle elle est la plus familière. Les exemples donnés par les élèves feront intervenir une des significations de la fraction : Partie-tout, Partie-ensemble, opérateur, rapport, quotient, mesure, nombre sur une droite numérique et nombre. De plus, ils peuvent donner un exemple d'écriture d'une fraction comme 1/4.

Chez les élèves des deux niveaux CM1 et CM2, la signification partie d'un tout se classe à la première place en tant que référence pour donner un exemple sur l'utilisation des fractions au quotidien. En effet, 41,88% d'élèves de CM1 et 33,04% d'élève de CM2 utilisent cette signification pour donner un exemple. Un bon nombre d'élèves de CM1 et de CM2 se réfèrent à la signification *Mesure* pour donner un exemple concret à l'emploi des fractions au quotidien. En effet, 13,13% des élèves de CM1 et 29,56% des élèves de CM2 utilisent cette signification pour donner un exemple. Par rapport à l'utilisation de la signification *Nombre*, 13,13% des élèves de CM1 et 9,57 % des élèves de CM2 font référence à cette signification pour donner un exemple. Nous constatons que les deux significations *Nombre sur une droite graduée* et *Rapport* ne sont pas utilisées par aucun élève des deux niveaux CM1 et CM2 pour donner un exemple sur l'utilisation des fractions au quotidien. De plus, nous avons 23,13% (37/160 élèves) des élèves de CM1 et 17,39% (20/115 élèves) des élèves de CM2 qui n'ont pas répondu pour donner une définition à la fraction. Enfin, 9,38% des élèves de CM1 et 15,68% des élèves de CM2 font référence aux autres significations de la fraction pour donner un exemple de l'utilisation des fractions dans la vie quotidienne.

6.3 Analyse des réponses des élèves de CM1 et de CM2 à : *Représente de deux façons différentes la fraction ?*

Le but de la question est de savoir quelle signification l'élève de CM2 est portée à utiliser pour représenter une fraction. En effet, il est probable que la première signification à laquelle l'élève songe est celle avec laquelle l'élève est le plus familier. Les illustrations que les élèves peuvent faire feront intervenir une des significations de la fraction : Partie-tout, Partie-ensemble, opérateur, rapport, quotient, mesure, nombre sur une droite numérique et nombre. De plus, les élèves peuvent écrire la fraction 3/4 sous une autre forme comme 6/8 ou 75/100, ce qui fait référence au symbole d'écriture d'un nombre rationnel.

Partie d'un tout est la première signification référée pour illustrer une fraction chez les élèves de CM1 et de CM2. En effet, 39,38% des élèves de CM1 et 60,87% de ceux de CM2 l'utilisent. La signification *Nombre sur une ligne graduée* vient à la deuxième place derrière celle de *Partie d'un tout* chez les élèves de CM1 et chez ceux de CM2. En effet, 20,63 % des élèves de CM1 et 6,09% des élèves de CM2 se réfèrent à cette

signification pour leur première illustration de la fraction $\frac{1}{4}$. Les élèves de CM1 et de CM2 font référence à la catégorie *Ecriture fractionnaire* pour illustrer la fraction $\frac{1}{4}$, elle se classe à la troisième place chez ceux de CM1 et en deuxième place avec la signification *Nombre sur une ligne graduée* chez les élèves de CM2. En effet, 13,13% des élèves de CM1 et 6,09% des élèves de CM2 utilisent cette catégorie pour leur première illustration de $\frac{1}{4}$. Par contre, nous constatons que les trois significations *Opérateur*, *Mesure* et *Rapport* ne sont pas utilisées par aucun des élèves des deux niveaux CM1 et CM2 afin d'illustrer la fraction $\frac{1}{4}$. Nous avons 13,13% (21/160 élèves) d'élèves de CM1 et 7,83% (9/115 élèves) des élèves de CM2 qui n'ont pas fourni une réponse afin d'illustrer ou représenter la fraction $\frac{1}{4}$. Enfin, 2, 15% des élèves de CM1 et 4,35% des élèves de CM2 font référence aux autres significations de la fraction pour leur première illustration de la fraction $\frac{1}{4}$.

6.4 Résultats concernant la première l'hypothèse de l'étude

Nous allons tenter de répondre à notre première hypothèse, celle-ci est « Pour répondre aux trois questions qui leur ont été posées, la signification *partie d'un tout* est celle qui est la plus utilisée chez les élèves de CM1 et CM2 ».

A travers les réponses données par les élèves de deux niveaux scolaires, CM1 et CM2, il apparaît que l'utilisation de la signification *Partie-Tout* de la fraction est celle que les élèves l'utiliser le plus. Ainsi, cela nous permet de constater que la première hypothèse posée est tout à fait vraisemblable. Pour expliquer : *Premièrement*, nous observons que la signification *Partie-Tout* est celle qui est la plus utilisée chez les élèves de CM1 et de CM2 pour définir la fraction. 43,75% d'élèves de CM1 et 33,91% de ceux de CM2 l'utilisent pour définir la fraction. *Deuxièmement*, chez les élèves des deux niveaux, la signification *Partie-Tout* se classe à la première place en tant que référence pour donner un exemple sur l'utilisation des fractions au quotidien. 41,88% d'élèves de CM1 et 33,04% d'élève de CM2 utilisent cette signification pour donner un exemple. *Troisièmement*, *Partie-Tout* est la première signification référée pour illustrer la fraction 1/4 chez les deux niveaux d'élèves. En effet, 39,38% des élèves de CM1 et 60,87% de ceux de CM2 l'utilisent.

6.5 Interprétation des résultats concernant la première hypothèse

Nous avons effectué une analyse sur 5 manuels de CM1 et 5 manuels de CM2 pour déterminer les différentes significations de la fraction qui sont exploitées dans ces manuels. Cette analyse montre que la signification *Partie-Tout* vient à la première place dans les manuels scolaires de CM2 et à la deuxième place dans ceux de CM2. Cela nous permet d'observer que la signification *Partie-Tout* est importante dans les activités portant sur les fractions. De plus, cette signification est également la plus utilisée chez les élèves des deux niveaux CM1 et CM2.

Partie d'un tout serait la première signification que les enfants auraient des fractions. Selon Kieren (1980) le modèle partie d'un tout s'avère le plus naturel. Behr et al. (1983)

suggèrent que le partitionnement et la signification partie/tout sont essentiels à l'apprentissage d'autres significations du nombre rationnel. Cela peut expliquer pourquoi la signification partie/tout de fractions a traditionnellement servi à initier les étudiants à l'instruction sur les fractions. Toutefois, certains auteurs Kieren (1980), Wearne-Hiebert et Hiebert (1983), Mack (1990), Sinicrope et Mick (1992), Adjiage et Pluvinage (2000), Rosar, Van Nieuwenhoven et Jonnaert (2001) et Blouin (2002) indiquent que les sens des activités portant sur les fractions qui se limitent souvent à la signification *Partie-Tout* entraînent un répertoire limité de procédures chez les élèves au moment de résoudre des problèmes. Les résultats obtenus convergent vers cette confirmation étant donné que la plupart des élèves utilisent souvent la signification *Partie-Tout* pour répondre à des questions portant sur les fractions comme le cas de nos trois questions posées dans notre étude.

7 Analyse et interprétation des résultats dans le cadre de l'Analyse Statistique Implicative pour la seconde hypothèse

L'approche Analyse Statistique Implicative (ASI) a été développée par Régis Gras (1979) et ses collaborateurs (Gras et *al*., 2009, 2013). L'analyse effectuée par cette méthode est instrumentée par le logiciel C.H.I.C sur lequel nous avons travaillé dans le séminaire doctoral ADATIC sous la direction de notre directeur du Professeur Jean-Claude Régnier. Grâce à cette méthode, nous exposons les résultats obtenus afin de voir si des liens existent entre les diverses variables préalablement explicitées (§ 5.3).

Les variables (V01 à V12) seront traitées comme variables binaires. Grâce à la méthode ASI, nous étudions l'arbre des similarités et le graphe implicatif. Nous allons présenter, tout d'abord, l'arbre des similarités avec tous les résultats obtenus. Puis, nous présentons le graphe implicatif et les résultats donnés et enfin, nous analysons et interprétons ces résultats. Nous commençons par la présentation de l'arbre des similarités.

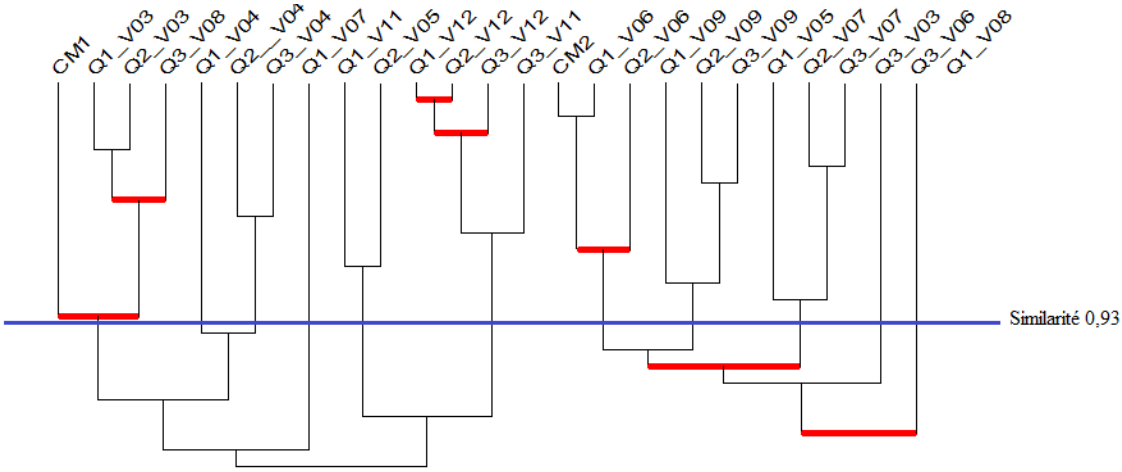


Figure 2 – Arbre des similarités

Comme le montre l'arbre des similarités, en se limitant au niveau de similarité de 0.93, nous distinguons 5 niveaux de classification présentés en lignes rouges sur le dessin) qui sont les suivants :

Tableau 2- Présentation les cinq niveaux de classification présentés dans l'arbre des similarités

Niveau de classification	Les variables concernées dans chaque niveau	Niveau de similarité
1	{(Q1_V12 Q2_V12)}	0,99
3	{(Q1_V12 Q2_V12) Q3_V12)}	0.99
7	{(Q1_V03 Q2_V03) Q3_V08)}	0.98
10	{(CM2 Q1_V06) Q2_V06)}	0.97
14	{(CM1 ((Q1_V03 Q2_V03) Q3_V08)}	0.93

Les caractéristiques des classes sont présentées dans le tableau suivant :

Tableau 3- Caractéristiques de niveau issu de l'arbre des similarités

Classification de niveau	Groupe optimal: cardinal	Classe qui a le plus contribué	sexe qui a le plus contribué	l'école qui a le plus contribué
1	19	CM1	Fille	F
3	6	CM1	Fille	H
7	12	CM1	Garçon	E
10	3	CM2	Garçon	C
14	11	CM1	Garçon	E

De ces données, nous constatons que les élèves de CM2 ont plus contribué à la classe « 10 » que ceux de CM1 alors que, les CM1 ont plus contribué aux classe « 1, 3, 7, 14 » que les CM2. De plus, les filles ont plus contribué aux classes « 1 et 3 » que les garçons alors que les garçons ont plus contribué aux classes « 7, 10 et 14 » que les filles. Enfin, parmi les écoles, l'école « E » a le plus contribué aux classes «7 et 14 », l'école « F » est celle qui a le plus contribué à la classe « 1 », celle de « H » est celle qui a le plus contribué à la classe « 3 » et l'école « C » est celle qui a le plus contribué à la classe « 10 ».

De ce qui précède, nous constatons que :

- Lorsque les élèves ne donnent pas de réponse à la première question (*Pour toi, une fraction, qu'est-ce que c'est ?*), alors, ils ne réussissent pas à donner de réponse à la deuxième question (*Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi ?*). La classe 1 confirme bien ce résultat.
- Lorsque les élèves ne donnent pas de réponse ni à la première question (*Pour toi, une fraction, qu'est-ce que c'est ?*) ni à la deuxième question (*Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi ?*), alors, ils ne réussissent pas à donner de réponse à la troisième question (*Représente la fraction* $\frac{1}{4}$ *?*). La classe 3confirme tout à fait ce résultat.

- Dans les deux classes « 7 et 14 » : lorsque les élèves utilisent la signification « Partie d'un tout » pour définir la fraction et pour donner un exemple sur l'utilisation des fractions au quotidien, ils ont tendance à utiliser la signification « Nombre sur la droite numérique » pour illustrer la fraction $\frac{1}{4}$.
- Nous constatons que lorsque les élèves de CM2 utilisent la signification de la fraction en tant que « Quotient » pour définir la fraction, ils ont tendance à utiliser cette même signification pour donner un exemple sur l'utilisation des fractions au quotidien.

Regardons maintenant les résultats que nous livre le graphe implicatif, sachant qu'il nous renseigne quant aux tendances entre les différentes variables.

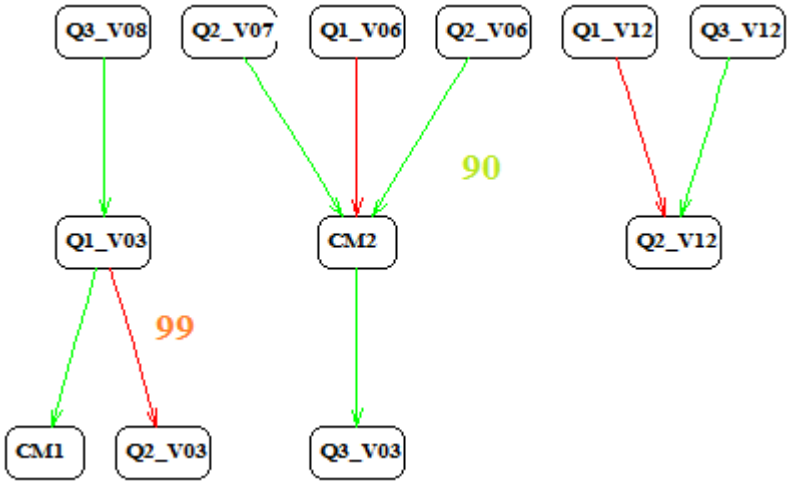


Figure 3 – Graphe implicatif

À un niveau de confiance de 0.99 (flèches rouges), nous avons les trois quasi-implications suivantes : la première est, si les élèves ne donnent pas de réponse à la question (« Donne un exemple où on peut utiliser les fractions dans la vie quotidienne et explique pourquoi? »), c'est que ces élèves ne donnent pas de réponse à la question (Pour toi, « Une fraction, qu'est-ce que c'est ?»). La deuxième quasi-implication est, lorsque les élèves utilisent la signification « Partie d'un tout » pour fournir une réponse à la première question concernant la définition de la fraction, ces élèves ont tendance à utiliser cette même signification « Partie d'un tout » pour fournir une réponse à la deuxième question concernant l'utilisation des fractions au quotidien. Un tel constat nous amène à confirmer la première partie de notre première hypothèse. Enfin, la troisième quasi-implication est, lorsque la réponse à la question (Pour toi, « Une fraction, qu'est-ce que c'est ?») est donnée en utilisant la signification « Quotient », il y a une tendance que le sujet répondant soit un élève du niveau CM2.

En descendant à un niveau de confiance plus bas, soit 0.90 (flèches vertes), nous voyons que lorsque les élèves ne donnent pas de réponse pour illustrer la fraction 1/4, c'est que ces élèves ne réussissent pas à donner de réponse pour fournir un exemple sur l'utilisation des fractions au quotidien. Un tel constat nous amène à confirmer la deuxième

partie de notre seconde hypothèse. Nous constatons que, si les réponses fournies pour illustrer la fraction 1/4 sont données en utilisation la signification « Partie d'un tout », les répondants sont des élèves du niveau CM2. Au même niveau de confiance, si les élèves utilisent la signification « Quotient » ou la signification « Mesure » pour donner un exemple sur l'utilisation des fractions au quotidien , c'est que ces élèves sont du niveau CM2. Si les élèves sont du niveau CM1, ils ont tendance à utiliser la signification « Partie d'un tout » pour donner une définition à la fraction. Cela nous peut-être dire que les élèves sont assez familiers avec cette signification ou que les élèves comprennent cette significations plus que les autres.

Nous allons maintenant donner une interprétation aux résultats obtenus par l'utilisation de la méthode Analyse Statistique Implicative (A.S.I): les activités qui portent sur les fractions se limitent souvent à la signification « Partie d'un tout » (Blouin, 200 ; Adjiage et Pluvinage, 2000 ; Kieren, 1980). Ceci entraîne comme conséquence un répertoire limité de procédures chez les élèves lorsqu'ils résolvent des problèmes. De plus, il nous semble que cette signification est la plus familière chez les sujets interrogés et nous pouvons dire que celle-ci est la première signification qui vient à l'esprit de l'élève lorsqu'il doit résoudre des problèmes portant sur les fractions. Par ailleurs, une exploitation très forte d'une des significations de la fraction dans les manuels scolaires et l'utilisation répétitive de cette signification de la part des enseignants influencent beaucoup sur l'utilisation de cette signification par les élèves. De plus, nous ne devons pas oublier l'importance de la signification « Nombre sur la droite numérique » qui peut renforcer la compréhension des enfants sur les concepts de l'ordre et sur l'équivalence des fractions.

Un bon nombre d'élèves de CM2 se réfèrent à la signification « Quotient » pour définir la fraction, soit 32,17% des élèves, alors que cette signification est complètement absente chez les élèves de CM1 pour la définition de la fraction. Pour donner un exemple sur l'utilisation des fractions au quotidien, 1,25% des élèves de CM1 et 11,3% des élèves de CM2 emploient effectivement cette signification. Cela pourrait s'expliquer par l'importance donnée à la signification « Nombre ». En effet, celle-ci se classe à la première place dans les manuels de CM2. De même, cette signification, nous l'avons vu selon notre analyse effectuée sur les manuels scolaires de CM2, domine l'ensemble des manuels étudiés de CM2. En effet, plus de la moitié des activités portant sur les fractions sont consacrées à la signification *Nombre*.

L'absence de réponses chez les élèves aux trois questions posées pourrait s'expliquer par leur manque de compréhension concernant la notion de fraction, par la complexité des fractions qui sont parmi les concepts mathématiques les plus complexes rencontrés par les enfants à l'école primaire (Charalambos et Pitta-Pantazi, 2005), par la limitation des situations d'apprentissage présentes aux élèves, etc..

8 Conclusion générale

L'apprentissage des fractions commence pleinement au troisième cycle de l'école primaire, surtout en CM1 et CM2. Plusieurs significations, attribuées à la fraction, sont possibles: *Partie d'un tout*, *Partie d'un ensemble*, *Opérateur*, *Rapport*, *Quotient*, *Mesure*, *Nombre sur une ligne numérique*, *Nombre* et *Probabilité*. L'analyse des réponses des 275 élèves (160 élèves de CM1 et 115 de CM2) nous permet de déterminer

les significations de la fraction manifestées par ces derniers. La signification *Partie d'un tout* est la plus utilisée par les élèves de CM1 et de CM2 pour définir la fraction, pour donner un exemple sur l'utilisation des fractions au quotidien et pour illustrer la fraction 1/4. La raison est que les activités qui portent sur les fractions se limitent souvent à la signification *Partie d'un tout*, ceci entraîne comme conséquence un répertoire limité de procédures chez les élèves au moment de résoudre des problèmes. L'analyse statistique implicative (A.S.I) nous permet d'étudier la relation existant entre les trois entrées suivantes ; définition de la fraction, donner un exemple sur l'utilisation des fractions dans la vie quotidienne et l'illustration de la fraction $\frac{1}{4}$. Pour répondre aux trois entrées précédentes, nos sujets de CM1 et de CM2 ont utilisé dix catégories concernant les fractions : Partie d'un tout, Partie d'un ensemble, Opérateur, Mesure, Quotient, Nombre sur une droite numérique, Nombre, Ecriture fractionnaire et Non réponse L'analyse des réponses de ces élèves nous montre que lorsqu'ils utilisent une des significations de la fraction pour définir la fraction, ils ont tendance à utiliser la même signification pour donner un exemple sur l'utilisation des fractions au quotidien et/ou illustrer la fraction 1/4. De plus, si les élèves ne donnent pas de réponse pour définir la fraction, ils ont tendance à ne pas donner de réponse ni pour donner un exemple sur l'utilisation des fractions au quotidien ni pour illustrer la faction 1/4.

Références

[1] Adjiage, R. et Pluvinage, F. (2000). Un registre unidimensionnel pour l'expression des rationnels. *Recherches en didactique des mathématiques*, *20* (1), 41-88.

[2] Behr, M. J., Lesh, R., Post, T. R. et Silver, E. A. (1983). Rational Numbers Concepts. In R. Lesh et M. Landau (dir.), *Acquisition of Mathematics Concepts and Processes* (P. 91-125). New York: Academic Press.

[3] Blouin, P. (2002). *Dessine-moi un bateau : la multiplication par un et demi*. Québec, Canada: Edition Bande Didactique.

[4] Brousseau, G. (2004). Les représentations : étude en théorie des situations didactiques. *Revue des sciences de l'éducation*, *30* (2), 241-277. Récupéré le 31 mai 2011 du site de la revue :

http://www.erudit.org/revue/rse/2004/v30/n2/012669ar.pdf

[5] Charalambos, Y. et Pitta-Pantazi, D. (2005). Revisiting a theoretical model on fractions: Implications for teaching and research. Dans H. L. Chick et J. L. Vincent (dir.), *Proceedings of the 29th PME Conference*, *2*, 233-240. Melbourne: University of Melbourne.

[6] Corrieu, L. (1999). *Dictionnaire du professeur des écoles, enseignement des mathématiques*. Paris, France: Vuibert.

[7] De Champlain, D., Mathieu, P., Patenaude, P. et Tessier, H. (1996). *Lexique mathématique : enseignement secondaire*. Les éditions du triangle d'or. Québec, Canada: Beauport.

[8] Gras R., Régnier J.-C. & Guillet F. (2009). *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités.* Toulouse: Cépaduès Editions.

[9] Gras R., Régnier J.-C., Marinica, C. & Guillet F. (2013). *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.* Toulouse: Cépaduès Editions.

[10] Hétu, J. C. et Desjardins, M. (1974). *L'activité mathématique dans l'enseignement des fractions*. Montréal, Canada : Presses de l'université du Québec.

[11] Kieren, T. E. (1980). Knowing rational numbers: ideas and symbols. Selected issues in mathematics education sous the direction of Mary Montgomery *Lindquist. Berkeley, Californie : Mccuthan Publishing Corporation*, USA, 69-81.

[12] Kieren, T. E. (1993). Rational and fractional numbers: From quotient fields to recursive understanding. Dans T. P. Carpenter, E. Fennema et T. A. Romberg (dir.), *Rational Numbers: An Integration of Research* (P. 49-84). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[13] Rosar, D., Van Nieuwenhoven, C. et Jonnaert, P. (2001). Les fractions, comment mieux comprendre les difficultés rencontrées par les élèves ? *Instantanées mathématiques, 32* (2), 4-16.

[14] Rouche, N. (1998). *L'esprit des sciences. Pourquoi ont-ils inventé les fractions?* Paris : Ellipses.

[15] Therrien, D., Dionne, J. et Mura, R. (1994). *La didactique de la mathématique*. Québec: Presses inter universitaires.

[16] Watanabe, T. (2002). Representations in teaching and learning fraction. T*eaching children mathematics, 8* (8), 457-463.

ANÁLISE DAS PESQUISAS DIDÁTICAS SOBRE FUNÇÃO AFIM NO ENSINO FUNDAMENTAL E MÉDIO NO QUADRO DA ANÁLISE ESTATÍSTICA IMPLICATIVA[1]

Aveilson José de SANTANA[2], Vladimir LIRA VERAS XAVIER DE ANDRADE[3], Jean-Claude RÉGNIER[4]

ANALYSE DES RECHERCHES DIDACTIQUES SUR LES FONCTIONS AFFINES AU COLLÈGE ET LYCÉE DANS LE CADRE DE L'ANALYSE STATISTIQUE IMPLICATIVE

ANALYSIS OF DIDACTIC RESEARCH STUDIES ON THE AFFINE FUNCTION IN ELEMENTARY AND HIGH SCHOOL EDUCATION IN THE FRAMEWORK OF STATISTICAL IMPLICATIVE ANALYSIS

RÉSUMO

O ensino da função afim tem uma importante contribuição para o desenvolvimento dos estudantes da educação básica. Apesar disso, observamos diversos estudos sobre as dificuldades dos alunos quando eles se deparam com situações que envolvem esse modelo matemático. Procuramos então fazer um levantamento das pesquisas que apresentam as dificuldades dos estudantes sobre a função afim. Para realizar este estudo, selecionamos artigos e dissertações produzidas no Brasil sobre o tema. Observamos que as dificuldades dos estudantes não são pontuais e muitas vezes se repetem em diferentes níveis e realidades de ensino. Neste estudo, utilizamos como ferramenta de análise dos dados a Análise Estatística Implicativa com a utilização do Software CHIC (Classificação Hierárquica Implicativa e Coesiva). Com os dados fornecidos pelo CHIC, verificamos que existem, em alguns casos, um índice de similaridade e implicação alto quando comparamos as dificuldades dos alunos nas pesquisas que foram levantadas.

Palavras-chave: *função afim, Ensino Médio, Ensino Fundamental, Analise Estatística Implicativa.*

RÉSUMÉ

L'enseignement des fonctions affines est une contribution majeure au développement des étudiants de l'école secondaire. Néanmoins, nous observons diverses études sur les difficultés des élèves quand ils son confrontés à des situations qui évoquent ce modèle mathématique. Nous cherchons alors à établir un état des lieux des recherches qui mettent en évidence les difficultés des élèves relativement aux fonctions affines. Pour réaliser cette étude, nous avons sélectionné des articles et thèses produites sur cet objet au Brésil. Nous avons observé que les difficultés des élèves ne sont pas ponctuelles et souvent se répètent à

[1] Esta pesquisa contou com o apoio da CAPES através da bolsa PVE para o programa PPGEC-UFRPE.com o Professor Jean-Claude Régnier UMR 5191 ICAR – Université Lyon2 França

[2] Programa de Pós-Graduação em Ensino das Ciências da Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros, s/n, Dois Irmãos - CEP: 52171-900 - Recife/PE, aveilsonsantana@gmail.com

[3] Departamento de Matemática/Programa de Pós-Graduação em Ensino das Ciências da Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros, s/n, Dois Irmãos - CEP: 52171-900 - Recife/PE, Brasil, vladiandrade@gmail.com

[4] UMR5191 – ICAR Université Lumiere Lyon 2 (FRA) – PVE - jean-claude.regnier@univ-lyon2.fr

différents niveaux et situations éducatives. Dans cette étude, nous avons utilisé comme un outil d'analyse de données l'Analyse Statistique Implicative avec l'aide du logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive). Avec les données fournies par CHIC, nous constatons qu'il y a, dans certains cas, un indice élevé de similarité et d'implication lorsque nous comparons les difficultés observées chez les étudiants dans les recherches identifiées.

Mots-clés : *fonctions affines, lycée, collège, l'analyse statistique Implicative.*

ABSTRACT

The teaching of the affine function has an important contribution to the development of students of basic education. Nevertheless, we note several studies on students' difficulties when they come across situations which involve this mathematical model. Therefore we seek to undertake a survey of research studies that show the difficulties that students have with the affine function. To conduct this study, we selected articles and dissertations that deal with the subject in Brazil. We observed that students' difficulties are not specific and are very often repeated at different levels and realities of teaching. In this study, the tool for analysis that we used was Statistical Implicative Analysis along with CHIC (Cohesive Hierarchical Implicative Classification) software. With the data provided by CHIC, we verified that there are, in some cases, an index of similarity and high implication when we compare the difficulties observed in the students in the research carried out.

Keywords: *the affine function, high school, elementary school, Statistical Implicative Analysis.*

1 Introdução

O conceito de função possui importância na formação matemática que faz parte da educação básica. Seja no Ensino Fundamental, seja no Ensino Médio, esse conteúdo possibilita a quem o estuda fazer interpretações diversas sobre conteúdos e problemas do cotidiano e também modelagens de diversos fenômenos presentes nas diversas áreas da educação. Nesse sentido, o conceito de função merece ser destacado dentre os diversos conceitos da Matemática abordados nos ensinos fundamental e médio. Santos, Silva e Almeida (2007) acreditam que o status ocupado por esse conteúdo está vinculado à necessidade do ser humano em explicar fenômenos relacionados à natureza e à sociedade através das regularidades existentes. Nessa perspectiva, é importante destacar que alguns documentos oficiais também fazem menção a esse conteúdo. Podemos citar como exemplo, Brasil (1998, p. 40) que diz: "é preciso que o aluno perceba a Matemática como um sistema de códigos e regras que a tornam uma linguagem de comunicação de ideias e permite modelar a realidade e interpretá-la". Ainda abordando a importância do estudo das funções, a pesquisadora Edna Maura Zuffi destaca a relevância do conceito de função para as ciências. Para essa pesquisadora, a ideia de função ultrapassa os domínios da Matemática há um bom tempo, oferecendo-se à Física, à Química, à Biologia, Economia e outras áreas. Isso é fortalecido particularmente em virtude do desenvolvimento tecnológico que se acentua na segunda metade do século XX (Zuffi, 2004).

Nesse contexto, o trabalho com funções possui um importante papel no ensino de matemática, pois este tópico, com sua gama de ferramentas e aplicações, possui em qualquer currículo escolar fundamental importância. Para Santos, Silva e Almeida (2007), o conceito de função é o que mais se destaca entre os conceitos matemáticos que são trabalhados no ensino médio. É importante ressaltar que na educação básica, os estudantes possuem várias dificuldades quando se deparam com situações que

envolvem o conceito de função, algumas delas são apontadas nos trabalhos de Braga (2009), Lopes (2003), Oliveira (1997), Maciel e Cardoso (2012) e Markovits, Eylon e Bruckheimer (1995).

Neste artigo, pretendemos realizar um estudo que consiste em analisar pesquisas realizadas no Brasil com estudantes que apresentaram dificuldades em desenvolver atividades que envolviam a função afim. Nosso interesse por esse tipo de função se deu pelo fato dele ser um dos primeiros modelos matemáticos a ser vivenciado pelos estudantes do ensino básico e também porque as dificuldades que os estudantes possuem com esse modelo matemático podem acompanha-los no Ensino Médio. Para realizar o tratamento dos dados, iremos utilizar a Análise Estatística Implicativa (A.S.I.). Para tanto, utilizaremos como ferramenta o software CHIC[5]. De acordo com Couturier, Bodin e Gras (2003, p. 01), o CHIC permite "extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecer um índice de qualidade de associação e de representar uma estruturação das variáveis obtida por meio destas regras". Dessa forma, o uso do CHIC nos permitiu, entre outras coisas, organizar os dados, descobrir semelhanças e implicações entre variáveis. Esperamos com essa pesquisa, trazer subsídios para que pesquisadores e professores de diversos níveis de ensino compreendam quais as principais dificuldades dos estudantes apontadas em pesquisas sobre função afim e também observar semelhanças e implicações entre as dificuldades apontadas pelas pesquisas. Procurando dessa forma, observar possíveis relações entre essas dificuldades.

2 Pesquisas sobre dificuldades na aprendizagem do conceito de função afim

Encontramos diversas pesquisas que mostram as dificuldades encontradas por estudantes do ensino básico, quando eles estavam trabalhando com a função afim. Nesse contexto, se inserem as pesquisas de Dornelas (2007), Delgado (2010) e Ferreira (2013). Esses trabalhos apontam para a dificuldade que os alunos possuem ao fazer a conversão entre os registros de representação de uma função. Nesse contexto, Dornelas (2007) formulou e aplicou uma sequência didática com momentos baseados na Teoria das Situações Didáticas que propiciou aos alunos situações de aprendizagens com o conceito de função afim. Essa pesquisadora propôs atividades que envolviam a conversão entre os registros de representação em linguagem natural, na forma gráfica, tabular e algébrica. Dornelas (2007) relata em seu trabalho algumas dificuldades enfrentadas pelos alunos, entre elas estão aquelas relacionadas ao registro em linguagem natural. Esse tipo de situação foi proposta em alguns problemas, de modo que os alunos se desvinculassem da linguagem algébrica. Outra dificuldade relatada por essa pesquisadora consiste na conversão do registro algébrico para o gráfico. Dessa maneira, durante o levantamento que fizemos para desenvolver esse capítulo, verificamos também que os pesquisadores também falam de outras limitações enfrentadas pelos estudantes. Como exemplo, podemos citar Postal (2009), que propôs aos alunos, entre outras atividades, algumas questões que envolviam a identificação dos coeficientes da função afim e também a representação de uma situação matemática através da função afim. Para Fonseca (2011), o trabalho com a função afim, para muitos alunos, se reduz a

[5] Abreviação de Classificação Hierárquica Implicativa e Coesiva.

operações, substituindo o número numa incógnita que está presente na lei de formação. De acordo com esse pesquisador, quando os sujeitos de sua pesquisa se depararam com situações que envolviam a representação de uma situação matemática através de uma função afim, eles não conseguiram compreender que os fenômenos que possuem um comportamento regular podem ser generalizados através de uma lei de formação.

Observamos nas pesquisas realizadas por Dornelas (2007), Delgado (2010), Ferreira (2013) e Fonseca (2011) que os alunos encontram muitas dificuldades quando precisam fazer a conversão entre as formas gráficas, algébrica, tabular e linguagem natural da função afim. Vale salientar que alguns dos percalços encontrados nas pesquisas em relação ao conceito de função, também estão presentes no trabalho com a função afim. Nesse caso, podemos citar o trabalho de Maciel (2012) que afirma que as dificuldades encontradas pelos alunos em relação ao conceito de função se referem à generalização de situações que envolviam o conceito de função. Esse mesmo entrave também é encontrado em alguns trabalhos que trataram sobre a função afim, como foi visto anteriormente. Isso também foi observado nos estudos de Fonseca (2011), Dornelas (2007), Delgado (2010), Ferreira (2013) e Postal (2009). Para Dornellas (2007), as dificuldades dos estudantes em generalizar situações que envolvem a função afim podem estar relacionadas ao fato de o registro numérico ter sido o recurso mais explorado na resolução de problemas durante a vida escolar dos alunos. Dornellas (2007) ainda complementa dizendo que vários estudantes não conseguem observar o crescimento ou decrescimento da função afim e nem organizar esses dados através de uma tabela. Dessa forma, alguns sujeitos de sua pesquisa somente conseguiram determinar o comportamento da função afim através da reta que representa o seu gráfico.

Para Delgado (2010), a abstração exigida quando os alunos lidam com representações algébricas, traz muita dificuldade ao trabalho com a função afim, já que os estudantes não possuem o hábito de se dedicar às situações que contribuem para desenvolver a capacidade de abstração. Seja em relação ao conceito de função ou a função afim, nas pesquisas que foram citadas no decorrer desse capítulo, chamou a nossa atenção a quantidade de trabalhos que citam a análise gráfica como sendo um dos principais entraves encontrados pelos estudantes. É importante deixar claro que, nesse caso, analisar o gráfico da função afim se refere à análise e interpretação de informações apresentadas na forma gráfica, como por exemplo, observar o seu comportamento, identificar pontos onde a reta corta os eixos e a determinação de sua lei de formação. Esse mesmo obstáculo é relatado nas pesquisas de Delgado (2010), Ferreira (2013), Delgado, Friedmann e Lima (2010), Luz (2010), Cardoso et al (2013) e Magalhães (2009). Conforme podemos observar no estudo realizado por Delgado Friedmann e Lima (2010, p. 8): “Em relação à forma gráfica, muitos alunos conseguem fazer a conversão da forma tabular para a gráfica com alguma facilidade, mas o inverso não ocorre da mesma maneira. Eles não conseguem analisar um gráfico de forma satisfatória. Em questões que envolveram interpretação de gráfico, a maioria dos erros ocorreu por não associar as variáveis, derivadas da situação-problema, com os valores representados por cada ponto pertencente à curva, no plano cartesiano”.

Nos trabalhos citados anteriormente, encontramos também outras dificuldades associadas à aprendizagem de função afim. Essas limitações dizem respeito à marcação de pontos no plano cartesiano, construção de gráficos somente vinculados a uma tabela, ao traçado de retas no plano, diferenciação entre as funções afim, constante e linear,

marcação de pontos nos quais as ordenadas ou abscissas são nulas e a verificação do sinal da função. Algumas dessas limitações estão presentes nos trabalhos de Menezes e Filho (2010). Para Delgado (2010), o fato de a função permitir aos estudantes trabalharem com diversas representações, e também por esse conceito envolver situações que estão em outras áreas do conhecimento, traz dificuldades extras ao seu entendimento. Para Andrade e Dias (2009), é preciso que os alunos trabalhem com situações que envolvam a resolução de problemas que tenham a função afim como modelo matemático. Esses pesquisadores concordam com Delgado (2010) quando dizem que uma parte dos estudantes não enxerga esse conceito como uma ferramenta que pode ser aplicada em outros contextos. Podemos observar nas pesquisas trazidas aqui, que as dificuldades relacionadas ao trabalho com a função afim não são pontuais e mostram que vários dos obstáculos encontrados são comuns em diferentes realidades e níveis de ensino. Assim, há uma preocupação dos pesquisadores não só em mostrar as dificuldades encontradas, mas também em tentar minimizá-las. No quadro 1, apresentamos uma síntese das dificuldades que foram levantadas pelos autores das pesquisas. Como apresentado nesse quadro, várias dificuldades são comuns à diferentes pesquisas.

Autor/Ano	Dificuldades
Luz (2010)	1. Distinguir incógnita de variável. 2. Converter da linguagem natural para a algébrica. 3. Analisar o gráfico.
Dornelas (2007)	1. Analisar o gráfico. 2. Observar o crescimento ou decrescimento da função a partir da relação entre o domínio e imagem da função. 3. Generalizar resultados. 4. Converter da linguagem numérica para a natural. 5. Converter da linguagem natural para a tabular. 6. Converter da linguagem natural para a numérica. 7. Localizar pontos nos eixos cartesianos. 8. Estabelecer relações de dependência entre as variáveis.
Fonseca (2011)	1. Converter da linguagem numérica para a natural 2. Converter da linguagem natural para a algébrica. 3. Converter da linguagem natural para a tabular. 4. Converter da linguagem natural para a numérica. 5. Generalizar resultados. 6. Estabelecer relações de dependência entre as variáveis.
Postal (2009)	1. Identificar o domínio e imagem. 2. Identificar da taxa de variação da função. 3. Generalizar resultados.
Lima (2013)	1. Construir gráficos. 2. Analisar o gráfico.
Schonardie (2011)	1. Generalizar resultados. 2. Construir gráficos. 3. Analisar o gráfico.
Delgado (2010)	1. Converter da linguagem natural para a algébrica. 2. Manipular a expressão algébrica da função. 3. Converter da linguagem algébrica para a gráfica. 4. Converter da linguagem algébrica para a tabular. 5. Converter da linguagem algébrica para a natural. 5. Analisar o gráfico. 6. Generalizar resultados.
Cardoso et al (2013)	1. Localizar pontos nos eixos cartesianos. 2. Analisar o gráfico.

Braga (2009)	1. Localizar pontos nos eixos cartesianos 2. Analisar o gráfico.
Filho (2011)	1. Localizar pontos nos eixos cartesianos. 2. Analisar o gráfico.
Menezes e Filho (2010)	1. Localizar pontos nos eixos cartesianos. 2. Analisar o gráfico.

Quadro 1– Dificuldades encontradas nas pesquisas sobre função afim.

Tratamos a seguir da metodologia utilizada para o desenvolvimento de nossa pesquisa.

3 Metodologia

Esta pesquisa tem como objetivo principal levantar as principais dificuldades encontradas por pesquisadores que trabalharam com a função afim no ensino básico no Brasil e observar a relação entre as dificuldades presentes nesses estudos. Para realizar este levantamento, fizemos uma busca no banco de dissertações e teses da CAPES, na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), no site do Centro de Estudos, Memória e Pesquisa em Educação Matemática (CEMPEM), no endereço eletrônico de algumas universidades públicas e privadas, assim como em alguns eventos como o Encontro Nacional de Educação Matemática (ENEM – 2007/2010/2013) e o Simpósio Internacional de Educação Matemática (SIPEM – 2012/2009/2006). Dentro da amostra indicada, procuramos observar apenas os trabalhos que envolviam os ensinos Fundamental e Médio e também aqueles que deixavam claro quais foram as dificuldades enfrentadas pelos alunos. Nesse contexto encontramos um total de onze trabalhos, sendo dois artigos (Cardoso et al (2013) e Menezes e Filho (2010)) e nove dissertações. É importante destacar também que do total de pesquisas somente duas (Schonardie (2011) e Braga (2009)) envolviam o Ensino Fundamental. No quadro 2, apresentamos a codificação das variáveis que utilizamos para o tratamento no CHIC.

Variável	A pesquisa identifica dificuldades dos alunos em:
V01	Analisar gráficos ? Sim=1; Não =0.
V02	Construir gráficos ? Sim=1; Não =0.
V03	Converter da linguagem algébrica para a gráfica? Sim=1; Não =0.
V04	Converter da linguagem algébrica para a natural? Sim=1; Não =0.
V05	Converter da linguagem algébrica para a tabular? Sim=1; Não =0.
V06	Converter da linguagem natural para a algébrica? Sim=1; Não =0.
V07	Converter da linguagem natural para a numérica? Sim=1; Não =0.
V08	Converter da linguagem natural para a tabular? Sim=1; Não =0.
V09	Converter da linguagem numérica para a natural? Sim=1; Não =0.
V10	Distinguir incógnita de variável? Sim=1; Não =0.
V11	Estabelecer relações de dependência entre as variáveis? Sim=1; Não =0.
V12	Generalizar resultados? Sim=1; Não =0.
V13	Identificar da taxa de variação da função? Sim=1; Não =0.
V14	Identificar o domínio e imagem? Sim=1; Não =0.
V15	Localizar pontos nos eixos cartesianos? Sim=1; Não =0.
V16	Manipular a expressão algébrica da função? Sim=1; Não =0.
V17	Observar o crescimento ou decrescimento da função a partir da relação entre o domínio e imagem da função? Sim=1; Não =0.

Quadro 2 – Codificação das variáveis desta pesquisa.

Vale salientar que apesar dessas dificuldades terem sido apontadas pelos pesquisadores, nós não utilizamos as mesmas palavras dos autores. Podemos citar como exemplo a dificuldade em Construir da função a partir de sua lei de formação, que foi

apresentada por Ferreira (2013). Nesse caso, inferimos que essa dificuldade se enquadra na conversão do registro algébrico para o gráfico. Para análise dos dados, utilizamos a Análise Estatística Implicativa e o software CHIC. Para o tratamento dos dados, consideramos a presença ou ausência de uma categoria transformando os resultados em binários. Os dados foram organizados no Excel e depois importados para o CHIC no formato CSV. Com o auxílio do CHIC, foram realizados três tratamentos distintos, são eles: árvore de similaridade; grafo implicativo e árvore coesitiva. Posteriormente, iremos abordar as principais características de cada um desses tratamentos. Nessa etapa, também observamos no CHIC a contribuição indivíduos (pesquisas que foram levantadas) para que determinada classe seja formada. Assim, com o objetivo de organizar a nossa análise, a mesma foi dividida em três etapas, na qual cada uma delas representa um tratamento distinto para as variáveis. A seguir, tratamos dos resultados, análises e discussões.

4 Resultados, análises e discussões

Para o tratamento no CHIC, selecionamos a opção de nós significativos com cálculo longo, segundo a teoria clássica com a utilização da lei binominal. Apresentamos a seguir os elementos utilizados na análise.

4.1 Árvore de similaridades

De acordo com Couturier, Bodin e Gras (2003), a árvore de similaridade apresentada pelo CHIC efetua uma análise das proximidades, segundo I.C. Lerman (1981), produzindo uma janela de resultados numéricos e outra janela apresentando uma árvore hierárquica de similaridades. Para esses pesquisadores, a similaridade é definida quando há um cruzamento entre os conjuntos de variáveis e de sujeitos. Essa análise, além de permitir ao usuário observar a existência ou não de similaridade, também mostra os níveis em que elas acontecem. Na figura 1, está representada a árvore de similaridade elaborada com base nos dados desta pesquisa.

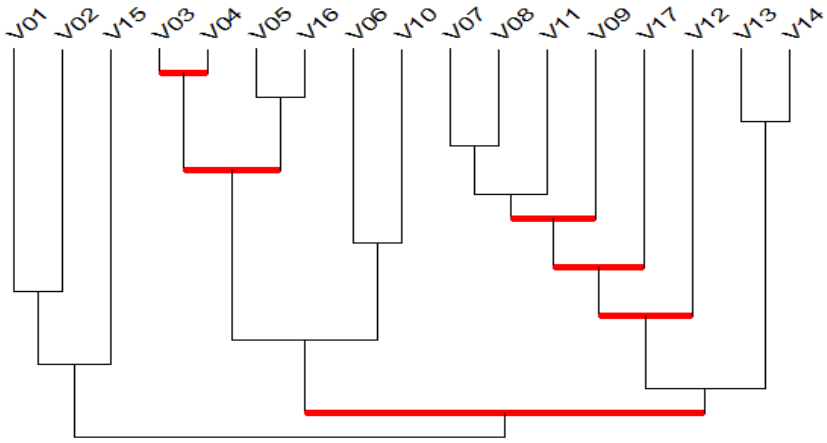


Figura 1 – Árvore de similaridade construída no CHIC.

De acordo com os dados da figura 1, podemos dizer que as variáveis V03 (converter da linguagem algébrica para a gráfica) e V04 (converter da linguagem algébrica para a natural) são as mais semelhantes, já que esse par está reunido no primeiro nível da árvore e tem um índice de similaridade de 0,99. Verificamos que nesse caso, o nível de

similaridade alto se deve ao fato das dificuldades representadas pelas variáveis V03 e V04 estarem presentes somente na pesquisa de Delgado (2010). Também existem outras variáveis que possuem um alto índice de similaridade porque estão presentes somente em uma das pesquisas. Esse é o caso das variáveis V16 (manipular a expressão algébrica da função) e também de V13 (identificar a taxa de variação) e V14 (identificar o domínio e a imagem da função), que formam o segundo e terceiro níveis de similaridade, respectivamente. O quarto nível de similaridade é formado pelas variáveis V07 (Converter da linguagem natural para a numérica) e V08 (Converter da linguagem natural para a tabular) e nesse contexto podemos dizer que existe uma possibilidade alta (0,996) das pesquisas que trazem a dificuldade V07 trazerem também V08, ou se não trouxerem V07 também não trazer V08. Nesse sentido, é importante destacar que embora essas dificuldades também possam estar presentes em alunos do Ensino Fundamental, observamos que elas se encontram somente em pesquisas que envolvem alunos do 1º ano do Ensino Médio. O décimo nível de similaridade possui um índice de 0,74 e é composto pelas variáveis V01 (analisar o gráfico) e V02 (construir gráficos). Nesse aspecto, temos um índice confiável para afirmar que as pesquisas que apontam para a dificuldade V01 também o fazem para a dificuldade V02. Nesse contexto, identificamos trabalhos que envolvem os ensinos Fundamental e Médio, isso nos permite dizer que essas dificuldades podem ser comuns a alunos de diferentes níveis de ensino. Isso confirma a nossa hipótese de que algumas dificuldades em relação ao trabalho com a função afim podem acompanhar os alunos durante a passagem para o ensino médio. Acrescentando a essa discussão, o CHIC também nos fornece os valores de similaridade entre todas as variáveis. Nesse aspecto, apresentamos no quadro 3, a classificação e o índice de similaridade das dificuldades descritas pelos pesquisadores.

Classificação ao nível 1	(V03 V04) similaridade : 0.998716
Classificação ao nível 2	(V05 V16) similaridade : 0.998716
Classificação ao nível 3	(V13 V14) similaridade : 0.998716
Classificação ao nível 4	(V07 V08) similaridade : 0.996672
Classificação ao nível 5	((V03 V04) (V05 V16)) similaridade : 0.994872
Classificação ao nível 6	((V07 V08) V11) similaridade : 0.993355
Classificação ao nível 7	(((V07 V08) V11) V09) similaridade : 0.928429
Classificação ao nível 8	(V06 V10) similaridade : 0.918133
Classificação ao nível 9	((((V07 V08) V11) V09) V17) similaridade : 0.894439
Classificação ao nível 10	(V01 V02) similaridade : 0.740426
Classificação ao nível: 11	(((((V07 V08) V11) V09) V17) V12) similaridade : 0.657069

Quadro 3 – Níveis e índices de similaridade.

Vale salientar que esses valores nos permitem analisar não só os níveis de similaridade entre variáveis ou classe de variáveis, mas também o valor do índice existente entre cada uma delas. Consideramos com esses resultados, que a classificação até o nível 11 apresenta um índice aceitável. Abaixo desse valor, não podemos indicar uma relação de similaridade confiável entre as dificuldades que são apresentadas pelas pesquisas.

Também iremos analisar no CHIC a contribuição dos sujeitos dentro de um determinado nível de similaridade. No quadro 4, tomamos como exemplo a contribuição em relação ao quarto nível.

Contribuição à classe : V07,V08 (4)
Contribuição dos indivíduos:
Braga (2009) : 0.000 Cardoso et al (2013) : 0.000 Delgado (2010) : 0.000 Dornelas (2007) : 1.000
Filho (2011) : 0.000 Fonseca (2011) : 1.000

Lima (2013) : 0.000 Luz (2010) : 0.000 Menezes e Filho (2010) : 0.000 Postal (2009) : 0.000 Schonardie (2011) : 0.000 Grupo ótima : card 2 Fonseca (2011) e Dornelas (2007)

Quadro 4 – Contribuições à V2, V5.

A classe formada pelas dificuldades V07 (converter da linguagem natural para a numérica) e V08 (converter da linguagem natural para a tabular) é formada por conversões que envolvem a linguagem natural e é mais influenciada por pesquisas desenvolvidas com alunos do Ensino Médio, que nesse caso são representadas por Dornelas (2007) e Fonseca (2011).

Conforme já citamos, também iremos mostrar a tipicalidade dos sujeitos (artigos e dissertações) de nossa pesquisa. Essa tipicalidade mostra quais são as pesquisas mais típicas em relação a determinado nível de similaridade.

Tipicalidade à classe : V03,V04 (1)
Tipicalidade dos indivíduos: Braga (2009) : 0.000 Cardoso et al (2013) : 0.000 Delgado (2010) : 0.999 Dornelas (2007) : 0.000 Filho (2011) : 0.000 Fonseca (2011) : 0.000 Lima (2013) : 0.000 Luz (2010) : 0.000 Menezes e Filho (2010) : 0.000 Postal (2009) : 0.000 Schonardie (2011) : 0.000 Grupo ótima : card 1 Delgado (2010)

Quadro 5 – Contribuições à V2, V5.

Consideramos somente a tipicalidade em relação à classe V03 – V04, pois a mesma possui o maior índice de similaridade. Percebemos que a pesquisa de Delgado (2010), que trabalhou com alunos do Ensino Médio, é a mais típica pra esse tipo de similaridade.

4.2 2ª Etapa de análise: Grafo implicativo

O grafo implicativo, conforme indicado por Couturier, Bodin e Gras (2003), realiza os cálculos dos índices de implicação utilizando a análise implicativa clássica ou entrópica. Na figura 2, temos o grafo que foi gerado utilizando a implicação segundo a teoria clássica.

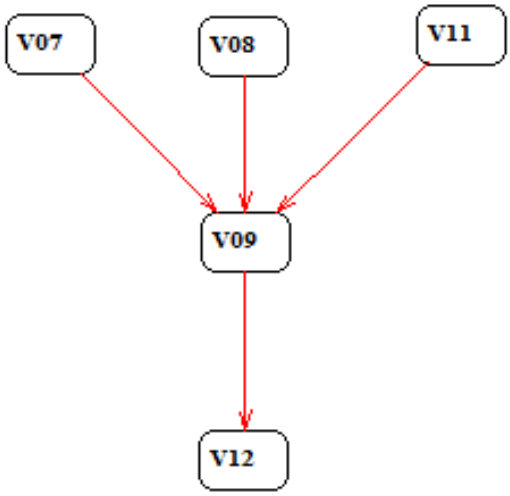


Figura 2 – Grafo implicativo.

Em relação à figura 2, podemos dizer, considerando os índices de implicação adotados, que as pesquisas que trazem a dificuldade representada pela variável V07 (converter da linguagem natural para a numérica) tendem a trazer a dificuldade V09 (converter da linguagem numérica para a natural). A implicação entre as variáveis V07 e V09 mostram que os estudantes podem ter dificuldades tanto na conversão que termina como na que começa com a linguagem natural. Essa informação fornecida pelo CHIC, corrobora com o que defende Duval (2011), pois para ele muitas vezes a passagem inversa da conversão entre os registros de representação traz dificuldades extras para os estudantes. Em relação a conversão de registros que envolvem a linguagem natural, Duval (2012, p. 295), diz que "A língua natural deve ser considerada, ao mesmo tempo, um registro de partida e um registro de chegada".

Em relação às implicações entre as variáveis V08 (converter da linguagem natural para a tabular) e V09 (converter da linguagem numérica para a natural) e também entre V11 (estabelecer relações de dependência entre as variáveis) e V09 (converter da linguagem numérica para a natural), percebemos mais uma vez a presença de dificuldades que envolvem a conversão em linguagem natural. Essas implicações reforçam a ideia de que é preciso que os professores tenham um olhar mais atento ao desenvolvimento de atividades que envolvam a conversão de registro com a utilização da linguagem natural, pois além de ser uma dificuldade de vários estudantes também, na visão de Dornelas (2007), essas atividades contribuem para que seja sanada a crença na qual as questões de matemática são somente resolvidas através de algoritmos pré-determinados. Na figura 2, temos também que quando se observa a dificuldade V09 (converter da linguagem numérica para a natural) nas pesquisas, também pode se observar a dificuldade V12 (generalizar resultados), ou seja, a dificuldade em fazer a conversão da linguagem numérica para a natural influencia nas que envolvem a generalização de resultados. Em síntese, chamou a nossa atenção a quantidade de implicações que envolvem as dificuldades de conversão entre os registros de representação da função afim. Isso corrobora com o que defende Duval (2005), pois para ele a conversão entre os registros de representação de objetos matemáticos traz várias dificuldades aos alunos. As implicações contidas na figura 2, também nos mostram que as dificuldades na realização da conversão de alguns registros de representação também podem acarretar em dificuldades de estabelecer relações de dependência entre variáveis e na generalização de resultados. De acordo com as

implicações fornecidas pelo CHIC, observamos que a dificuldade na conversão entre alguns registros de representação da função afim contribuem para que os estudantes tenham dificuldades em compreender o conceito desse tipo de função. Isso vai de encontro ao que pensa Andrade e Kaiber (2011), pois para elas a capacidade de dominar e manipular notações matemáticas é uma habilidade importante para que os sujeitos se apropriem de conhecimentos matemáticos. Prosseguindo com a nossa análise, trouxemos o quadro 6, onde mostramos as contribuições dos indivíduos (pesquisas) presentes no grafo implicativo. Vale ressaltar que na contribuição fornecida pelo CHIC observamos a contribuição ao caminho V09 – V12, já que esse é o caminho que possui o maior índice de confiança (0,83).

Contribuição ao caminho: V09-V12
Contribuição dos indivíduos:
Braga (2009) : 0.500 Cardoso et al (2013) : 0.500 Delgado (2010) : 1.000 Dornelas (2007) : 1.000
Filho (2011) : 0.500 Fonseca (2011) : 1.000
Lima (2013) : 0.500 Luz (2010) : 0.500 Menezes e Filho (2010) : 0.500 Postal (2009) : 0.682
Schonardie (2011) : 0.682
Grupo ótima : card 3
Delgado (2010) Dornelas (2007) Fonseca (2011)

Quadro 6 – Contribuição dos indivíduos em relação ao caminho V09-V12.

De acordo com o quadro 6, podemos perceber que a implicação V09 (converter da linguagem numérica para a natural) - V12 (generalizar resultados) é mais influenciada por pesquisas que foram desenvolvidas no Ensino Médio. Nesse sentido, poderíamos pensar em quais foram as pesquisas que mais influenciaram na formação dessa quase implicação e nesse caso, temos as pesquisas de Delgado (2010), Dornelas (2007) e Fonseca (2011). Assim podemos dizer que a implicação V09 – V12 foi mais influenciada por pesquisas que foram realizadas no Ensino Médio.

Apresentamos a seguir a tipicalidade em relação ao caminho V09 – V12.

Tipicalidade ao caminho: V09-V12
Tipicalidade dos indivíduos:
Braga (2009) : 0.000 Cardoso et al (2013) : 0.000 Delgado (2010) : 0.485 Dornelas (2007) : 0.485
Filho (2011) : 0.000 Fonseca (2011) : 0.485
Lima (2013) : 0.000 Luz (2010) : 0.000 Menezes e Filho (2010) : 0.000 Postal (2009) : 0.553
Schonardie (2011) : 0.553
Grupo ótima : card 5
Delgado (2010) Dornelas (2007) Fonseca (2011) Schonardie (2011) Postal (2009)

Quadro 7 – dos indivíduos em relação ao caminho V09-V12.

Conforme podemos perceber no quadro 7, as pesquisas mais típicas para esse tipo de implicação (V09-V12) são as de Delgado (2010) Dornelas (2007) Fonseca (2011) Schonardie (2011) Postal (2009). Desse modo, o fato dos alunos que possuíam dificuldade V09 (converter da linguagem numérica para a natural) também poderem ter V12 (generalizar resultados), é típico de estudantes dos ensinos Fundamental e Médio.

4.3 3ª etapa de análise: árvore coesitiva

A árvore coesiva ou coesitiva, segundo Couturier, Bodin e Gras (2003, p.08), "efetua os cálculos dos índices de coesão implicativa no sentido da análise implicativa". Nessa linha, apresentamos a seguir a árvore coesitiva que foi construída a partir dos dados da tabela 1.

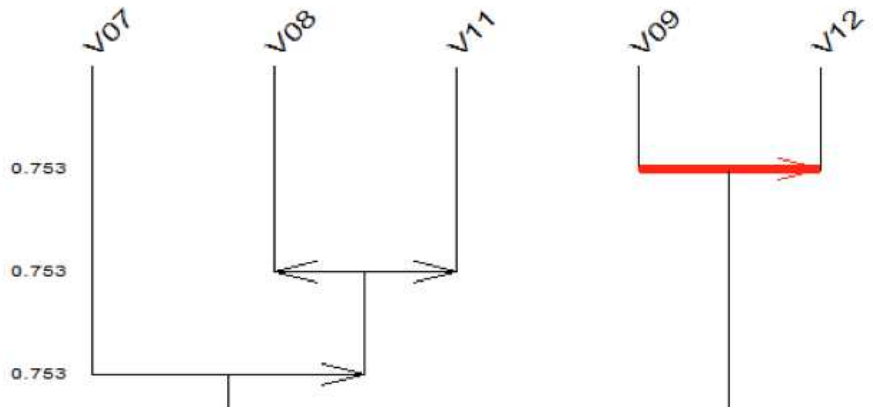


Figura 3 – Árvore coesitiva.

Nessa árvore coesitiva temos várias implicações e nós significativos. Vale salientar que apresentamos na figura 3, apenas os índices a partir de 0,75. Nesse aspecto, considerando as variáveis, descrevemos a seguir o significado dessas coesões, a partir dos dados mostrados na figura 2.

V09 → V12: Esse caso é o que possui uma relação estatística mais significativa e indica que as pesquisas levantadas que apontam para a dificuldade dos alunos em fazer conversões da linguagem numérica para a natural, também podem apontar para a dificuldade na generalização de resultados.

V08 ↔ V11: Nesse caso temos uma equivalência entre as variáveis e significa dizer que as pesquisas que trazem a dificuldade dos alunos em converter da linguagem natural para a tabular, também podem trazer a dificuldade em estabelecer relações de dependência entre variáveis. É importante destacar que como temos uma equivalência, pode-se também ter a implicação na ordem inversa, ou seja, de V11 para V08. Consideramos esse achado importante, porque ele nos mostra que para os estudantes conseguir estabelecer relações entre variáveis da função afim, eles precisam compreender a conversão existente entre linguagem natural e a construção de uma tabela.

V07 → (V08 ↔ V11): Essa implicação possui índice de confiança de 0,753 e significa dizer que as pesquisas que mostram a dificuldade dos alunos em fazer a conversões da linguagem natural para a numérica e da natural para a tabular, também podem trazer as dificuldades relacionadas ao estabelecimento de relações de dependência entre as variáveis. Nesse contexto, observamos mais uma relação direta entre a conversão de registros que envolvem a linguagem natural. Nesse caso percebemos mais uma vez a importância que esse registro de representação possui para o desenvolvimento de atividades pelos alunos.

Conforme fizemos com os outros tratamentos realizados no CHIC, também iremos mostrar a contribuição e a tipicalidade dos indivíduos. Esses casos estão ilustrados nos quadros 8 e 9, respectivamente. Para essa análise, consideramos somente a contribuição e a tipicalidade em relação ao maior nível de coesão que é representado pelas variáveis . Também mostramos a seguir a figura 3 que representa o índice de coesão entre as variáveis.

Contribuição à classe : V09,V12 (1)
Contribuição dos indivíduos:
Braga (2009) : 0.500 Cardoso et al (2013) : 0.500 Delgado (2010) : 1.000 Dornelas (2007) : 1.000
Filho (2011) : 0.500 Fonseca (2011) : 1.000
Lima (2013) : 0.500 Luz (2010) : 0.500 Menezes e Filho (2010) : 0.500 Postal (2009) : 0.682
Schonardie (2011) : 0.682
Grupo ótima : card 3
Delgado (2010) Dornelas (2007) Fonseca (2011)

Quadro 8 – Contribuição dos indivíduos à classe V09,V12.

Tipicalidade à classe : V09,V12 (1)
Tipicalidade dos indivíduos:
Braga (2009) : 0.000 Cardoso et al (2013) : 0.000 Delgado (2010) : 0.485 Dornelas (2007) : 0.485
Filho (2011) : 0.000 Fonseca (2011) : 0.485
Lima (2013) : 0.000 Luz (2010) : 0.000 Menezes e Filho (2010) : 0.000 Postal (2009) : 0.553
Schonardie (2011) : 0.553
Grupo ótima : card 5
Delgado (2010) Dornelas (2007) Fonseca (2011) Schonardie (2011) Postal (2009)

Quadro 9 – Tipicalidade dos indivíduos à classe V09,V12.

No quadro 8, percebemos que as pesquisas de Delgado (2010) Dornelas (2007) e Fonseca (2011) são as que mais contribuem para a formação da coesão da classe V09 – V12. Em relação à tipicalidade verificamos que as pesquisas mais típicas para esse nível de coesão são Delgado (2010) Dornelas (2007) Fonseca (2011) Schonardie (2011) Postal (2009). Temos ainda que somente pesquisas que trabalharam com alunos do Ensino Médio contribuem ou são típicas desse tipo de coesão.

5 Conclusão

Ao longo desse trabalho, trouxemos algumas pesquisas que envolviam a função afim para mostrar as principais dificuldades identificadas pelos pesquisadores quando eles trabalharam com estudantes dos ensinos Fundamental e Médio. Também realizamos vários tratamentos no CHIC, afim de observar como as dificuldades estavam relacionadas. De acordo com os dados que foram fornecidos pelo software, chegamos a algumas conclusões que a nosso ver são importantes. Nessa discussão, pudemos verificar que algumas das dificuldades encontradas pelos pesquisadores possuem um alto grau de semelhança, com por exemplo a classe formada pelas variáveis V03 (Converter da linguagem algébrica para a gráfica e V04 (Converter da linguagem algébrica para a natural). Em relação às implicações também encontramos resultados que julgamos relevantes, como por exemplo, o fato da presença das variáveis V07, V08 e V11 poderem implicar na presença de V09. Com esse resultado, inferimos que existe uma relação entre as dificuldades que envolvem a conversão entre registros que possuem a linguagem natural. Verificamos também que o registro de representação em linguagem natural é uma dificuldade que os alunos enfrentam quando, esse tipo de registro, é de partida ou de chegada, ou seja, os estudantes apresentam dificuldades nos dois sentidos da conversão. Percebemos ainda que pesquisas que mais apresentam relações entre si, foram aquelas desenvolvidas com alunos do Ensino Médio e também que em alguns casos, as mesmas dificuldades quem foram apresentadas por alunos do Ensino Fundamental, também estiveram presentes em estudantes do Ensino Médio.

Referências

[1] Andrade, L.S. e Kaiber, C. T. (2011), Registros de representação semiótica e o estudo das funções. *Anais do XIII CIAEM, Recife, Brasil, 2011.*

[2] Braga, E. (2009), A compreensão dos conceitos das funções afim e quadrática no ensino fundamental com o recurso da planilha. Dissertação de mestrado. Pontifícia Universidade Católica do Rio Grande do Sul.

[3] Brasil. (1998), Parâmetros Curriculares Nacionais. Secretaria da Educação Básica.

[4] Cardoso, M. B.; Santana, L. E. L.; SILVA, S. H.; FERREIRA, M. A. S. (2010), Função afim: uma análise dos procedimentos de conversão de aluno Couturier, Bodin e Gras Couturier, Bodin e Gras s do 2º ano do ensino médio. *Anais do Encontro Nacional de Educação Matemática (ENEM)*, 11, 1-15.

[5] Couturier, R.; Bodin, A.; Gras, R. (2003), A Classificação Hierárquica Implicativa e Coesiva. *Colóquio Internacional sobre Análise Estatística Implicativa*, 2, 1-22.

[6] Delgado, C. J. B. (2010), O Ensino da função afim a partir dos registros de representação semiótica. Dissertação de mestrado. Universidade do Grande Rio.

[7] Delgado, C. J. B.; Friedmann, C. V. P.; Lima, J. C. P. (2010), As dificuldades apresentadas por alunos do 1º ano do ensino médio em relação às diferentes representações da função afim. *Anais do Encontro Nacional de Educação Matemática (ENEM)*, 10, 1-11.

[8] Dornelas, J. J. B. (2007), Análise de uma sequência didática para a aprendizagem do conceito de função afim. Dissertação de mestrado. Universidade Federal Rural de Pernambuco.

[9] Duval, R. (2005), Transformations de représentations sémiotiques et démarches de pensée em mathématiques. *Anais do XXXII Colloque COPIRELEM Des Professeurs et des Formateurs de Mathématiques charges de la Formation des Maîtres,2005, 67 – 89.*

[10] Duval, R. (2011), Gráficos e equações: a articulação de dois registros. *Anais da REVEMAT, Florianópolis (SC), v. 6, n. 2, p. 96-112, 2011.*

[11] Duval, R. (2012), Registros de representação semiótica e funcionamento cognitivo do pensamento. *Anais da REVEMAT, Florianópolis (SC), v.7, n. 2, p. 266-297, 2012.*

[12] Ferreira, R. D. (2013), Contribuições do geogebra para o estudo de funções afim e quadrática em um curso de licenciatura em Matemática. Dissertação de mestrado. Pontifícia Universidade Católica de São Paulo.

[13] Filho, L. G. (2011), Modelagem Matemática e o ensino de função do 1º grau. 2011. 140 f. Dissertação de mestrado. Pontifícia Universidade Católica de São Paulo.

[14] Fonseca, V. G. (2011), O uso de tecnologias no ensino médio: a integração de mathlets no ensino da função afim. Dissertação de mestrado. Universidade Federal do Rio de Janeiro.

[15] Lima, C. E. O. (2013), A utilização do software geogebra como ferramenta para o ensino de funções. Dissertação de mestrado. Universidade Federal do Ceará. Fortaleza.

[16] Lopes, W. S. (2003), A importância da utilização de múltiplas representações no desenvolvimento do conceito de função: uma proposta de ensino. Dissertação de mestrado. Pontifícia Universidade Católica de São Paulo.

[17] Luz, S. V. (2010), Aprendizagem significativa de função do 1º grau: uma investigação por meio da modelagem Matemática e dos mapas conceituais. Dissertação de mestrado. Universidade Estadual de Maringá. Maringá.

[18] Maciel, P. R. C. (2011), A construção do conceito de função através da história da Matemática. Dissertação de mestrado. Centro Federal de Educação Tecnológica Celso Suckow da Fonseca.

[19] Maciel, P. R. C. e Cardoso, T. F. L. (2012), O conceito de função através da história da Matemática. *Anais do Seminário Internacional de pesquisa em educação matemática (SIPEM)*, 5, 1-19.

[20] Magalhães, A. R. (2009), Mapas conceituais digitais como estratégia para o desenvolvimento da metacognição no estudo de funções. Tese de doutorado. Pontifícia Universidade Católica de São Paulo.

[21] Markovits, Z. e Eylon, B. S.; Bruckheimer, M. (1995), Dificuldades dos alunos com o conceito de função. As ideias da álgebra. (tradução de Hygino H. Domingues). 49-69.

[22] Menezes, J. E. e Filho, M. S. M. (2010), Como os alunos do ensino médio estão construindo e interpretando gráficos de funções polinomiais de 1º e 2º graus. *Anais do Encontro Nacional de Educação Matemática (ENEM)*, 10, 1-10.

[23] Oliveira, N. (1997). Conceito de função: uma abordagem do processo ensino-aprendizagem. Dissertação de mestrado. Pontifícia Universidade Católica de São Paulo.

[24] Postal, R. F. (2009), Atividades de modelagem Matemática visando a uma aprendizagem significativa de funções afins, fazendo uso do computador como ferramenta de ensino. Dissertação de mestrado. Centro Universitário Univates.

[25] Santos, F. V.; Silva, K. A. P. e Almeida, L. M. W. (2007), O uso do computador no estudo de função no ensino médio. *Anais do Encontro Nacional de Educação Matemática (ENEM)*, 9, 1-22.

[26] Schönardie, B. (2011), Modelagem matemática e introdução da função afim no ensino fundamental. Dissertação de mestrado. Universidade Federal do Rio Grande do Sul.

[27] Zuffi, E. M. (2004), Uma sequência didática sobre funções para a formação de professores do ensino médio. *Anais do Encontro Nacional de Educação Matemática (ENEM)*, 8, 1-16.

REPRESENTATIONS DES ENSEIGNANTS GABONAIS SUR L'EDUCATION A LA SANTE

Laurence NDONG[1]

REPRESENTATIONS OF GABONESE TEACHERS IN HEALTH EDUCATION

RÉSUMÉ

A travers cette étude, menée dans le cadre de l'enquête européenne « BIOHEAD Citizen », nous avons cherché à identifier les représentations que les enseignants et futurs enseignants gabonais ont de l'éducation à la santé et les valeurs implicites qu'elles induisent. Nous nous sommes interrogés sur ce que sont pour eux les buts de l'éducation à la santé, quels devraient en être les principaux acteurs et les enjeux qu'elle représente. Les représentations des enseignants et futurs enseignants de notre échantillon ont été mises en relation avec leur discipline d'enseignement, leur niveau de formation et la pratique ou non d'une religion. Les résultats obtenus à l'aide notamment de l'Analyse Statistique Implicative (ASI) montrent que les enseignants et futurs enseignants de notre échantillon ont plutôt une approche positiviste de l'éducation à la santé.

Mots-clés : *Education à la santé, enseignants, Gabon, modèle KVP, représentations, valeurs.*

ABSTRACT

Through this study led within the framework of the European survey "BIOHEAD Citizen", we tried to identify the representations that the teachers and the future Gabonese teachers have of the health education and the implicit values which they lead. We wondered about the fact that are for them the purposes of the health education, which should be the main actors and the stakes which it represents. The representations of the teachers and the future teachers of our sample were got in touch with their discipline of education, their level of formation and the practice or not of a religion. The results obtained with the help of the Statistical Analysis Implicative (ASI) show that the teachers and the future teachers of our sample have rather a positivist approach of the health education.

Keywords: *Health education, teachers, Gabon, KVP model, representations, values.*

1 Introduction

1.1 But de l'étude

Le travail que nous présentons ici a été réalisé dans le cadre d'un projet européen dénommé « BIOHEAD-CITIZEN, Education à la biologie, à la santé et l'environnement pour une meilleure citoyenneté » dont le but était de voir comment sont enseignés des thèmes sensibles comme l'éducation à la santé, l'évolution, l'éducation à l'environnement... Et analyser les interactions entre connaissances scientifiques et systèmes de valeurs. En fonction de quels paramètres ces connaissances

[1] Laboratoire de Recherche en Education (LARED), ENS de Libreville, BP : 17009 laurence.ndong@gmail.com

et leur enseignement varient ? Du contexte socio-économique, culturel, religieux, de l'histoire récente du pays, de la discipline enseignée… ?

Le projet BIOHEAD-CITIZEN comparait les données recueillies dans 19 pays d'Europe, du Moyen-Orient, du Maghreb et d'Afrique de l'Ouest lors de sa phase initiale. D'autres pays tels que le Gabon y ont été rajouté ultérieurement.

1.2 Problématique et question de recherche

Depuis son émergence dans les années 70, la recherche en didactique des sciences s'est appliquée à produire un certain nombre de concepts ou en utiliser d'autres venus d'ailleurs comme de la didactique des mathématiques, ou d'autres domaines comme la psychologie sociale pour être repris, transformés ou encore créés *de novo* pour les besoins théoriques d'analyse (Astolfi et al. 1997). C'est le cas du concept de représentations sociales, emprunté à la sociologie de Durkheim et à la psychologie sociale (Moscovici, Abric, Jodelet, Dortier…).

Les représentations sont « sociales » en un double sens. D'une part, elles portent sur des phénomènes sociaux (le travail, la politique, les groupes humains, l'art…), et d'autre part elles sont issues et héritées de la société (Dortier, 1999). Ainsi, on peut admettre qu'elles se construisent par l'interaction entre l'individu et le monde qui l'entoure (Clément 2010). C'est en s'appuyant sur ce constat que Clément (1994, 2004, 2008), dans le cadre des recherches en didactique, propose d'analyser les représentations comme de possibles interactions entre trois pôles : les connaissances scientifiques (K comme knowledge), les valeurs (V) et les pratiques sociales (P). Il propose le modèle KVP que nous avons retenu pour cette étude.

La didactique portant sur des contenus de savoirs précis, dans une recherche en didactique des sciences les représentations sont toujours relatives à un thème précis qui est souvent un objectif d'enseignement (Clément, 2010).

Longtemps étudiées au niveau des élèves, depuis quelques années, les représentations des enseignants sont l'objet de plusieurs études (Désautels et Larochelle, 1999). En effet, l'élaboration d'interventions pédagogiques ne se fait pas dans le vide. Elle suppose que l'on se soit forgée une image (représentation) de la discipline que l'on enseigne, une image de ce que sont les élèves, de ce qu'ils savent déjà et une image de la pratique enseignante (Astolfi, 1998). C'est pourquoi nous nous intéressons aux représentations des enseignants sur l'éducation à la santé et les valeurs qui sous-tendent leur choix dans le cadre de cette dernière.

Pour cet article, nos questions de recherche sont :

- Quelle perception les enseignants et futurs enseignants de notre échantillon ont-ils de la santé d'une part ;
- D'autre part, quels sont pour eux les buts de l'éducation à la santé et quels devraient en être les principaux acteurs ?

1.3 Cadre de référence et recension des écrits

1.3.1 Le modèle « KVP »

Notre cadre de référence s'articule autour du modèle KVP de Pierre Clément (1994, 2004, 2008). Dans le modèle KVP (figure 1), les connaissances scientifiques de l'enseignant (le pôle K) proviennent de son cursus scolaire et de ses autres apprentissages provenant de diverses sources (manuels scolaires, revues spécialisées, etc.). Connaissances, tout en étant assez proches dans leur contenu des connaissances spécialisées, s'en différencient aussi fortement. Le pôle des pratiques sociales (P) introduit par Martinand (1986) renvoie à des activités sociales diverses (activités de recherche, de production, d'ingénierie, mais aussi activités domestiques et culturelles) pouvant servir de référence à des activités scolaires et devenant elles-mêmes un objectif d'enseignement. Outre cette fonction de référence à l'enseignement, le pôle P renvoie aussi aux pratiques professionnelles, citoyennes et à toute autre pratique sociale des acteurs dont les représentations sont étudiées. Ainsi, les connaissances scientifiques d'un enseignant dépendent aussi de ses pratiques (interactions KP). Il s'agit des pratiques personnelles et/ou sociales de l'enseignant (ses pratiques en tant que citoyen, en dehors de l'école et de ses pratiques professionnelles dans l'école y compris ses pratiques pédagogiques dans la classe).

Pour ce qui est des valeurs, elles sous-tendent la plupart des pratiques sociales. Chaque personne fait des choix qui reflètent son système de valeurs. Dans ce modèle, Clément reprend la définition des valeurs comme étant « ce qui fonde le jugement » c'est-à-dire ce qui peut être retrouvé à la base des opinions, des croyances, des idéologies, des positions philosophiques, morales ou éthiques, mais aussi de la science elle-même. En effet, les choix des chercheurs sont fondés sur des valeurs que la science est censée cultiver. Des valeurs telles que la modestie, l'esprit critique, la rationalité de la preuve, de la démonstration etc. (interaction KV). Les valeurs de l'enseignant ou de l'apprenant font référence à ses opinions, son idéologie et ses croyances.

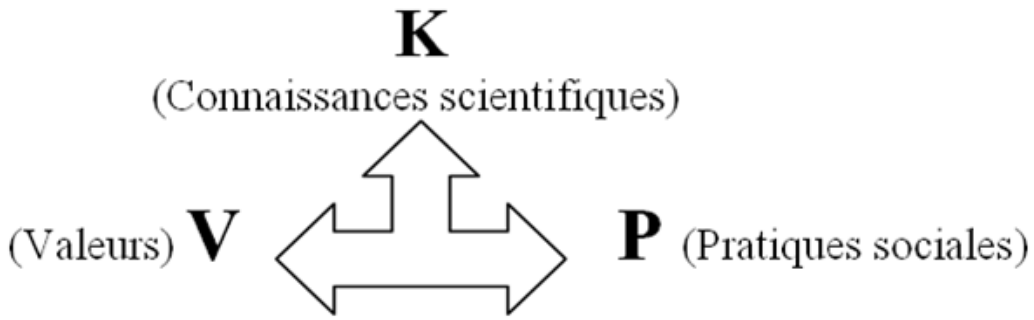


Figure n°1 : Modèle KVP (Clément, 1998, 2004)

Nous analyserons donc les représentations des enseignants et futurs enseignants gabonais sur le thème de l'éducation à la santé en interaction avec leurs valeurs. Nous privilégions donc le pôle KV. Pour ce qui est des connaissances, nous allons comparer

les représentations des enseignants de notre échantillon avec les paradigmes admis dans la communauté scientifique.

1.3.2 Education à la santé : définition, buts et acteurs

Souvent, les termes « éducation à la santé », «éducation pour la santé » et « éducation en santé » sont utilisés de façon interchangeable. Eymard (2004), dans un essai de modélisation entre éducation et santé propose de débattre sur la pertinence de la différenciation grammaticale et langagière entre ces termes.

Pour elle, les éducations « à » et « pour » la santé semblent offrir le même sens en ce qui concerne les liens entre éducation et santé. En effet, ces expressions relatent le plus souvent un lien de but entre éducation et santé, la préposition « pour » indiquant l'intention de l'éducateur envers l'éduqué à savoir celle d'acquérir une bonne santé ou un meilleur état de santé et la préposition « à » précisant la direction de l'acte d'éducation ou l'objet auquel elle se rapporte à savoir la santé. Dans ce cadre là, la santé de l'autre est alors l'aboutissement de l'acte éducatif. Elle devient une chose, un objet à acquérir. Il s'agit donc de faire acquérir un comportement du sujet ou du groupe conforme aux normes assurant une bonne santé. Ainsi, la mise en œuvre d'une démarche d'éducation pour la santé est liée à un déficit de santé réel ou supposé. La durée de la démarche éducative est donc liée à la correction des comportements de l'éduqué.

Dans cet article, nous considérons et employons les termes « éducation à la santé » et « éducation pour la santé » tels que définit par Eymard à savoir la mise en œuvre d'une démarche d'éducation visant à faire acquérir un comportement du sujet ou du groupe conforme aux normes assurant une bonne santé.

Pour ce qui concerne les buts de l'éducation à la santé, comme pour l'ensemble des « éducations à… », Lange et Victor (2006) reviennent sur la typologie proposée par Robottom et Hart en 1993 (Fortin-Debart et Girault, 2004) dans le cadre des recherches en éducation relatives à l'environnement (ERE). Ils distinguent trois orientations différentes de la recherche en éducation sur ce domaine :

- l'approche interprétative, centrée sur les relations, les rapports entre l'apprenant et l'environnement ;
- l'approche positiviste qui vise surtout l'acquisition de connaissances avec en perspective un changement de comportement ;
- l'approche critique sociale, qui se donne pour objectif de permettre à l'apprenant de développer une analyse critique des valeurs et intérêt sous-jacents (p.88).

Si Lange et Victor (op.cit.) inscrivent leurs travaux de recherche dans le courant de l'approche « critique sociale » car, selon eux, cette dernière permet de garder un équilibre indispensable entre « éducation » et « savoirs » mis au service d'une visée sociale, ils pensent aussi que l'approche éducative, tout en étant centrée sur les savoirs, est aussi rattachée à des valeurs, des objectifs plus étendus qui sont ceux de l'idéal démocratique. Et, l'idée de science engagée permet de situer les savoirs dans cette perspective avec comme visée : la construction d'opinions chez les apprenants.

Le but de l'éducation à la santé devrait donc être à la fois de faire acquérir des connaissances avec en perspective un changement de comportement, mais également de permettre à l'apprenant de développer une analyse critique des valeurs et intérêts sous-jacents. Car, comme l'affirme Pritchard et Buckland (1986), *« la science n'est pas neutre, et dans une très large mesure, ne saurait être enseignée sans référence à des valeurs* ».

1.3.3 Education à la santé : acteurs

Qui peut intervenir dans le cadre de l'éducation à la santé ? Pour Lemmonier et al. (op.cit.), les intervenants et les publics – que nous pouvons appeler « apprenants »- sont divers et variés. Pour eux, est intervenant, « *toute personne en posture d'éducateur dans une action ou intervention d'éducation pour la santé* », (p.19). Il peut donc s'agir : « *d'un enseignant qui mène une action d'éducation nutritionnelle auprès de ses élèves ; d'un animateur sportif qui propose une action de prévention du dopage aux licenciés d'un club de sport ; d'une infirmière scolaire qui propose des débats ou jeux de rôle sur le thème des relations affectives et sexuelles dans le cadre d'un projet d'établissement s'étant fixé pour objectif la promotion de la santé auprès des jeunes ; d'un médecin généraliste qui entreprend une démarche éducative auprès des personnes âgées ; d'un pharmacien qui propose une réflexion sur la consommation d'alcool ou de tabac ; d'une diététicienne qui met en place des activités d'éducation du patient…* » (Lemmonier et al. Op.cit., p.19).

1.3.4 Education à la santé et formation des enseignants

Étant donné que notre échantillon comporte des enseignants en formation, il nous a paru judicieux de considérer les travaux qui traitent de l'éducation à la santé dans la formation des enseignants. Nous nous attarderons sur les obstacles que rencontre cet objet d'enseignement à partir du travail réalisé par Cardot (2011) dans le cadre de sa thèse de doctorat qui portait sur les formateurs d'enseignants et l'éducation à la santé en France. Cardot (op.cit.) identifie deux principaux obstacles qui sont : la place de l'éducation à la santé dans la formation des enseignants et la place prépondérante des disciplines. Concernant la place qui lui est accordée, l'éducation à la santé est noyée dans de multiples priorités et elle occupe une place périphérique dans l'identité professionnelle des enseignants (Jourdan, 2004). Cela est dû au fait que les enseignants et les formateurs ont du mal à appréhender cet objet d'enseignement dans toutes ses dimensions. Certains sous-estiment la dimension globale de l'éducation à la santé la réduisant à un apport de connaissances dévolu à certaines disciplines (SVT, EPS, etc.) et occultant la dimension éducative, ils considèrent qu'elle relève du cadre familial, la priorité étant donné à leur discipline respective pour les enseignants du second degré et au français et aux mathématiques pour ceux du premier degré. D'autres enseignants trouvent le sujet trop complexe et ne se sentent pas compétents pour l'aborder. Ils préfèrent donc qu'il soit confié à des experts. Pour ce qui est de la place prépondérante des disciplines, Cardot (op.cit.) montre que pour l'éducation à la santé, il est nécessaire de s'écarter de l'enseignement disciplinaire, traditionnel, basé sur l'apport de connaissances et en référence aux savoirs, pour adopter une démarche interdisciplinaire fondée sur un projet éducatif qui s'appuie sur un ensemble de valeurs. Il ne s'agit pas,

pour les enseignants de s'appuyer sur un savoir figé dans une logique médicale mais au contraire d'agir dans un contexte d'incertitude par rapport à un objet « mouvant » qui résiste à toute transmission descendante. L'éducation à la santé, comme l'éducation à l'environnement, est une forme scolaire non disciplinaire (Martinand, 2003). Il s'agit de permettre aux différents acteurs de travailler ensemble et de s'impliquer dans des actions où chacun a un rôle à jouer, en instaurant un partenariat réel entre les différents acteurs potentiels. Pour cela, il convient de développer la formation des enseignants en élaborant une culture commune afin d'homogénéiser les concepts de santé et d'éducation à la santé.

2 Approche méthodologique

2.1 . Méthode

Dans le cadre de cette étude, nous nous sommes servie du questionnaire élaboré dans le cadre de ce projet pour pouvoir comparer les résultats obtenus au Gabon avec ceux des autres pays ayant pris part à ce dernier. Le questionnaire a été rempli de manière anonyme par les enquêtés en présence des chercheurs. Le questionnaire comprend les différents thèmes suivants : l'éducation à la santé ; la reproduction humaine et l'éducation sexuelle ; l'écologie et l'éducation à l'environnement ; le cerveau humain ; la génétique humaine et l'évolution. Dans le questionnaire, les questions sont organisées en deux sections principales notées A et B. La section A concerne les questions relatives à l'ensemble des thèmes et des compléments sur l'Environnement ; tandis que dans la section B, il est question des compléments sur la santé et de questions biologiques. Les réponses aux questions ont été recueillies et codées suivant une échelle quantitative, en général de 1 à 4 sur une gradation entre « d'accord » et « pas d'accord » (échelle de Likert), ou de 1 à 5 lorsqu'il s'agit d'un différenciateur sémantique d'Osgood entre deux adjectifs antonymes (Munoz et Clément, 2007). Dans cet article, nous présenterons les résultats de certaines questions relatives à l'éducation (Tableau 1). Elles sont toutes codées de 1 à 4 sur une gradation entre « d'accord » et « pas d'accord », sauf la question A63 qui porte sur la conception que les enquêtés ont de la santé. Elle est formulée comme suit :

La santé peut être perçue de différentes manières. Dans la liste suivante, cocher les trois expressions qui sont les plus associées à votre perception de la santé :

1. Ne pas souffrir de maladies graves ;
2. Etre en paix avec moi-même ;
3. Profiter de la vie, sans trop de stress ;
4. Avoir toutes les parties de mon corps qui fonctionnent bien ;
5. Etre en forme pour être socialement actif ;
6. Ne pas avoir besoin d'aller chez le docteur pour être soigné.

Les enquêtés devaient choisir trois expressions parmi ces propositions. Ce sont les réponses 2, 3 et 5 qui ont été retenues dans la codification. Pour chaque enquêté, nous avons noté le nombre de ces réponses (aucune, une, deux, les trois).

A55.	Selon vous, l'objectif principal de l'éducation à la santé dans le cadre scolaire devrait être :	Faire acquérir des connaissances	1	2	3	4	Faire acquérir des comportements respectueux de sa propre santé
B1.	L'éducation à la santé en milieu scolaire améliore le comportement des élèves.	D'accord					Pas d'accord
B15.	Il revient principalement aux infirmières et aux médecins scolaires d'assurer l'éducation à la santé.	D'accord					Pas d'accord
B21.	Dans le cadre scolaire, l'éducation à la santé doit se limiter à des informations scientifiques (alimentation, sommeil, drogues…).	D'accord					Pas d'accord
B22.	Les enseignants ne devraient pas être obligés d'enseigner l'éducation à la santé s'ils ne se sentent pas à l'aise pour le faire.	D'accord					Pas d'accord
B23.	Les établissements scolaires doivent prendre en compte les politiques de santé publique.	D'accord					Pas d'accord
B26.	L'éducation à la santé consiste principalement à développer les aptitudes personnelles des élèves, par exemple l'estime de soi ou la gestion du stress.	D'accord					Pas d'accord
B27.	L'éducation à la santé relève uniquement de la responsabilité des familles.	D'accord					Pas d'accord

Tableau 1 - Les questions relatives à l'éducation à la santé

Les questions que nous avons sélectionnées pour cet article peuvent être regroupées en 3 catégories :

- **Catégorie 1**- Perception de la santé (A63) ;
- **Catégorie 2**- Les buts de l'éducation à la santé et relation avec les institutions (B1, B21, B23)
- **Catégorie 3**- Les acteurs de l'éducation à la santé (B15, B22, B23, B27)

2.2 Echantillonnage

L'échantillonnage a été conçu de manière à représenter de façon équilibrée six catégories d'enseignants, suivant deux classes de statuts (enseignants en service, *In* vs. Futur enseignant, *Pre*) et trois classes de disciplines (Primaire, *P*, Secondaire Lettres, *L* et secondaire biologie, *B*). Le but étant, entre autre, de voir l'impact des connaissances des enseignants (celles acquises au cours de leur formation et en dehors) sur leurs représentations, les enseignants de Lettres et ceux du Primaire servent, un peu de témoins par rapport aux enseignants de biologie. Le tableau 2 indique les effectifs des individus échantillonnés par groupe d'enseignants.

Groupes d'enseignants	Code	Effectifs
Enseignants de SVT en formation	PreB	52
Enseignants de Français en formation	PreL	52
Enseignants sur primaire en formation	PreP	53
Enseignants de SVT en service	InB	31
Enseignants de Français en service	InL	35
Enseignants du Primaire en service	InP	46
	Total	269

Tableau 2- Composition de l'échantillon

2.3 Caractéristiques de l'échantillon

Nous avons choisi de mettre en relation les réponses des enseignants de notre échantillon avec leur niveau d'études (Tableau 3).

Niveaux d'études	PreB	PreL	PreP	InP	InB	InL	Total
Enseignement secondaire	-	-	4	5	-	-	9
Bac +1 ou 2	39	46	45	40	-	-	170
Bac +3 ou 4	11	5	2	1	12	9	40
Bac +5 ou 6	2	1	1	-	18	24	46
Plus élevé	-	-	1	-	1	2	4
Total	52	52	53	46	31	35	269

Tableau 3- Niveau d'études des enseignants interrogés

Parmi les enseignants interrogés, il apparaît que les enseignants du primaire sont ceux qui ont le niveau d'étude le plus faible. 9 d'entre eux ont un niveau d'études qui correspond à l'enseignement secondaire ; tandis que les niveaux les plus élevés se retrouvent chez les professeurs de biologie et de français en service. Pour ce qui concerne les enseignants du Primaire, autrefois, ils étaient effectivement recrutés avec un niveau de l'enseignement secondaire, avec ou sans le baccalauréat. Depuis quelques années, maintenant, le baccalauréat est exigé pour entrer en formation. S'agissant des enseignants de Biologie (SVT) et de français en formation, les modalités d'entrée à l'ENS de Libreville sont assez diversifiées. En effet, depuis sa création au début des années 70 jusqu'en 1996, on entrait à l'ENS avec le baccalauréat ; la sélection des étudiants se faisait sur dossier et on y passait soit 3 ans pour être professeur de collège, soit 5 ans pour être professeur de lycée. Il y avait évidemment des passerelles qui faisaient que des étudiants pouvaient entrer directement en 3ème, 4ème ou 5ème année selon qu'ils étaient titulaires d'un DEUG, d'une Licence ou d'une Maîtrise (Master 1). Après 1996, suite à la création des IUFM en France, l'ENS a adopté le même mode de recrutement que ces derniers. L'entrée se faisait désormais sur concours et ne pouvaient s'y présenter que les étudiants détenteurs d'une licence pour préparer le diplôme de professeur de collège ou ceux titulaires d'au mois une Maîtrise (Master 1) pour le diplôme de professeur de lycée. Cependant, le déficit d'enseignants dans le secondaire a fait que depuis deux ou trois ans, l'ENS a ouvert à nouveau des classes de premières années auxquelles on accède juste après le Baccalauréat, en plus des classes auxquelles on accède par le concours comme nous venons de l'expliquer. Ce qui explique les différents niveaux d'études qui apparaissent dans notre échantillon. Pour parler du

rapport à la religion des enseignants de notre échantillon, sur 269 enseignants interrogés, une large majorité d'entre eux (250) se dit croyants, plus de la moitié d'entre eux sont chrétiens, majoritairement catholique ou protestant. Et, parmi ceux qui se disent croyants, plus de la moitié pratique leur religion. Nous avons voulu voir si la dimension du sexe intervenait aussi dans la construction des représentations de ces enseignants sur le déterminisme génétique du phénotype. L'échantillon compte 61% de femmes et 39% d'hommes.

2.4 Analyse des données

Pour traiter les données recueillies, nous avons choisi l'Analyse Statistique Implicative (ASI) réalisée à l'aide du logiciel CHIC. Initiée et développée par Gras (2009), l'ASI vise l'extraction de connaissances, d'invariants, des règles inductives non symétriques consistantes, et accorde une mesure à des propositions du type « quand a est choisi, on a tendance à choisir b » Pour cela, elle quantifie la qualité de ces règles sur la base statistique d'un nombre significatif de contre-exemples où la règle n'est pas vérifiée et où certains déséquilibres cardinaux sont observés parmi les exemples et les contre-exemples à l'implication et à sa contraposée[2] (Gras 2013). En d'autres termes, la modélisation de la quasi-implication mesure l'étonnement de constater la petitesse des contre-exemples en regard du nombre surprenant des instances de l'implication (Couturier et Gras, 2005) (Couturier, 2008). L'A.S.I. permet d'obtenir des graphes implicatifs, des arbres hiérarchiques de similarités et une hiérarchie cohésitive. Dans cet article, nous présentons l'analyse du graphe implicatif aux seuils[3] d'implication de 0,99, 0.95, 0.90 et 0.85, il va donc s'agir de quasi-implication et l'analyse cohésitive orientée. L'analyse de l'arbre cohésitif traduit ce que Kuntz (2005) a appelé « la hiérarchie orientée ». Les niveaux de la hiérarchie orientée sont des règles ou des R-règles et contrairement au modèle hiérarchique classique, une hiérarchie orientée ne contient que des règles significatives selon un critère statistique préalablement défini. En fait, ces règles ou R-règles font référence à des variables se structurant de façon ascendante en classes emboîtées et orientées. Or, l'existence d'une certaine **cohésion** entre les variables qui constituent une classe est indispensable pour qu'une règle ait véritablement son sens à l'intérieur de cette classe dont on examine la relation avec d'autres.

> « *Comme en classification hiérarchique classique* (hiérarchique de similarités), *étant donné la multiplicité des niveaux de la hiérarchie orientée, il est*

[2] Si a → b est une implication, sa contraposée est non b→ non a. Cette implication est, a priori, équivalente à la première. Ce qui veut dire que les deux sont ou vraies ou fausses en même temps. Par ex. dans une pièce de la maison où l'électricité fonctionne bien, si l'on appuie sur le bouton (a), alors la lampe s'allume (b). Cette implication a → b est équivalente à : si la lumière ne s'allume pas (non b) alors c'est que l'on n'a pas appuyé sur le bouton (non a).soit non b→ non a.

[3] Pour mesurer la qualité de la règle d'implication, Gras définit une intensité d'implication. Celle-ci repose sur la probabilité que, si seul le hasard et l'indépendance a priori des variables intervenaient, la probabilité que le nombre de contre-exemples observés dans le cas du hasard (ensembles de représentation X et Y de a et b aléatoires) serait plus grand que celui qui est observé dans la contingence (ensembles A et B de représentation de a et b observés), compte tenu des effectifs de A, B et E. Par exemple, si l'on fixe 0,95 comme seuil d'acceptabilité de l'implication a->b et si la probabilité mesurée est 0,98, on accepte l'implication. Si elle est égale à 0,89, on peut réfuter l'implication, voire l'accepter au seuil 0,88

nécessaire de dégager ceux qui sont les plus pertinents par rapport à l'intention classificatrice de l'utilisateur et eue égard aux critères de construction choisis » (Kuntz, 2005, p. 59).

- *Notion de niveau significatif*

Etant donné la multiplicité des niveaux de formation des classes, il est indispensable de dégager ceux qui sont les plus pertinents par rapport à l'intention classificatrice du chercheur et eu égard aux critères choisis. Ces niveaux semblent, dans des applications psycho- didactiques ou sociologiques, correspondre à des conceptions consistantes et stables d'où leur intérêt pour l'expert (Gras, op.cit.). Lorsqu'on a un niveau significatif, cela signifie que les modalités ou regroupements de modalités qu'il contient ont une cohérence très forte. Ainsi, les niveaux non significatifs ne peuvent valablement intervenir dans l'interprétation (Comiti, op. cit.). Pour savoir si un niveau est significatif, on forme deux parties disjointes dans l'ensemble des paires de modalités : d'une part les paires de modalités séparées par ce niveau, et d'autre part, les paires de modalités qu'il réunit. Le niveau est significatif si le plus fort indice de similarité des paires de la première partie est inférieur au plus faible indice de similarité des paires de la deuxième partie. Le logiciel C.H.I.C. dont nous nous sommes servie, fait automatiquement ces calculs et affiche en rouge les niveaux significatifs.Dans notre étude, nous souhaitons voir les relations d'implication entre les variables. En d'autres termes, nous cherchons à identifier des règles du type : « si l'enseignant donne la réponse a, alors généralement il donnera la réponse b ». L'A.S.I. permet de dégager, au sein de groupes bien identifiés, des structures, des liaisons entre les variables. Un autre avantage de l'A.S.I. est que la taille de l'échantillon la taille n'est pas un handicap car, sur la cohésion des éléments structurants, telle que l'ASI et CHIC la mettent en évidence, le nombre des sujets semble importer peu dans la mesure où les réseaux résistent à des perturbations numériques. Compte tenu de la taille imposée à cet article, les représentations des enseignants seront analysées uniquement au regard des connaissances qu'ils déclarent avoir (voir tableau 2) de leur niveau d'études et leurs valeurs (interaction KV).

2.5 Les limites de notre enquête

La méthodologie adoptée peut présenter des limites en ce sens qu'il s'agit de connaissances déclaratives qui semblent ne pas avoir été vérifiées par ailleurs. Or, et c'est là aussi que les relations d'implication entre les variables sont importantes parce qu'elles vont faire ressortir les cohérences et les incohérences entre les réponses déclaratives des enseignants, d'une part. D'autre part, le questionnaire utilisé n'a pas été testé sur des enseignants gabonais. Il se peut donc que cela introduise des biais dans la compréhension des questions et donc dans les réponses fournies.

3 Résultats

3.1 Les graphes implicatifs

Au seuil d'implication de 0,99, nous obtenons le graphe ci-dessous qui compte un seul chemin implicatif :

Figure 2- Graphe implicatif 1 au seuil d'implication de 0.99

Ce chemin implicatif montre que les enseignants qui ne sont pas d'accord avec le fait que l'éducation à la santé devrait revenir principalement aux infirmières et aux médecins scolaires, ne sont pas aussi d'accord avec le fait que cette éducation devrait relever uniquement de la responsabilité des familles. Ce sont les professeurs de français en service qui contribuent le plus à la formation de ce chemin, suivis des professeurs de biologie en service. En baissant le seuil d'implication à 0.95, 0.90 et 0.85, on obtient le graphe suivant :

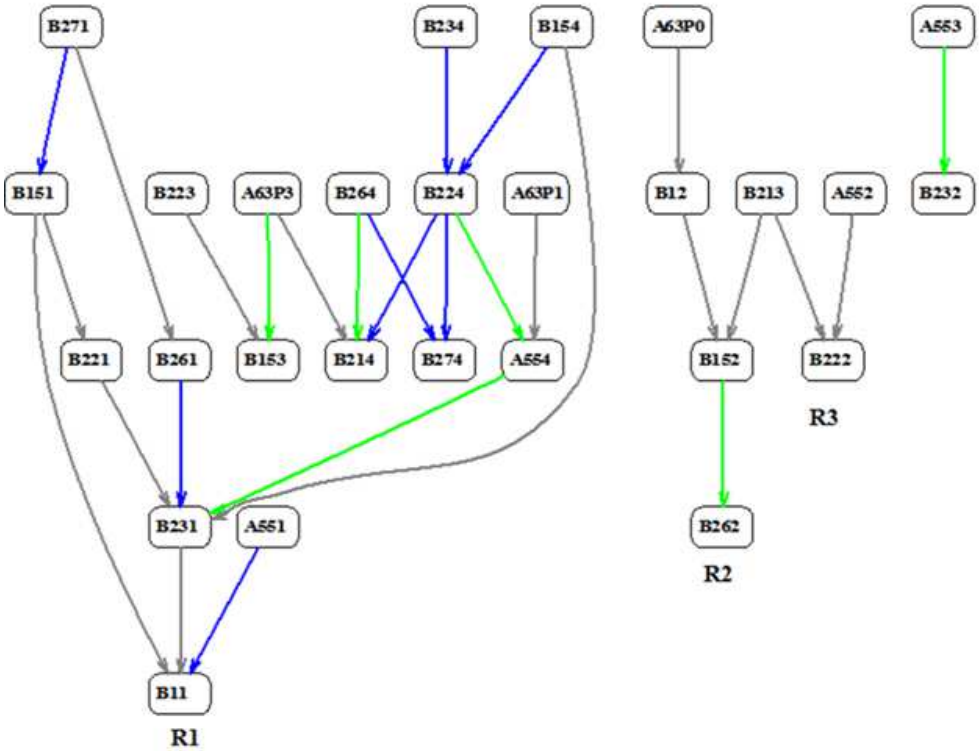


Figure 3- Graphe implicatif au seuil d'implication de 0.95, 0.90 et 0.85.

99 95 90 85

Compte tenu du nombre de pages limitées de cet article, nous n'analyserons que les réseaux qui nous semblent les plus pertinents.

Dans le réseau R1, le chemin B271 →B151 →B221 →B231 →B11 montre qu'il y a des enseignants qui sont d'accord avec le fait que l'éducation à la santé relève uniquement de la responsabilité des familles (B271) et que ces derniers sont aussi d'accord avec le fait que cette éducation revient principalement aux infirmières et médecins scolaires. Ils sont d'accord avec le fait que les enseignants ne devraient pas être obligés d'enseigner l'éducation à la santé s'ils ne se sentent pas à l'aise de le faire (B221), que les établissements scolaires doivent prendre en compte les politique de santé publique (B231). Ils sont aussi d'accord avec le fait que l'éducation en milieu scolaire améliore le comportement des élèves (B11). Ces enseignants sont d'accord avec le fait que l'éducation à la santé consiste principalement à développer les aptitudes personnelles des élèves, par exemple l'estime de soi ou la gestion de stress (B261). Ce sont les enseignants du primaire en formation qui contribuent le plus à ce chemin implicatif suivi des enseignants de français en formation et des professeurs de Biologie en service.

Toujours dans le réseau R1, le chemin A551 →B11 montre que les enseignants du primaire en formation pensent que l'objectif principal de l'éducation à la santé devrait être de faire acquérir des connaissances et qu'elle améliore le comportement des élèves.

D'après le chemin A63P1 →A554 →B231→B11, les enseignants qui considèrent que la santé est le fait d'être en paix avec soi-même, considèrent aussi que l'éducation à la santé dans le cadre scolaire devrait avoir pour objectif principal de faire acquérir des comportements respectueux de sa propre santé, que les établissements scolaires devraient suivre les politiques de santé publique et que l'éducation à la santé en milieu scolaire améliore le comportement des élèves. La variable qui contribue le plus à la formation de ce chemin implicatif est le fait d'être plutôt non croyant. Les enseignants qui ont donné ces réponses sont plutôt non croyants.

Il y a aussi des enseignants pour lesquels l'éducation à la santé ne revient pas principalement aux infirmières et médecins scolaires, les enseignants devraient être obligés d'enseigner l'éducation à la santé même s'ils ne se sentent pas l'aise. Pour eux, l'éducation à la santé ne relève pas uniquement de la responsabilité des familles et son contenu ne devrait pas se limiter à des informations scientifiques (alimentation, sommeil, drogue…). Ces enseignants pensent aussi que l'objectif principal de l'éducation à la santé devrait être de faire acquérir des comportements respectueux de sa propre santé. C'est ce que montre les chemins B154 →B22→B274, B224→B214 et B224→A554. Le chemin B234→B224→B274 montre que ces enseignants ne sont pas d'accord avec le fait que les établissements scolaires doivent suivre les politiques de santé publique. Les enseignants de biologie et les enseignants du primaire en service sont ceux qui contribuent le plus à la formation de ces chemins implicatifs.

Les réseaux R2 et R3 montrent des positions intermédiaires entre d'accord et pas d'accord pour les questions que nous avons posées.

3.2 L'analyse cohésitive orientée

L'analyse de l'arbre cohésitif (figure 4) confirme les résultats obtenus par l'analyse des graphes implicatifs. On voit des règles s'établir entre certaines variables. Pour plus d'efficacité, nous n'analyserons que les niveaux significatifs.

Ainsi, le niveau le plus significatif s'établit entre les variables B154 et B274. Le graphe implicatif a lui aussi montré une quasi implication entre ces deux variables. La variable B154 correspond aux enseignants qui ne sont pas d'accord avec le fait que l'éducation à la santé devrait revenir principalement aux infirmières et aux médecins scolaires tandis que la variable B274 désigne ceux qui ne sont pas d'accord avec le fait que cette éducation devrait relever uniquement de la responsabilité des familles. On a une règle entre ces deux variables.

Le second niveau significatif s'établit entre les deux précédentes variables et B224 et B214. Ces deux dernières variables parlent respectivement des enseignants qui ne sont pas d'accord avec le fait qu'ils ne devraient pas être obligés d'enseigner l'éducation à la santé s'ils ne sont pas à l'aise pour le faire (B224) et ceux qui ne sont pas d'accord avec le fait que dans le cadre scolaire, l'éducation à la santé doit se limiter à des informations scientifiques. Il s'établit des règles entre ces variables. Donc si on résume, l'éducation à la santé ne doit pas revenir principalement aux infirmières et aux médecins scolaires, elle ne doit pas relever uniquement de la responsabilité des familles ; et dans le même temps, les enseignants ne devraient pas être obligés de l'enseigner s'ils ne sont pas à l'aise pour le faire et son contenu, dans le cadre scolaire, ne doit pas se limiter au contenu scientifique. A qui revient la charge de l'éducation à la santé ? Les intervenants en charge de cette éducation ne sont pas clairement désignés, son contenu également.

Le troisième niveau significatif établit une règle entre les variables B213, B152 et B262. On voit ici que les enseignants qui ne sont plutôt pas d'accord avec le fait que les établissements scolaires doivent prendre en compte les politiques de santé publique sont plutôt d'accord avec le fait qu'en milieu scolaire devrait être assuré par les médecins et les infirmières. Ils sont aussi plutôt d'accord avec le fait que l'éducation à la santé consiste principalement à développer les aptitudes personnelles des élèves telles que l'estime de soi ou la gestion du stress.

Dans le quatrième niveau significatif, les enseignants qui ne sont pas d'accord avec le but de l'éducation à la santé qui consiste à développer principalement les aptitudes personnelles des élèves, ne sont pas aussi d'accord avec le fait que cette éducation devrait être assurer par des médecins et des infirmières en milieu scolaire. Ils ne sont pas d'accord avec le fait qu'elle relève uniquement de la responsabilité des familles et qu'elle doit se limiter à des informations scientifiques.

L'analyse de l'arbre cohésitif orienté montre des positions tranchées au niveau des enseignants de notre échantillon. Les niveaux les plus significatifs apparaissent entre les enseignants qui ne sont pas d'accord avec les propositions relatives aux responsables de la formation et à son contenu, d'une part et ceux qui sont plutôt d'accord avec ces propositions. Aucun niveau significatif ne s'établit entre les enseignants qui sont d'accord avec les propositions qui leur ont été faites.

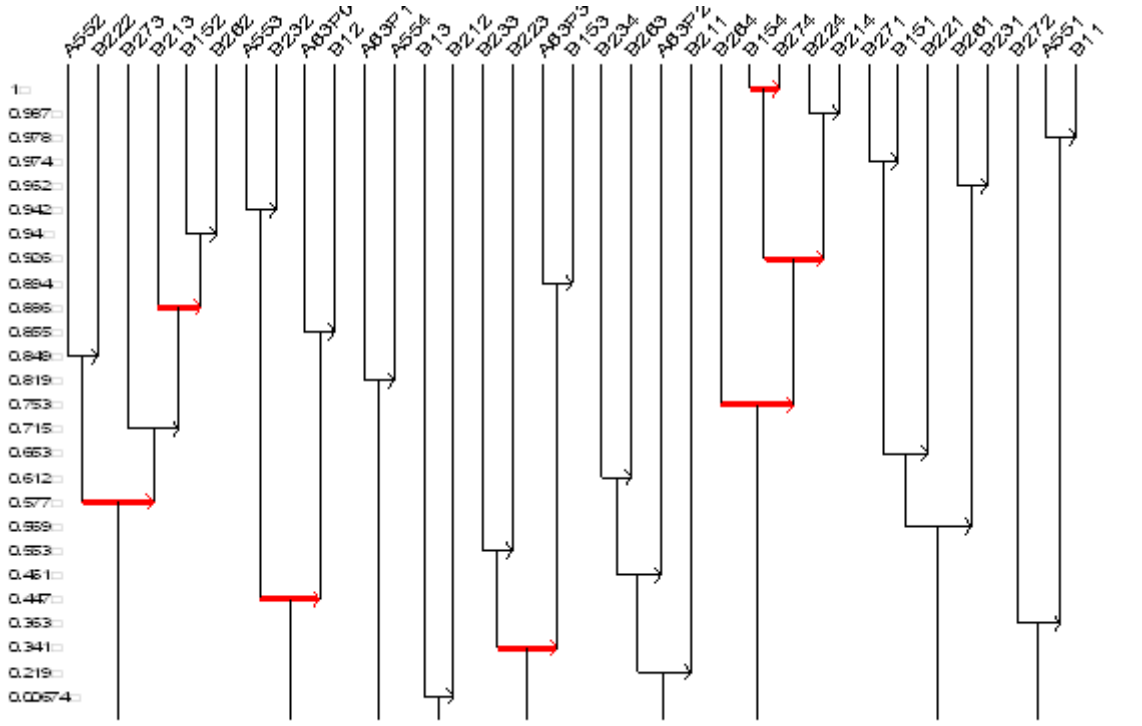


Figure 4- L'arbre cohésitif

4 Discussion

En considérant le seuil d'implication le plus élevé entre les variables, il est apparu que les professeurs de français en service pensent que l'éducation à la santé ne revient pas principalement aux infirmières et aux médecins scolaires et qu'elle ne relève pas uniquement de la responsabilité des familles. On aurait pu attendre ce point de vue des professeurs de biologie en service ou en formation. Se sentiraient-ils pas assez outillés pour assurer cette éducation ? On retrouve ce positionnement chez les enseignants en formation de notre échantillon. Le problème qui semble se poser est celui de la place de l'éducation à la santé dans la formation de ces enseignants, de son contenu et de ceux qui devraient l'assurer. On retrouve là le constat que fait Cardot (op.cit) lorsqu'il dit que pour parmi les obstacles à l'éducation à la santé, il y a le fait que les enseignants et les formateurs d'enseignants ont du mal à appréhender cet objet d'enseignement dans toutes ses dimensions. Ils réduisent alors l'éducation à la santé à certaines disciplines (SVT, EPS, etc.) et en occultant la dimension éducative, ils considèrent alors qu'elle relève du cadre familial. Il relève aussi le fait que certains enseignants se sentent incompétents et par conséquent, ils confient cette éducation à des experts (infirmiers, médecins scolaires). Il propose donc de travailler en formation sur les conceptions qu'ont les enseignants de leurs missions par rapport aux questions de santé et d'enraciner l'éducation à la santé au cœur de leur identité professionnelle. On retrouve l'approche positiviste lorsque des enseignants de notre échantillon pensent que l'éducation à la santé doit avoir comme objectif principal l'acquisition des connaissances. L'éducation à la santé doit transmettre des connaissances, mais sans viser l'acquisition de comportements respectueux de leur propre santé, les

connaissances seules devraient suffire à amener les élèves à changer de comportement. Or, il a été démontré que les connaissances à elles seules ne suffisent pas à améliorer les comportements. C'est pourquoi Rugumayo (1985) propose que pour réussir, les professeurs de sciences des pays en développement doivent rendre leur enseignement, utile et pratique, propre à résoudre des problèmes, tout en formant des individus qui aient assez d'imagination pour adapter leurs acquis à des situations nouvelles, les compétences requises pour mener les choses à bien, la curiosité nécessaire pour découvrir et comprendre le monde qui les entoure et aussi la sensibilité voulue pour mettre l'enseignement des sciences au service des besoins de l'homme. Nous pouvons appliquer cela à l'éducation à la santé. Il serait donc souhaitable d'aller au-delà de la simple transmission des connaissances. Et ce, d'autant plus que pour l'Organisation Mondiale de la Santé, l'éducation à la santé se situe dans une perspective d'apprentissage de la liberté et du « vivre ensemble » plutôt que d'intégration de préceptes normatifs (Berger, 2015[4]). Les enseignants de notre échantillon qui pensent que la santé est le fait d'être en paix avec soi-même, pensent aussi que l'éducation à la santé doit viser l'acquisition de comportements respectueux de sa propre santé, et qu'elle devrait suivre les politiques de santé publique du pays. Il y a là une certaine reconnaissance de la nécessité de faire en sorte que l'éducation à la santé soit en phase avec les problèmes posés par les politiques de santé publique. Les professeurs de biologie et les enseignants du primaire en service pensent que l'éducation à la santé ne revient pas principalement aux infirmières et médecins scolaires, ni d'ailleurs aux seules familles et que les enseignants devraient être obligés de l'enseigner même s'ils ne se sentent pas à l'aise. Est-ce pour autant que l'on peut affirmer qu'ils perçoivent toutes les dimensions de l'éducation à la santé ? A ce stade de notre enquête nous ne pouvons l'affirmer, mais nous constatons que pour eux, les enseignants sont aussi des acteurs de l'éducation à la santé en milieu scolaire. Par contre, ils affichent un certain détachement face à l'institution publique puisque pour eux, les établissements scolaires ne doivent pas suivre les politiques de santé de publique. L'acte éducatif s'inscrivant dans un contexte socio-économique et culturel, comment l'éducation à la santé pourrait-elle se détacher des politiques de santé publique du pays ?

Les positions intermédiaires (entre d'accord et pas d'accord) témoignent pour nous du manque de conviction de certains de nos enquêtés par rapport à cet objet d'enseignement qu'est l'éducation à la santé. Cette étude a été menée dans le cadre de l'enquête européenne « BIOHEAD CITIZEN » qui a été réalisée dans plusieurs pays. Les résultats pour 15 d'entre-eux (Algérie, Burkina-Faso, Brésil, Chypre, Estonie, Finlande, France, Hongrie, Italie, Liban, Maroc, Portugal, Roumanie, Sénégal et Tunisie) ont montré que les enseignants en formation ont une conception qui se rapproche davantage des conceptions les plus ouvertes de l'éducation à la santé contrairement aux enseignants en service qui paraissent plus réticents à aborder ces sujets. Ils ont tendance à penser qu'il revient aux familles de prendre en charge ces questions (Berger, op.cit.). Ces conceptions rejoignent celles que nous avons identifiées au cours de notre enquête. On note un fort clivage entre les représentations des enseignants en service et ceux en formation. Dans l'étude de Berger (op.cit) La discipline enseignée a aussi un impact sur ces représentations puisque les professeurs de biologie sont les plus proches des acceptions actuelles de l'éducation à la santé. Nous

[4] Il s'agit d'un document numérique. La référence de l'article n'y figurant pas, nous indiquons ici l'année de consultation.

avons fait le même constat. Dans notre étude, les enseignants de biologie pensent aussi que l'éducation à la santé ne relève pas uniquement de la responsabilité des familles, qu'elle ne revient pas principalement aux médecins et aux infirmières en milieu scolaire et que le contenu de cette éducation ne doit pas se limiter à des informations scientifiques, mais qu'elle doit contribuer à faire acquérir des comportements respectueux de sa propre santé. Toutefois, ces positions se retrouvent aussi chez les enseignants du primaire en service. Par contre, les enseignants du primaire en formation ont des positions contraires. Pour eux, l'éducation à la santé devrait relever uniquement de la responsabilité des familles, qu'elle devrait revenir aux infirmières et aux médecins en milieu scolaire. Ils sont d'accord avec le fait que les enseignants ne devraient pas être obligés d'enseigner l'éducation à la santé s'ils ne sont pas à l'aise de le faire. Ces représentations peuvent être mises en lien avec leur niveau de formation car comme Berger (op.cit.) l'a aussi montré les enseignants dont le niveau de formation académique est le plus faible ont souvent des conceptions qui s'éloignent des acceptions les plus actuelles de l'éducation à la santé. Nous avons choisi d'utiliser le modèle KVP qui dit que les représentations d'un enseignant sont de possibles interactions entre ses connaissances, ses pratiques professionnelles et ses valeurs, ces dernières pouvant être influencées par le fait d'être croyant ou non, ou de pratiquer ou pas sa religion. Ainsi, il ressort de notre enquête que les enseignants non-croyants sont ceux pour lesquels la santé est le fait d'être en paix avec soi-même.

5 Conclusion

Nos questions étaient regroupées en trois catégories : la perception de la santé, les buts de l'éducation à la santé et relations avec les institutions et les acteurs de l'éducation à la santé. Il apparaît d'après nos résultats que la perception qui de la santé qui ressort le plus est le fait que celle est le fait d'être en paix avec soi-même. Pour ce qui concerne les buts de l'éducation à la santé, la représentation qui domine est celle pour laquelle l'objectif principal de l'éducation à la santé c'est de faire acquérir des connaissances et que cela devrait entrainer des changements de comportement chez les élèves. On est donc clairement dans une visée positiviste de l'éducation à la santé. S'agissant des relations avec l'institution, on retrouve deux positions assez tranchées. Il y a ceux qui pensent que les établissements scolaires doivent suivre les politiques de santé publique et d'autres qui pensent que non. Ce qui nous a surpris c'est que ce sont les enseignants du primaire et les professeurs de biologie en service qui pensent que les établissements ne doivent pas suivre les politiques de santé publique. Nous pensons que cela traduit un certain détachement vis-à-vis de l'institution comme si l'école devait définir elle-même ses contenus d'enseignement sans tenir compte de la politique du pays. Concernant les acteurs de l'éducation à la santé, les positions sont là aussi bien tranchées entre ceux qui pensent qu'il faut la laisser aux experts (infirmiers et médecins) et ceux qui pensent que les enseignants devraient aussi assurer cet enseignement. Nous retenons, en guise de conclusion, que pour donner sa pleine mesure à l'éducation à la santé, il semble primordial de redéfinir sa place dans la formation des enseignants et de l'enraciner au cœur de leur identité professionnelle pour qu'elle cesse d'être une activité périphérique.

Références

[1] ASTOLFI, J.-P., DAROT, L., GINSBURGER-VOGEL, Y., et TOUSSAINT, J. (1997). *Mots-clés en Didactique des sciences : Repères, définitions, bibliographies.* Paris – Bruxelles : De Boeck, coll. Pratiques pédagogiques.

[2] BEITONE A., LEGARDEZ, A. et al. (1997). *Travaux en didactique des sciences économiques, sociales et de gestion,* Publication de l'Université de Provence.

[3] BERGER, D. Les Représentations des enseignants en éducation à la santé et à la sexualité. http://espe.univ-reunion.fr/fileadmin/Fichiers/ESPE/bibliotheque/expression/36/Berger.pdf, consulté le 05 septembre 2015.

[4] CARDOT, J.-P., 2011. *Formateurs d'enseignants et éducation à la santé : analyse des représentations et identité professionnelle.* Education. Université Blaise Pascal - Clermont- Ferrand II; Clermont Ferrand 2, 2011. https://hal.archives-ouvertes.fr/tel-00751865/document, consulté le 20 mars 2015.

[5] CARVALHO, G., CLEMENT, P., BOGNER, F., CARAVITA, S. (2008). *BIOHEAD-Citizen : Biology, Health and Environmental Education for better Citizenship, Final Report*. Brussels : FP6, Priority 7, Project N° CITC-CT-2004-506015.

[6] CLEMENT, P. (1998). La biologie et sa didactique. Dix ans de recherches. *Aster*, 27, 57-93.

[7] CLEMENT, P. (2000). La Recherche en Didactique de la Biologie. *Actes du Colloque International de Didactique de la Biologie*. Alger, 11-28.

[8] CLEMENT, P. (2004). Science et idéologie : exemples en didactique et en épistémologie de la biologie. In Actes du colloque Science-Médias-Société. Lyon : ENS-LSH. (http://sciences-medias.ens-lsh.fr), consulté le 08 juillet 2013.

[9] CLEMENT P., (2010). Conceptions, représentations sociales et modèle KVP. *Skholê (Univ. de Provence, IUFM), 16*, 55-70.

[10] COUTURIER, R. (2008). Statistical implicative analysis. In *CHIC : Cohesive Hierarchical Implicative Classificatio*n, Volume 127 of Studies in Computational Intelligence, 41–52. Springer Verlag

[11] COUTURIER R, GRAS R. (2005), *CHIC : Traitement de données avec l'analyse implicative, Extraction et Gestion des Connaissances*, Volume I1, RNTI, Cépaduès, Toulouse, 679-684

[12] DORTIER, J.-F., 1999. Les représentations sociales. L'image de la psychanalyse. *Sciences Humaines*, (91), 44-46.

[13] EYEANG, E. (1997), *L'enseignement / Apprentissage de l'espagnol au Gabon. Eléments de recherche pour une adaptation socio-didactique*, Thèse de doctorat NR, Université Stendhal – Grenoble III, Tome 1.

[14] EYMARD, C. (2004). Essai de modélisation des liens entre éducation et santé, *Questions Vives*, (5)

http://www.ofep.inpes.fr/apports/pdf/Chantal%20Eymard.pdf, consulté le 20 mars 2015

[15] GRAS, R. (dir) réd. invités R. Gras, J.C. Régnier, C. Marinica, F. Guillet (2013). *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités,* Toulouse : Cépuadès Ed.

[16] JOURDAN, D., (2010). *Éducation à la santé. Quelle formation pour les enseignants* ? Saint-Denis : INPES, coll. Santé en action. http://www.inpes.sante.fr/CFESBases/catalogue/pdf/1272.pdf, consulté le 20 mars 2015.

[17] LANGE, J.-M. et VICTOR, P., 2006. Didactique curriculaire et « éducation à…la santé, l'environnement et au développement durable » : quelles questions, quels repères ? *Didaskalia* (28), 85-100.

[18] LEMMONNIER, F., BOTTERO, J., VINCENT, I., FERRON, C., (2005). *Référentiel de bonnes pratiques. Outils d'intervention en éducation pour la santé : critères de qualité*. Saint-Denis : INPES éditions, http://www.inpes.sante.fr/CFESBases/catalogue/pdf/883.pdf, consulté le 20 mars 2015.

[19] MUNOZ, F., CLEMENT, P. (2007). Des méthodes statistiques originales pour analyser les conceptions d'enseignants de plusieurs pays à partir d'un questionnaire sur des questions vives. *Actes Colloque AREF (Actualité de la Recherche en Education et en Formation),* Strasbourg.

[20] PRITCHARD, A.,J. et BUCKLAND, D., J., (1996). *Loisirs, valeurs et enseignement de la biologie. Paris* : UNESCO

[21] RUGUMAYO, E. (1985). Science Education and the Needs of Developing Countries. *In Science Education and Social Needs*. Paris: ICSU Press.

ENSEIGNEMENT DE LA RESOLUTION DE PROBLEMES MATHEMATIQUES A L'ECOLE PRIMAIRE AU MAROC : REPRESENTATIONS DES ENSEIGNANTS A L'EGARD DE LEURS PRATIQUES

Brahim EL MEKAOUI [1] et Fadhila FARHANE HORRIGUE[2]

MATHEMATIC PROBLEM SOLVING TEACHING IN MOROCCAN PRIMARY SCHOOL: TEACHERS' INTRODUCTIONS WHILE PRACTISING SYNOPSIS

RÉSUMÉ

La résolution de problèmes occupe une place importante dans la didactique des mathématiques parce qu'elle constitue le critère principal de la maîtrise des notions mathématiques et elle est le moyen d'en assurer une appropriation qui en garantit le sens. L'enseignant ne doit pas communiquer les savoirs qu'il veut enseigner, mais plutôt proposer des situations qui vont permettre à l'élève la construction d'une première solution en mobilisant ses connaissances antérieures. Et, pour trouver de telles solutions l'élève collabore avec ses pairs, s'appuie sur des ressources internes et externes… Ainsi, l'élève devient actif dans la construction de ses connaissances en mathématiques qu'il va transférer dans des situations nouvelles et complexes. Est-il le cas dans l'école primaire marocaine ? Les enseignants marocains croient-ils en la capacité de l'élève à résoudre des problèmes mathématiques de manière autonome ? Quelles sont leurs pratiques de préparer, présenter et corriger des problèmes mathématiques ?

Mots-clés : *résolution de problèmes, construction, situation, enseignant, élève, mathématiques.*

ABSTRACT

Problem solving plays an important role in the didactic of mathematics because it constitutes the principal criterion in the grasp of the mathematical notions. Also, it is the medium that assures an appropriacy which guarantees the meaning. The teacher should not communicate the notions he wants to teach but he should suggest some situations, which allow the student to construct a first solution by mobilising his /her prior knowledge. To find such situations, the student collaborates with his/her peers focusing on internal and external resources. So, the student becomes active in the construction of mathematical knowledge which he/she will activate in new and complex situations. Is it the case in the Moroccan primary school? Do Moroccan teachers believe in the student' ability to solve mathematical problems in an autonomous way? What are their practices in preparing and correcting mathematical problems?

Keywords : *problem solving, construction/building, situation, teacher, student, mathematics.*

1 Introduction

En tant que professeur d'enseignement primaire au Maroc depuis 2002, j'ai remarqué que les élèves rencontrent de grandes difficultés en mathématiques surtout

[1] Université de Lyon, ED 485 EPIC, UMR 5191 ICAR, brahimabosalah@yahoo.com
[2] Université de Lyon, ED 485 EPIC, UMR 5191 ICAR, fadhilamsh@yahoo.fr

dans la résolution de problèmes. Cette constatation est confirmée par des rapports officiels[3] et des résultats de tests internationaux[4] qui ont montré que les élèves marocains, à la fin de leurs cursus scolaires, n'arrivent pas à acquérir les compétences spécifiques et transversales requises pour l'insertion professionnelle. Or, les orientations de la politique éducative considèrent la réforme du système éducatif comme l'une de ses premières préoccupations et elles donnent la grande importance à l'amélioration de la qualité de formation. L'analyse profonde de cette situation montre que cette politique n'a pas réalisé son objectif. Car, malgré les efforts fournis, l'élève est considéré toujours en difficulté. Cette constatation est le point de départ de l'idée de cette recherche. Pourquoi nous continuons à qualifier l'élève « sans niveau » et qu'il ne possède pas les compétences spécifiques et transversales bien prescrites dans les curricula ? Si, au cours de plus d'un demi-siècle, le système scolaire marocain a connu différentes tentatives de réforme qui ont visé, à la fois, la quantité et la qualité du service éducatif, comment pouvons-nous expliquer les mauvais résultats obtenus par les élèves surtout dans les disciplines scientifiques ?

Certes, de telles questions sont très vastes, très problématiques et nécessitent l'étude de différentes composantes du système éducatif. Pour cela nous limitons notre champ de questionnements à la dimension didactique de ce problème. Autrement dit, notre recherche consiste à rechercher les éventuelles sources des difficultés rencontrées par les élèves dans l'apprentissage de concepts mathématiques. Et pour bien cerner ce sujet, nous mettons l'accent sur la question de l'enseignement - apprentissage de la résolution de problèmes mathématiques à l'école primaire. Ainsi, notre projet de recherche est à double facette : D'un côté, l'enseignant et ses pratiques d'enseignement, et de l'autre côté, l'élève avec ses processus d'apprentissage. Toutefois, dans le cadre de cette étude nous allons étudier seulement la première facette en examinant les pratiques d'enseignement dans la résolution de problème mathématiques à l'école primaire au Maroc. Le but de cette recherche est de faire des investigations sur les pratiques d'enseignement de la résolution de problèmes mathématiques dans l'école primaire au Maroc. Ainsi, nous visons à repérer et analyser les points de vue des enseignants relatifs à l'enseignement de la résolution de problèmes.

2 Aperçu historique sur le système éducatif marocain

Pour retracer le développement historique de l'institution scolaire dans la société marocaine, il convient de le lier avec les trois périodes qui marquent l'histoire contemporaine du Maroc, à savoir : avant le protectorat français, durant le protectorat et après l'indépendance. Toutefois, les changements[5] qu'a connus le Maroc, à partir de

[3] Rapport thématique 2009 sur les résultats du programme national d´évaluation des acquis PNEA 2008, fascicule des mathématiques

[4] On peut citer comme exemple l'article « Education : Enquête PISA de l'OCDE : Le Maroc ignoré » http://www.lopinion.ma/def.asp?codelangue=23&ref=2, consulté le 24/01/2015

[5] Ces changements sont marqués sur le plan politique par l'arrivé du Roi Mohamed VI et le gouvernement d'alternance, sur le plan économique par le lancement des projets de restructuration de l'économie marocaine, sur le plan juridique et de droit par une '' novelle conception du pouvoir'' et le renforcement de l'Etat du droit, et sur le plan d'enseignement et de formation par l'élaboration de la charte nationale de l'éducation et de la formation en 1999 et la mise en application progressive de ses dispositions dès les années 2000.

1999, exige une quatrième période qui sera consacrée à l'enseignement scolaire sous la charte nationale de l'éducation et de la formation.

2.1 L'enseignement au Maroc avant le protectorat français

L'histoire du Maroc montre que l'enseignement a toujours été une préoccupation prioritaire. En effet, avant l'instauration du protectorat français en 1912, le Maroc disposait déjà d'un réseau d'enseignement de différents niveaux avec un contenu essentiellement religieux. L'enseignement primaire était assigné aux jeunes enfants dès l'âge de cinq ans par l'école coranique. Cette dernière, appelée dans les villes MSID[6] et dans les compagnes « JAMA' », assurait une formation fondée sur la mémorisation des sourates (versets) du Coran et du Hadith (les paroles du prophète) et leurs interprétations. L'enseignant, appelé « FKIH » ou « TALEB » est supposé avoir appris le Coran par cœur. Il était considéré comme la source de savoir, le détenteur de celui-ci et il jouissait d'un pouvoir et d'un respect indiscutables. La méthode d'enseignement se caractérisait par l'apprentissage par cœur du texte saint au moyen de l'écriture et de la récitation ; les apprenants, appelés « MHADRA », mémorisent et restituent les savoirs à haute voix devant le maître (Boulahcen Ali, 2002, p. 12-13). Ouvert à une petite minorité d'enfants, l'enseignement secondaire se faisait dans des mosquées, des Zaouïas ou des médersas où un seul professeur enseignait toutes les disciplines. Cet enseignement est basé sur l'explication des versets coraniques, l'interprétation des paroles et actes du prophète Mohammed et sur la grammaire arabe (Boulahcen, 2002, p. 14-15). Concernant la pédagogie, c'est, toujours, la mémorisation et l'apprentissage "par cœur". Le fonctionnement de ces établissements scolaires était financé par les tribus qui payaient le salaire annuel du maître (ECHCHARTE), et les élèves venant de l'extérieur étaient pris en charge par les familles riches. Dès le IXe siècle, l'enseignement supérieur était garanti par université « Al-Qaraouyine » fondée à Fès en 859 par « Fatima ALFIHRIYA ». Il était consacré à l'étude du droit musulman (AL FIQH), la littérature, la grammaire et la rhétorique (EL Bayane), ainsi que les mathématiques, la philosophie et l'astronomie.

2.2 L'enseignement sous le protectorat français

Les autorités françaises ont mis en place un système éducatif caractérisé par une multiplicité et une diversité basées sur des différences ethniques (arabes, berbères, européennes), religieuses (musulmans, juifs, chrétiens), linguistiques (arabophones, amazighophones) et socioprofessionnelles (notables, artisans, paysans) (El Ayadi M, 2009, p. 38). L'objectif de cette politique visait la formation d'une élite intellectuelle avec laquelle ces autorités entendent coopérer et qui va être médiateur entre la société marocaine et les européens. A cette époque, différents types d'écoles coexistaient pour fournir un enseignement assez élitiste. Il s'agit des écoles de fils de notables, collèges musulmans, écoles libres dites « Madaris Horra », écoles professionnelles, lycées français et des écoles franco-israélites.

[6] MSID est une pièce situé près de la mosquée ou en fait partie.

2.3 L'enseignement au Maroc de l'indépendance (1956) jusqu'en 1999

Le Maroc indépendant a hérité des autorités occupantes une situation marquée par la dualité, voire la diversité, dans tous les domaines : juridique, administratif, juridictionnel et éducatif… Dans le domaine éducatif, l'Etat national marocain a adopté le modèle des écoles de Jules Ferry en développant l'école des fils de notables mises en place par le protectorat (El Ayadi M., 2009, p 38). En fait, l'objectif était la construction d'une école nationale marocaine ayant pour mission de la démocratisation de l'enseignement et la formation de cadres nationaux capables de répondre aux exigences du développement du pays. Ainsi, au lendemain de l'indépendance (1956) la politique éducative se basait sur les principes suivants : la démocratisation de l'enseignement, son unification, l'arabisation du contenu et la marocanisation des cadres.

2.4 L'état actuel du système scolaire au Maroc

A la fin des années quatre-vingt-dix les autorités marocaines avaient lancé une large concertation entre toutes les composantes de la société marocaine pour élaborer la charte d'une profonde réforme du système éducatif. Ces consultations ont abouti à l'adoption de la Charte Nationale d'Education et de Formation. A la lumière de cette charte, le Maroc a placé la réforme et le développement du système éducatif parmi les premières priorités de l'Etat, et la période 2000-2009 fut déclarée décennie nationale de l'éducation et de la formation. La scolarité dans l'école primaire a connu une grande augmentation au fil des années. En effet, le taux de scolarisation des enfants de 6 à 11 ans est passé de 54,1% en 1991-1992 à 92,2% en 2003-2004 pour atteindre 99,5% en 2013-2014. Concernant l'évolution du taux de la scolarité selon le genre, les statistiques des dernières années nous amènent à dire qu'il y a une égalité de chance entre les filles et les garçons.

Plusieurs facteurs ont contribué à la généralisation de la scolarité au primaire, dont les principaux on peut citer :

- le soutien social à travers l'élargissement du réseau des cantines scolaires, la distribution des manuels et des fournitures scolaires en faveurs des élèves des familles vulnérables.

- Le programme de transferts monétaires conditionnels (programme TAYSSIR).

- Au niveau de l'infrastructure, l'élargissement du réseau des établissements scolaires par la création de nouvelles écoles et satellites ; ce qui a réduit les disparités de scolarisation entre milieux urbain et rural.

- La création des écoles communautaires avec internats et/ou transport scolaire. Leur nombre a augmenté de 13 entre 2009-2010 à 68 entre 2013-2014.

- An niveau juridique, la promulgation de la loi n° 04-00 qui reconnaît l'éducation comme étant à la fois un droit et une obligation pour tous les enfants marocains des deux sexes ayant atteint l'âge de six ans. Par ailleurs, si l'enseignement primaire a connu un grand accroissement au niveau quantitatif, il connaît de grands problèmes au niveau de sa qualité. Théoriquement, la question de la qualité d'enseignement constitue la pierre angulaire de la réforme introduite depuis les années 2000, et la charte nationale de l'éducation et la formation a prescrit tout un processus pour améliorer la qualité des

apprentissages acquis à l'école. Ainsi, les curricula ont été révisés selon l'approche par compétences avec introduction de nouvelles matières, la réforme des manuels scolaires concrétisée par la libération du livre scolaire et l'introduction des valeurs de démocratie et de droit de l'homme dans ces manuels, l'introduction des technologies de l'information et de la communication dans l'enseignement-apprentissage.

Or, des études nationales et internationales sur le fonctionnent du système éducatif et le niveau de performance des élèves montrent la persistance d'un certain nombre de dysfonctionnements :

- Les méthodes d'enseignement ne sont pas centrées sur l'élève et se fondent sur la transmission directe des savoirs, sans donner de place aux démarches par découverte qui ont le potentiel de construire les connaissances et de s'adapter à des situations nouvelles.

- Les programmes scolaires sont prédominés par des savoirs livresques au détriment des savoirs pratiques, ce qui empêche le développement des compétences supérieures chez l'élève.

- La problématique des langues est considérée comme le principal obstacle à l'amélioration de la qualité du système éducatif, car elle concerne, à la fois, les langues d'enseignement et l'enseignement des langues. En outre, après la constitutionnalisation de la langue amazighe comme langue national en plus de la langue arabe, la question de langues a donné lieu à un vif débat dans la société et les médias ces dernières années.

La conséquence directe de ces problèmes est une faiblesse évidente des élèves marocains en sciences, mathématiques et lecture. Cette faiblesse est constatée par des enquêtes internationales d'évaluation des acquis scolaires auxquelles le Maroc a participé (CSEFRS du Maroc, 2008), et une étude récente sur la lecture au Maroc (surtout dans les premiers niveaux de l'école primaire) faite par le ministère de l'éducation nationale en partenariat avec l'agence américaine de la coopération internationale USAID[7] .

3 L'importance de l'enseignement des mathématiques dans les programmes scolaires

L'enseignement des mathématiques joue un rôle déterminant dans le développement des capacités intellectuelles et affectives des élèves, ce qui justifie la place privilégiée qu'occupe cette discipline dans le système scolaire (Elfassi Elfihri O., 2012, p. 271). Pour illustrer cette importance nous exposons l'enveloppe horaire réservée à cette discipline dans les cycles de l'école primaire.

L'organisation de l'année scolaire est répartie sur 34 semaines dont 24 semaines (70%) seront exploitées à la présentation des contenus des programmes scolaires, et 10 semaines (30%) seront réservées à l'évaluation et le soutien.

[7] Ministère de l'éducation nationale du Maroc Espace partenaire ; Bulletin d'information et de communication, N°10, 1 juillet 2014.

Concernant l'enseignement des mathématiques, le livre blanc lui consacre un volume horaire annuel de 170 heures[8]

Quant à la semaine pédagogique, elle se compose :

- en cycle premier, de 10 séances à raison de deux leçons par semaine, présentées en 4 séances de 30 minutes chacune et 2 séances de 30 minutes chacune pour le soutien des deux leçons.

- En cycle intermédiaire, de 7 séances à raison de 2 leçons par semaine, présentées en 3 séance de 45 minutes chacune et une séance de 30 minutes pour le soutien des deux leçons.

L'analyse de ces données montre l'importance donnée à l'enseignement des mathématiques dans le curriculum scolaire marocain. En effet, presque un cinquième (17.86%) de l'enveloppe horaire d'enseignement est consacré à l'enseignement des mathématiques. En outre, les élèves confrontent chaque jour une séance de mathématiques. Or, la performance des élèves en mathématiques ne reflète pas cette importance, et des tests nationaux et internationaux font la preuve de ce constat en montrant des disparités de performance selon le sexe, le milieu, les régions et les domaines de contenu. D'où vient le problème alors ? C'est cette question qui donne naissance à la problématique de notre recherche.

4 Le rôle de la résolution de problème dans l'enseignement des concepts mathématiques

Dans une conception constructiviste (et aussi socioconstructiviste), un élève ne peut acquérir une nouvelle connaissance qu'à partir du moment où il constate l'insuffisance de ses connaissances actuelles. En effet, selon cette conception, il est important d'amener l'élève à former les notions et à découvrir lui-même les relations et les propriétés mathématiques, plutôt que de lui imposer une pensée adulte toute faite. Ainsi, le rôle de l'enseignant ne consiste pas à communiquer les savoirs qu'il veut enseigner, mais de proposer des situations qui vont permettre à l'élève la construction d'une première solution en mobilisant ses connaissances antérieures. Cette démarche permet, d'une part, de repérer les obstacles à l'apprentissage et, d'autre part, de sélectionner parmi ceux-ci l'obstacle à franchir. Selon la théorie de situations didactiques de Brousseau, l'enseignant doit intégrer dans la situation didactique initiale une situation a-didactique où il se met à côté pour laisser l'élève agir libre en face du milieu. Seul ou en collaboration avec ses collègues, l'élève va mobiliser des connaissances et choisir entre plusieurs voies pour trouver une réponse au problème posé. Mais si cette situation est qualifiée a-didactique cela ne veut pas dire d'annuler le cadre didactique où est traitée la situation initiale, car, d'une part, c'est en fonction d'objectifs prédéfinis que l'enseignant laisse l'élève travailler seul et, d'autre part, ce dernier sait bien que cette situation est soumise par le professeur et à l'école. En fait, l'enjeu est de mettre le plus possible à distance l'intention didactique pour laisser jouer au maximum les mécanismes d'appropriation par les élèves du problème et de son dépassement (Dupin et Joshua, 1993, p. 261). Dans le même sens, et selon la théorie de champs conceptuels

[8] - Ministère de l'éducation nationale du Royaume du Maroc, juin2002, le livre blanc, volume 2, p 12-13

de Vergnaud, la construction et l'appropriation de toutes les propriétés d'un concept est un processus long et ne peut être appréhendé si le concept est considéré dans sa globalité. Il faut donc procéder à des coupures parce qu'on ne peut tout étudier à la fois.

5 Méthodologie

Pour réaliser cette recherche, nous avons procédé à une investigation par questionnaire auprès des enseignants d'écoles primaire.

5.1 Terrain de la recherche

Dans le cadre de ces investigations relatives aux représentations des enseignants sur l'enseignement de la résolution de problèmes mathématiques, nous avons constitué un échantillon de professeurs de l'enseignement primaire issus des écoles de trois délégations provinciales de l'éducation nationale au Maroc : Guerif, Taza et Jerada, relevant de deux académies régionales de l'éducation et de la formation : l'académie de la région Taza-Al Hoceima-Taounate et l'académie de la région orientale. Ainsi, 73 enseignants ont été enquêtés par un questionnaire anonyme au cours des mois de février et mars de l'année 2015.

5.2 Hypothèses

A travers le recueil de données et l'analyse des réponses aux questions relatives aux variables, que nous exposerons au paragraphe suivant, nous cherchons à vérifier trois hypothèses principales qui orientent cette recherche :

Nos hypothèses sont les suivantes :

Hypothèse 1 :

Le parcours scolaire de l'enseignant a un impact sur ses pratiques d'enseignement relatives aux concepts mathématiques (modalités de préparation, de présentation et de correction de problèmes mathématiques).

Hypothèse 2 :

Les enseignants n'ont pas de représentations homogènes relatives à l'autonomie des élèves lors de la résolution de problèmes mathématiques.

Hypothèse 3 :

L'ancienneté des enseignants dans l'enseignement des mathématiques permet l'amélioration de leurs pratiques d'enseignement de la résolution de problèmes mathématiques.

5.3 Variables et résultats

Notre questionnaire de recherche se compose de 23 variables (V01 à V23) se rapportant à 23 items. La réponse à chaque item sera par « oui » ou « non ». Les réponses par « oui » seront codées par « 1 » et celles par « non » le seront par « 0 ». Les 23 variables sont :

Codes	Libellés/modalités
V01	Sexe : Homme / Femme
V02	Vous êtes titulaire d'un Bac scientifique
V03	Vous êtes titulaire d'un Bac littéraire
V04	Vous êtes titulaire d'un Bac en d'autre spécialité
V05	Votre ancienneté générale dans l'enseignement primaire est de [1 ; 10[
V06	Votre ancienneté générale dans l'enseignement primaire est de [10. 20[
V07	Votre ancienneté générale dans l'enseignement primaire est de [20. 30[
V08	Votre ancienneté générale dans l'enseignement primaire est de [30. 40[
V09	Votre ancienneté dans l'enseignement des mathématiques est de [1 ; 4[
V10	Votre ancienneté dans l'enseignement des mathématiques est de [4. 12[
V11	Votre ancienneté dans l'enseignement des mathématiques est de [12. 20[
V12	Votre ancienneté dans l'enseignement des mathématiques est de [20. 30[
V13	Votre ancienneté dans l'enseignement des mathématiques est de [30. 40[
V14	Adaptez-vous les situations problèmes aux spécificités cognitives et socioculturelles de vos élèves
V15	Pensez-vous que vos élèves soient-ils capables de résoudre des problèmes mathématiques en autonomie
V16	Lors de la résolution de problème vous préférez que les élèves travaillent individuellement
V17	Lors de la résolution de problème vous préférez que les élèves travaillent en petits groupes
V18	Lors de la résolution de problème vous préférez que les élèves travaillent collectivement
V19	En préparant vos séquences relatives à la résolution de problème, vous le résolvez préalablement par écrit
V20	En préparant vos séquences relatives à la résolution de problème, vous le résolvez le plus souvent mentalement
V21	En préparant vos séquences relatives à la résolution de problème, vous recherchez par écrit plusieurs façons de résoudre le problème
V22	Un élève corrige au tableau, ses pairs le suivent
V23	Un élève corrige au tableau, ses pairs l'aident

Tableau 1 – Variables

Les résultats obtenus sont exposés dans les tableaux situés en annexe.

6 Analyse des résultats et discussions

Pour savoir si des liens existent entre ces divers items nous analysons les résultats obtenus en appliquant la méthode Analyse Statistique Implicative (ASI).

Les variables Vi (i = 1 à 23) seront traitées comme variables binaires. Grâce à la méthode ASI, nous étudiant l'arbre de similarité, le graphe implicatif et l'arbre cohésitif.

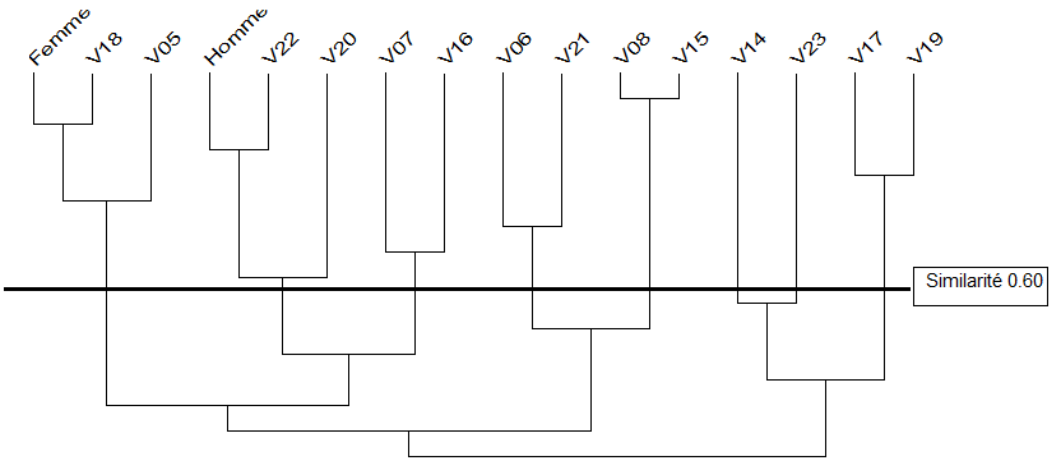


Figure 1 – Arbre des similarités

En se limitant à une similarité de 0.6, nous distinguons huit niveaux de classification, à savoir :

Classification au niveau : 1 : (V08 V15) similarité : 0.979845
Classification au niveau : 2 : (Femme V18) similarité : 0.976707
Classification au niveau : 3 : (Homme V22) similarité : 0.871728
Classification au niveau : 4 : (V17 V19) similarité : 0.855584
Classification au niveau : 5 : ((Femme V18) V05) similarité : 0.816245
Classification au niveau : 6 : (V06 V21) similarité : 0.674984
Classification au niveau : 7 : (V07 V16) similarité : 0.656273
Classification au niveau : 8 : ((Homme V22) V20) similarité : 0.635305

Tableau 2 – Classes de similarité retenues avec un indice supérieur à 0.60

Nous voyons quatre classes isolées, à savoir (V08, V15) ; (V17, V19) ; (V06, V21) ; (V07, V16) et deux autres grandes classes qui sont ((Femme, V18), V05)) et ((Homme, V22), V20)).

Contribution à la classe : V08, V15 (1)
Groupe optimal: Ind 30 Ind 43
Card GO = 2 p = 0.0274 1-p = 0.973
La variable qui contribue le plus à cette classe est V13 avec un risque de : 0

Les enseignants qui jouissent d'une ancienneté comprise entre 30 et 40 ans dans l'enseignement primaire en général et dans l'enseignement des mathématiques en particulier voient leurs élèves aptes à résoudre des problèmes mathématiques en autonomie. Ceci pourrait bien ***confirmer notre troisième hypothèse*** dans la mesure où, dans un premier temps, un travail autonome, est en faveur de l'élève. En effet, il lui offre l'occasion de se confronter avec lui-même en face d'une situation problème donnée et d'essayer d'apporter des solutions en cas d'obstacles. Par opposition au travail collectif qui ne permet pas à l'élève de se situer au sein du problème et encore moins de repérer ses défaillances car très tôt la bonne réponse sera donnée par un pair ou par l'enseignant.

D'un autre angle, ce point de vue est nettement adopté avec un risque de 0.0238 par les enseignants titulaires d'un baccalauréat qui n'est ni littéraire ni scientifique mais

plutôt d'un bac d'une autre spécialité (des spécialités ayant un caractère professionnel : fabrication mécanique, secrétariat, comptabilité,…). En revanche, ceux issus d'un bac scientifique croient très peu en la capacité des élèves à travailler en autonomie.

Ces observations ***confirment a priori notre première hypothèse*** qui stipule que le parcours scolaire des enseignants influence leurs pratiques enseignantes.

Contribution à la classe : V17,V19 (4)
Groupe optimal :
Ind 61 Ind 68 Ind 67 Ind 6 Ind 64 Ind 25 Ind 39 Ind 43 Ind 32 Ind 27 Ind 35 Ind 45 Ind 52 Ind 53 Ind 60
Ind 19 Ind 47 Ind 50
Card GO = 18 p = 0.247 1-p = 0.753
La variable qui contribue le plus à cette classe est V04 avec un risque de : 0.0463

Les enseignants qui préfèrent que leurs élèves résolvent les problèmes mathématiques en petits groupes, autrement dit, ceux qui sont en faveur d'une co-construction de savoirs, résolvent eux-mêmes les mêmes problèmes préalablement par écrit. Cette pratique est acquiescée essentiellement par les enseignants titulaires d'un baccalauréat d'une spécialité autre que la littéraire ou la scientifique et ce avec un risque de 0.0463, ce qui pourrait encore ***confirmer notre première hypothèse***

Contribution à la classe : V06,V21 (6)
Groupe optimal :
Ind 53 Ind 72 Ind 66 Ind 65 Ind 59 Ind 60 Ind 71 Ind 56 Ind 69 Ind 57 Ind 52
Ind 13 Ind 14 Ind 17 Ind 18 Ind 19 Ind 5 Ind 8 Ind 9 Ind 11 Ind 12 Ind 23
Ind 37 Ind 39 Ind 42 Ind 48 Ind 51 Ind 24 Ind 25 Ind 27 Ind 28 Ind 29
Card GO = 32 p = 0.438 1-p = 0.562
La variable qui contribue le plus à cette classe est V11 avec un risque de : 0.0006

Les enseignants dont l'ancienneté dans l'enseignement primaire est comprise entre 10 et 20 ans résolvent les problèmes mathématiques qu'ils proposent à leurs élèves par écrit et de plusieurs façons. Une telle méthode est approuvée par la quasi-totalité des enseignants ayant une ancienneté entre 12 et 20 ans dans l'enseignement des mathématiques et est très peu adoptée par ceux titulaires d'un baccalauréat littéraire.

Contribution à la classe : V07, V16 (7)
Groupe optimal :
Ind 38 Ind 73 Ind 3 Ind 31 Ind 46 Ind 22 Ind 21 Ind 20 Ind 64 Ind 62 Ind 44
Ind 58
Card GO = 12 p = 0.164 1-p = 0.836
La variable qui contribue le plus à cette classe est V12 avec un risque de : 7.9e-05

Les enseignants d'une ancienneté dans l'enseignement primaire entre 20 et 30 ans préfèrent le travail individuel des élèves, autrement dit, ils sont en faveur du travail autonome des élèves. Une telle pratique est choisie par la presque totalité des enseignants ayant une ancienneté dans l'enseignement des mathématique comprise entre 20 et 30 ans, pour autant elle est nettement réfutée par ceux qui n'ont qu'une ancienneté entre 1 et 4 ans.

Ce fait vient confirmer l'idée mise en exergue précédemment et prouvant que plus l'ancienneté dans l'enseignement des mathématiques est élevée plus les enseignants

sont confiants en le travail autonome de leurs élèves ce qui ***confirme bien notre deuxième hypothèse.***

Contribution à la classe : Femme,V18, V05 (2,5)
Groupe optimal :
Ind 67 Ind 68 Ind 14 Ind 13 Ind 12 Ind 26 Ind 63 Ind 37 Ind 36
card GO = 9 p = 0.123 1-p = 0.877
La variable qui contribue le plus à cette classe est V04 avec un risque de : 0.0174

Les femmes d'une ancienneté dans l'enseignement primaire comprise entre 1 et 10 ans préfèrent une correction collective des problèmes mathématiques, technique pratiquée par la majorité des enseignants titulaires d'un bac d'une spécialité autre que scientifique ou littéraire, ce qui ***confirme davantage notre première hypothèse.***

Nous exposons, maintenant, ce que nous rapporte le graphe implicatif (figure 2) :

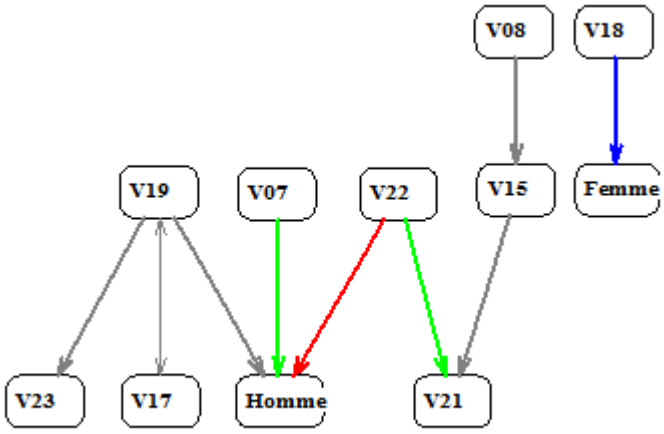


Figure 2 – Graphe implicatif

Au niveau de confiance de 0.92 donné par la flèche rouge, une quasi implication se voit nette entre le fait qu'un élève corrige le problème au tableau et ses pairs le suivent passivement et le fait d'être homme. Autrement dit cette pratique de correction de problème est adaptée particulièrement par les enseignants de sexe masculin.

A un niveau de confiance plus bas, de 0.87 (flèche bleue), encore une quasi implication s'installe entre le fait que : pour résoudre un problème mathématique, les élèves travaillent collectivement et le fait que l'enseignante soit une femme.

En descendant à un niveau de confiance encore plus bas, soit 0.84 (flèches vertes), on voit une quasi implication entre le procédé de correction du problème mathématique et la façon qu'adopte l'enseignant pour préparer préalablement cette correction. Autrement dit, si dans sa classe la correction est faite par un élève en la présence passive de ses pairs, l'enseignant prépare préalablement la correction par écrit et par plusieurs manières. Ceci pourrait s'interpréter par le fait que l'enseignant est persuadé que la manière par laquelle procèdera l'élève est imprévisible vu que plusieurs façons seront envisageables, et l'enseignant devrait être prêt pour chacune d'entre elles.

A un niveau de confiance de 0.75 (flèches grises), on voit que les enseignants qui préparent les séquences relatives à la correction de problèmes mathématiques

préalablement par écrit sont ceux qui, au cours de la correction de ce même problème désignent un élève pour assurer cette correction avec la coopération de ses pairs.

Une quasi implication, et dans les deux sens, est nette au même degré de confiance (0.75) annonçant que les enseignants qui résolvent les problèmes mathématiques préalablement par écrit sont les mêmes qui préfèrent que la correction des problèmes soient faite par un élève au tableau avec l'aide de ses pairs.

En mettant femme et homme en variables supplémentaires, nous obtenons le graphe suivant (figure 3) qui, à un niveau de confiance 0.75 (flèches vertes) montre une quasi implication entre le fait d'avoir une ancienneté importante dans l'enseignement primaire (entre 30 et 40 ans) et le fait d'être convaincu de la compétence des élèves dans la résolution de problèmes mathématiques en autonomie. Ces mêmes enseignants résolvent le problème posé préalablement par écrit et par multiples façons. Ce qui ***confirme mieux notre troisième hypothèse*** au même niveau de confiance de 0.75.

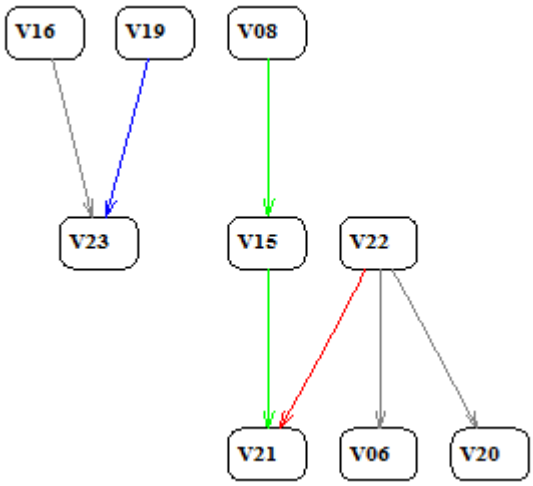


Figure 3 – Graphe implicatif avec homme et femme variables supplémentaires

A un niveau de confiance plus élevé, soit 0.80 (flèche bleue), on peut dire que les enseignants qui, en préparant les séquences relatives à la résolution de problème, le résolvent préalablement par écrit optent pour qu'un élève corrige le problème au tableau tout en étant guidé et aidé par ses pairs.

Nous analysons, enfin, les données de l'arbre cohésitif qui nous renseigne sur le degré de cohésion entre les différentes variables préalablement fixées. En gardant la variable sexe comme variable supplémentaire, nous obtenons l'arbre de cohésion suivant :

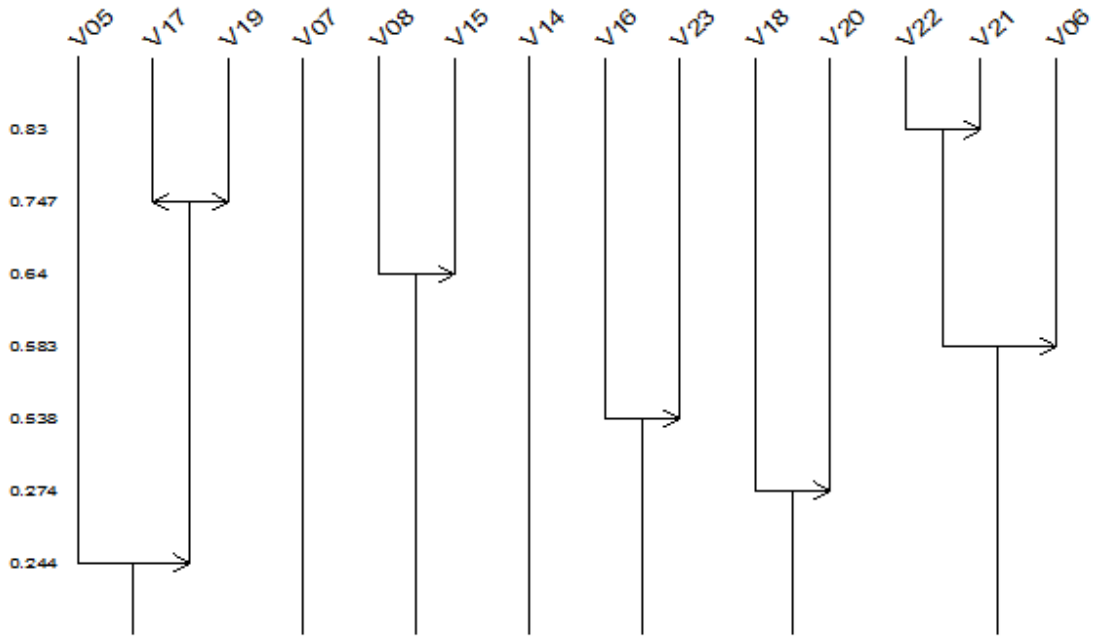


Figure 4 – Arbre cohésitif

Sachant que plus le degré de cohésion est proche de 1, plus la cohésion entre les variables est forte.

En effet, une forte cohésion (0.83) est décelable entre V22 (Un élève corrige au tableau, ses pairs le suivent) et V21 (En préparant vos séquences relatives à la résolution de problème, vous recherchez par écrit plusieurs façons de résoudre le problème).

Une cohésion notable (0.747) mais moins importante que la précédente est visible entre V17 (Lors de la résolution de problème vous préférez que les élèves travaillent en petits groupes) et V19 (En préparant vos séquences relatives à la résolution de problème, vous le résolvez préalablement par écrit).

Un degré de cohésion de 0.64 est entre V08 (Votre ancienneté générale dans l'enseignement primaire est de [30. 40[) et V15 (Pensez-vous que vos élèves soient-ils capables de résoudre des problèmes mathématiques en autonomie), ce qui ***confirme encore notre deuxième hypothèse.***

L'ensemble des résultats fournis par l'arbre des similarités, le graphe implicatif et l'arbre cohésitif nous amènent à confirmer nos trois hypothèses.

7 Conclusion générale

A partir de ces résultats, nous constatons clairement que les enseignants n'ont pas les mêmes les mêmes représentations à propos des modalités de préparer, présenter et corriger une séquences de résolution de problèmes mathématiques. Ainsi nous voyons que leurs pratiques d'enseignement dépendent à la fois de leurs parcours scolaires et de leurs anciennetés dans l'enseignement des mathématiques. En effet, plus l'enseignant a une ancienneté importante dans l'enseignement primaire en général, et dans l'enseignement de mathématiques en particulier, plus il croie en la capacité de ses élèves à résoudre des problèmes de manière autonome ; et pour cela, il résout le

problème posé préalablement par écrit et par multiples façons et préfère que ses élèves travaillent en petit groupe pour la Co-construction des connaissances mathématiques.

Concernant le parcours scolaire de l'enseignant, nous voyons qu'il joue un rôle important sur les représentations dont l'enseignant pense l'enseignement de la résolution de problèmes mathématiques. Et ce sont les enseignants titulaires d'un baccalauréat professionnel (fabrication mécanique, secrétariat, comptabilité…) qui voient leurs élèves aptes à résoudre des problèmes mathématiques en autonomie et en travaillant en petits groupes. Ce constat peut se justifie par le fait que ces enseignants avaient, eux-mêmes, l'habitude d'apprendre en collaboration avec leurs collègues pour réaliser des tâches prédéterminées.

L'hétérogénéité des représentions des enseignants nous amènent à poser la question sur l'importance et la qualité de la formation initiale et continue qu'ils ont reçus dans les centres de formations des professeurs de primaire, sachons nous que des enseignants sont recrutés sans recevoir aucune formation pédagogique ou didactique. Ce qui peut être à l'origine des difficultés rencontrées par les élèves dans l'apprentissage des concepts mathématiques.

Références

[1] Boulahcen A. (2002), *Sociologie de l'éducation : Les systèmes éducatifs en France et au Maroc, étude comparative*, Casablanca : 2° édition 2010, Afrique Orient,.

[2] Dupin J.J. et S. Johsua (1993), *Introduction à la didactique des sciences et des mathématiques*, Paris : PUF

[3] El Ayadi M. (2009), L'école marocaine entre la cohérence du modèle et l'anarchie du réel, en arabe, *AlMadrassa*, n° 1, 38-41.

[4] Elfassi Elfihri O. (2012), Les sciences exactes à l'école marocaine, en arabe, *AlMadrassa*, n° 4-5, 271-286.

[5] Rapport thématique 2009 sur les résultats du programme national d'évaluation des acquis PNEA 2008, fascicule des mathématiques, le conseil supérieur de l'éducation de la formation et de la recherche scientifique du Maroc, http : //www.csefrs.ma (consulté le 10/01/2015).

Annexe

	F	H	V02 s	V03 s	V04 s	V05	V06	V07	V08	V09 s	V10 s	V11 s	V12 s	V13 s	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23
I01	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1
I02	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0
I03	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	1
I04	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1
I05	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	1	1	1
I06	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1
I07	0	1	0	1	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1	1	1	0
I08	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1
I09	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	1	0	1
I10	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	1	0	1
I11	1	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1
I12	1	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	1
I13	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	1
I14	1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	1
I15	0	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	1	1
I16	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1
I17	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1	1
I18	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0	1
I19	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0	1	1	1	1	1
I20	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0	1
I21	0	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0	1
I22	1	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0	0	1	0	1
I23	0	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1	1
I24	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1
I25	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1
I26	1	0	0	1	0	1	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	1
I27	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1
I28	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1
I29	0	1	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	1
I30	0	1	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	1
I31	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	0	0	1	0	1
I32	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1
I33	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	1	1
I34	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0
I35	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	1
I36	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0
I37	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	1

Tableau 3 - Résultats du questionnaire

	F	H	V02 s	V03 s	V04 s	V05	V06	V07	V08	V09 s	V10 s	V11 s	V12 s	V13 s	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23
I38	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	0	1
I39	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1	0	1
I40	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1
I41	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1
I42	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1
I43	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1	0	1	0	1	0	0	0	1
I44	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	1	0	1
I45	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	1	1	1
I46	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	1
I47	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	1
I48	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	0
I49	0	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1
I50	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	1	0	1	0	1	0	1
I51	1	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	1
I52	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0
I53	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	0	1	0	1	0	1
I54	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1
I55	1	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0	1
I56	0	1	0	0	1	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	1	0	1
I57	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	1
I58	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0	1
I59	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	1	1
I60	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	1	1
I61	0	1	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	0	1	1
I62	0	1	0	1	0	0	0	1	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0	1
I63	1	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	1	1
I64	0	1	0	0	1	0	0	1	0	0	0	0	1	0	1	1	1	1	0	1	1	1	1	1
I65	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	1
I66	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	1
I67	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	1	1	0	1	0	0	0	1
I68	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	1	0	1
I69	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	1	1	1	0	1
I70	0	1	0	1	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	1	1
I71	0	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	1	0	1
I72	1	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0
I73	0	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	1	0	1	1	1

Tableau 4 – Résultats du questionnaire

MESTRADO PROFISSIONAL EM ENSINO DE MATEMÁTICA: ARTICULAÇÃO DE COMPETÊNCIAS

Silvia Maria ISAIA[1] Eleni BISOGNIN[2] Vanilde BISOGNIN[3] Jean-Claude REGNIER[4] Nadja ACIOLY-RÉGNIER[5] Andréia CARDOSO SILVEIRA[6]

MASTER PROFESSIONNEL EN ENSEIGNEMENT DE MATHÉMATIQUE: ARTICULATION DE COMPÉTENCES

PROFESSIONAL MASTER'S DEGREE IN TEACHING OF MATHEMATICS: ARTICULATION OF COMPETENCES

RESUMO

Neste artigo, discutem-se os resultados oriundos da realização de um projeto de pesquisa que teve como intuito oferecer suporte teórico-metodológico para acompanhar e compreender a dinâmica formativa do Curso de Mestrado Profissional em Ensino de Matemática de uma IES comunitária do Rio Grande do Sul, Brasil. Para tanto, entende-se que o curso precisa oferecer condições necessárias para que o desenvolvimento profissional docente de seus protagonistas se efetive. Os sujeitos envolvidos são estudantes e egressos do curso, que atuam na educação básica e superior, num total de 20 participantes. Para essa discussão, o problema proposto é o seguinte: Como estudantes e egressos percebem as competências decorrentes do curso de Mestrado e das Diretrizes Curriculares Nacionais das Licenciaturas de Matemática? Esse problema justifica-se pelo fato de o curso ter, entre outros, o objetivo de capacitar, em nível de pós-graduação *stricto sensu*, professores de Matemática em exercício nos sistemas de ensino nos níveis fundamental, médio e licenciatura. Buscou-se, assim, analisar as competências esperadas na formação continuada dos sujeitos participantes. Como resultado parcial, é possível verificar que o cruzamento entre os quadros e gráficos relativos às competências desenvolvidas pelo mestrado e as indicadas pelas DCN de Matemática demonstram a coerência dos sujeitos participantes em assinalar o grau de importância de cada uma, evidenciando correlação entre elas.

***Palavras-chave**: ensino de matemática, competências, desenvolvimento professional docente.*

---

1 Professora doutora do Programa de Pós-Graduação em Ensino de Ciências e Matemática do Centro Universitário Franciscano, Rio Grande do Sul, Brasil, silviamariaisaia@gmail.com

2 Professora doutora do Programa de Pós-Graduação em Ensino de Ciências e Matemática do Centro Universitário Franciscano, Rio Grande do Sul, Brasil, eleni.bisognin@gmail.com

3 Professora Doutora do programa de Pós-Graduação em Ensino de Ciências e Matemática do Centro Universitário Franciscano, Rio Grande do Sul, Brasil, vanilde.bisiognin@gmail.com

4 Pesquisador Visitante Especial PVE/CAPES no PPGEC/UFRPE (Programa de Pós-Graduação em Ensino das Ciências da Universidade Federal Rural de Pernambuco) – UMR5191 – ICAR Université Lumiere Lyon 2 (FRA), jean-claude.regnier@univ-lyon2.fr

5 Universidade Lumière Lyon 1; França; nadja.acioly-regnier@univ-lyon1.fr

6 Universidade Federal da Bahia; Brasil; Membro do Projeto "Determinantes da Equidade no Ensino Superior" financiado pelo Programa OBEDUC/CAPES; Universidade Lumière Lyon 2 ; França; andreia-car@hotmail.com.

RESUME

Dans cet article, nous discutons les résultats provenant de la réalisation d'un projet de recherche qui avait pour but d'offrir un support théorique-méthodologique pour accompagner et comprendre la dynamique de formation d'un cours de Mestrado Professionnel d'enseignement des Mathématiques dans Rio Grande du Sul - Brésil. Pour ce faire, nous considérons que le Mestrado doit offrir des conditions nécessaires, pour que le développement professionnel de l'enseignant en formation soit effectif. Les sujets impliqués sont des étudiants en cours de formation et des diplômés qui travaillent dans l'enseignement aux niveaux primaire, secondaire et supérieur. L'échantillon est 20 sujets. Pour cette discussion, le problème proposé est: Comment les étudiants et les diplômés perçoivent les compétences issues du Master et des Directives Curriculaires Nationales des « Licenciaturas » en Mathématiques? Ce problème est justifié par le fait que le cours du Mestrado doit avoir, parmi d'autres, l'objectif de qualification, au niveau du Mestrado *stricto sensu*, des professeurs de Mathématiques pour exercer leur métier. Nous avons ainsi cherché à analyser les compétences attendues dans la formation continue des sujets. Comme résultat partiel, il est possible de mettre en évidence que le croisement de tableaux et des graphiques relatifs aux compétences développées par le Mestrado et indiquées par DCN en Mathématique démontre la cohérence des sujets au travers la façon de traduire le degré d'importance de chaque compétence.

***Mots-clés**: enseignement des mathématiques, compétences, développement professionnel enseignant.*

ABSTRACT

In this article we discuss the results that come from a research project whose aim to offer theoretical and methodological support to follow and understand the formative dynamics of the Professional Master in Mathematics Teaching at a community college in Rio Grande do Sul, Brazil. Therefore, it is understood that the course needs to provide necessary conditions for the professional development of teachers of its protagonists become effective. The subjects involved are students and course graduates who work in primary and higher education, a total of 20 participants. For this discussion, the proposed problem is this: As students and graduates realize the responsibilities arising from the Master course and the National Curriculum Guidelines for Undergraduate Mathematics? This issue is justified by the fact that the course has, among others, the aim of training at the level of post-graduate studies, mathematics teachers working in education in primary, secondary and higher education. It sought therefore to analyze the competencies present in the continuing education of participating subjects. As a partial result, you can verify that the intersection between the tables and charts on skills developed by masters and indicated by National Curriculum Guidelines for Undergraduate Mathematics demonstrate the coherence of research subjects in indicating the degree of importance of each, showing correlation between them.

***Keywords**: Mathematics Teaching. Continuing Education of Teachers. Competences.*

1 Introduction

Neste artigo, discutem-se os resultados oriundos de um projeto de pesquisa que teve como intuito oferecer suporte teórico-metodológico para acompanhar e compreender a dinâmica formativa do Curso de Mestrado Profissional em Ensino de Matemática de uma IES Comunitária do Rio Grande do Sul, Brasil, tendo por horizonte as competências para o ensino, vinculadas às Diretrizes Curriculares Nacionais (DCN) para a Licenciatura de Matemática (2001), bem como o Projeto Pedagógico (PP) do Curso de Mestrado Profissional em Ensino de Física e Matemática, que foi aprovado pela CAPES em 2003 e está em funcionamento desde 2004. Os sujeitos envolvidos

foram professores (mestrandos) e mestres (egressos) que atuam na educação básica e superior.

O objeto de estudo, aqui colocado, circunscreve-se às competências decorrentes do mestrado e das DCN para a Licenciatura de Matemática (2001), constantes em uma enquete respondida por 20 sujeitos. Nesse sentido, o problema que se coloca é: *Como estudantes e egressos percebem as competências decorrentes do curso de mestrado e das Diretrizes Curriculares Nacionais da Licenciatura de Matemática?*

O problema justifica-se por considerar a formação continuada de professores, desembocando no desenvolvimento profissional docente. Com esse intuito, procurou-se aliar as competências das DCN e as do curso de mestrado para investigar se esse está cumprindo com os objetivos, colocados no seu PP (2003), a saber: a) capacitar, em nível de pós-graduação *stricto sensu*, profissionais do ensino de Física e de Matemática que estejam em exercício nos sistemas de ensino, nos níveis fundamental, médio e superior (licenciaturas); b) oportunizar aos docentes que atuam na educação básica (ensino fundamental e ensino médio) e na licenciatura de Matemática, tendo como ponto de referência a realidade sociocultural e seus determinantes histórico-culturais, pedagógicos e científicos, com vistas a proposições de alternativas para a melhoria da qualidade do ensino no contexto social de abrangência do Curso.

2 Referencial teórico

Os referenciais teóricos que dão sustentação à discussão empreendida neste artigo envolvem temas relativos: às competências profissionais docentes; às *práticas pedagógicas* que darão base às competências em construção; ao desenvolvimento profissional dos professores, tendo por núcleo a formação docente; à compreensão sobre as *áreas específicas de conhecimento*, envolvendo conhecimento científico, acadêmico e escolar; ao *processo de ensinar e aprender*.

2.1 Competências docentes

Quando se pensa na profissão professor, subentende-se a noção de saber, percebida em seu sentido amplo, como aquela que “engloba os conhecimentos, as competências, as habilidades (ou aptidões) e as atitudes, isto é, aquilo que, muitas vezes, foi chamado de saber, saber-fazer e saber-ser” (TARDIF, 2002, p.255).

O conceito de competências para Esteves (2009) e Perrenoud (1999, 2000), apesar de sua polifonia, pode ser utilizado para a formação e desenvolvimento profissional dos professores, bem como para aferir a aprendizagem dos alunos. Parte-se, portanto, da ideia de que competências envolvem um complexo de conhecimentos e habilidades que permite a resolução de situações para as quais não se tem uma solução única. Por sua vez, para Zabala e Arnau (2009) as competências consistem na capacidade do professor e do aluno em mobilizar atitudes, habilidades e conhecimentos de forma inter-relacionada, a fim de poderem resolver de forma eficaz questões decorrentes do âmbito educativo. Consequentemente, elas envolvem componentes atitudinais, procedimentais e conceituais.

Nesse sentido, é necessário que a edução básica e superior contribuam para que os estudantes possam desenvolver competências que lhes permitam viver e conviver em

uma sociedade cada vez mais complexa, envolvendo, entre outras, ferramentas para pensar: a linguagem, as tecnologias, os símbolos e pricipalmente a capacidade para atuar em um grupo diversificado e de maneira autônoma.

As diretrizes para formação de professores de Matemática (BRASIL, 2001), visando ao seu desenvolvimento profissional, consideram que o conhecimento matemático, habilidades, hábitos e competências são importantes para a formação de cidadãos competentes. O ambiente social, o papel do professor e as práticas de sala de aula são considerados essenciais para obtê-las.

Conforme essas mesmas diretrizes, os cursos precisam desenvolver nos formandos, entre outras habilidades e competências, as de

> estabelecer relações entre a Matemática e outras áreas do conhecimento; trabalhar na interface da Matemática com outros campos de saber; identificar, formular e resolver problemas na sua área de aplicação, utilizando rigor lógico-científico na análise da situação-problema; compreender, criticar e utilizar novas ideias e tecnologias para a resolução de problemas (BRASIL, 2001, p. 4).

Especificamente, em relação ao educador matemático, o mesmo documento sugere que o licenciado em Matemática necessita ter a capacidade de

> desenvolver estratégias de ensino que favoreçam a criatividade, a autonomia e a flexibilidade do pensamento matemático dos educandos; buscando trabalhar com mais ênfase nos conceitos do que nas técnicas, fórmulas e algoritmos; elaborar propostas de ensino-aprendizagem de Matemática para a Educação Básica; analisar, selecionar e produzir materiais didáticos; perceber a prática docente da Matemática como um processo dinâmico, carregado de incertezas e conflitos, um espaço de criação e reflexão, onde novos conhecimentos são gerados e modificados continuamente (BRASIL, 2001, p. 4).

2.2 Desenvolvimento profissional docente

Os achados teóricos e investigativos de diversos pesquisadores (ABRAHAM, 2000; NÓVOA, 2009; ZABALZA, 2004; MARCELO GARCIA, 1999, 2013; IMBERNÓN, 2006, 2009; ISAIA, 2007; ISAIA e BOLZAN, 2009) que têm se debruçado sobre o desenvolvimento profissional docente, articulado ao processo formativo, levam à compreensão de que esse desenvolvimento envolve um processo amplo que engloba a dimensão pessoal, pedagógica e profissional dos professores, entendidos como seres unitários entretecidos pelo percurso pessoal (ciclo vital) e também pelo profissional (os diversos caminhos construídos ao longo da profissão).

Dessa forma, é necessário indagar qual é a natureza da profissão docente e quais são os condicionantes capazes de estabelecê-la. Entende-se que o percurso docente assumido precisa levar em conta um conjunto de ideias e experiências organizadas institucionalmente, capazes de proporcionar um desenvolvimento profissional docente consistente, o que se caracteriza por um processo mediado por espaços institucionais em que redes de interações e mediações possibilitem aos professores refletir, compartilhar e reconstruir experiências e conhecimentos próprios à especificidade da Educação Básica e Superior para desenvolver-se na profissão.

Em consonância com Marcelo Garcia (1999, 2013), Imbernón (2006, 2009, 2010), Zabalza (2004), Pimenta e Anastasiou (2002), Tardiff e Lessard (2005) o desenvolvimento profissional docente constitui um processo amplo, multifacetado, compreendendo questões de ordem teórica, profissional, atitudinal, valorativa, sociocultural e política.

Assim, é necessário que seja levado em conta o intercâmbio de conhecimentos, a criação de redes de relações capazes de proporcionar, tanto a professores quanto a estudantes, a construção de conhecimentos e competências, voltados para o ensino de matemática.

Ao se pensar em desenvolvimento profissional docente, é importante relacioná-lo com a aprendizagem docente e refletir sobre como se aprende a ser professor e como se ensina esse ofício. Em ambos os casos, instaura-se um processo de natureza inter/intrapessoal, abrangendo a apropriação de procedimentos, conhecimentos, competências, saberes e fazeres que estejam vinculados à realidade concreta da atividade docente em seus diversos campos de atuação e em seus respectivos domínios, bem como em um contexto institucional peculiar. Nessa perspectiva, a aprendizagem docente é condição para o desenvolvimento e concretização do fazer-se professor ao longo da trajetória pessoal e profissional, implicando a possibilidade de o docente estar aberto e receptivo a aceitar novas formas de se constituir com conhecimentos tanto específicos quanto pedagógicos e profissionais.

Os processos de aprender e ensinar a ser professor vinculam-se também à dinâmica entre conhecimento científico, saber acadêmico e saber escolar. No caso do conhecimento científico, o processo exige uma série de critérios de rigor em que precisam ser consideradas as condições espaço-temporais concretas, compreendendo uma situação problematizadora, uma relação próxima entre o sujeito e o objeto, uma explicitação dos processos e das formas de construção lógica do conhecimento, abarcando tanto as questões quanto as respostas decorrentes dessa construção. As respostas ou resultados obtidos desses processos passam também por critérios coletivos de verificação (GAMBOA, 2009).

Por outro lado, para esse mesmo autor, a acumulação de respostas sobre um determinado fenômeno, isto é, informações diferentes sobre algo, constituem o mundo dos saberes acadêmicos e escolares. Essas respostas podem ser divulgadas na forma de informações padronizadas e selecionadas, tais como: livros didáticos, esquemas, resumos e fórmulas, e são transmitidas no contexto da organização acadêmica e escolar. Dessa forma, a característica dos saberes é a de se apresentar como respostas, separadas de suas perguntas originárias. Nesse sentido, elas são produtos que não exigem sua relação imediata com o processo da relação pergunta/resposta, inerente ao conhecimento científico. Contudo, tanto os saberes acadêmicos e escolares como o conhecimento científico podem ter origem nos mesmos problemas e partir das mesmas perguntas, mas os primeiros se constituem a partir de respostas/produtos, desvinculados do processo de sua construção. Ademais, em contrapartida do conhecimento científico, os saberes acadêmicos e escolares são construídos na constante busca de convergência entre os interesses da área específica e da área pedagógica.

Nessa dinâmica, entende-se que os saberes acadêmicos e escolares precisam ser problematizados na aula. A problematização está na base da aprendizagem docente e discente, indicando a incompletude do professor e do estudante. Nesse processo, estão

implicados a dinamização dos conhecimentos específicos da área disciplinar e os modos de construção de estratégias pedagógicas para o desenvolvimento de atividades de estudo a serem implementadas, compreendidos como elementos fundantes do processo de aprender a docência e, consequentemente, de construir-se como professor (BOLZAN, ISAIA, 2006; ISAIA, BOLZAN, 2007, 2009).

Pensando no conhecimento científico e no saber acadêmico e escolar, entendem-se como importantes as afirmações de Shulman (1968, 1987, 1989), para quem, o fazer docente envolve três dimensões: conhecimento de conteúdo específico, ou seja, de conceitos básicos de determinada área que implica o modo de entender seu processo de construção (conhecimento específico e saber acadêmico e escolar); conhecimento pedagógico geral, isto é, aquele que, entre outros, vai além, pois inclui os objetivos, as metas e os propósitos educacionais de manejo de classe e interação com os alunos, de estratégias instrucionais, da maneira como os alunos aprendem, de outros conteúdos, de conhecimento curricular; conhecimento pedagógico do conteúdo, que integra o conteúdo tanto específico quanto pedagógico de cunho geral e envolve a forma como o professor concebe os propósitos de ensinar determinada matéria, a relevância do que é aprendido pelos alunos, as possíveis concepções errôneas ou falsas que eles apresentam em relação à matéria, entre outros aspectos.

2.3 Ensinar e aprender matemática

Um dos desafios no processo de ensino e aprendizagem da Matemática é o fato de que o pensamento matemático pressupõe regras sólidas advindas da Lógica. A partir desta concepção muitas práticas de ensino são construídas, tendo como base a estrutura: axiomas, definições, proposições e teoremas. Os professores acreditam que esta estrutura formal do pensamento é suficiente para os alunos entenderem o significado matemático de um conceito. Esta prática traz como consequência a ideia de que basta explicar claramente um conteúdo para os alunos entenderem e, a partir deste entendimento, estabelecer relações. O exemplo mais significativo deste tipo de ensino está no chamado movimento da Matemática Moderna que teve seu auge na década de 1960. A Matemática Moderna levou, por meio dos livros didáticos, esta estrutura de pensamento lógico para a sala de aula. Este pensamento reducionista do trabalho de sala de aula, a partir de uma lógica preestabelecida, tem elevado os índices de reprovação e evasão da escola e ao fracasso da Matemática em todos os níveis de ensino.

O desenvolvimento de pesquisas na área da Educação Matemática tem apresentado, nos últimos anos, caminhos alternativos para as práticas de sala de aula de modo a promover, de fato, a compreensão dos conceitos matemáticos e de suas relações. Entre as diferentes alternativas destaca-se a teoria de Imagem de Conceito e Definição de Conceito de Tall e Vinner (1981). Estes autores defendem que o trabalho de sala de aula não deve seguir apenas a lógica do pensamento estruturado da Matemática, mas deve levar em consideração, também, os processos cognitivos relacionados com a aprendizagem dos alunos. Segundo os autores,

> comparada com outras áreas de atividade humana, Matemática é usualmente considerada como uma ciência de muita precisão em que conceitos podem ser definidos seguramente para fornecer uma base firme para a teoria matemática. As realidades psicológicas são sutilmente diferentes. (TALL; VINNER, 1981, p. 151).

Para os autores, partir-se da definição formal de um determinado conceito matemático, definição esta como concebida pela comunidade de matemáticos profissionais, pode causar um conflito entre a forma como a matemática é estruturada e os processos cognitivos de aquisição de conceitos. Isto é, para a comunidade de matemáticos profissionais, a matemática é uma ciência dedutiva que começa com as noções simples de definições e axiomas e que, a partir dessas noções, os teoremas e demais relações são estabelecidas. Pensar o ensino da Matemática a partir dessa estruturação pode ter sérias consequências para a aprendizagem dos alunos, pois os processos cognitivos de aquisição de um novo conceito não seguem esta lógica. Portanto, é fundamental que professores comecem a pensar sobre uma pedagogia apropriada para o ensino e a aprendizagem da matemática que leve em consideração os processos psicológicos de aquisição de conceitos e de formação do raciocínio lógico.

Esta pedagogia deve levar em consideração a distinção entre o modo como é gerado o conceito na mente humana e a formalização matemática do mesmo. De acordo com Vinner (1991, p. 69), "adquirir um conceito significa formar uma imagem do conceito", ou seja, a formação de conceitos é precedida de uma imagem mental e, após, a formação dessa imagem conceitual é que o indivíduo passa a descrevê-la por meio de palavras e do uso das estruturas matemáticas para então formalizar uma definição do conceito. O autor defende que a formação da imagem do conceito é fundamental para a compreensão dos conceitos matemáticos.

Segundo Tall e Vinner (1981, p. 152), o conceito imagem pode ser definido como aquele que

> [...] descreve toda estrutura cognitiva que está associada ao conceito, inclui todas as imagens mentais e propriedades a elas associadas e os processos. É desenvolvido ao longo dos anos por meio de experiências de todos os tipos, mudando tanto quando o indivíduo encontra novos estímulos quanto quando amadurece.

No processo de construção da imagem do conceito, o indivíduo associa ao conceito algo que o reporte a ele sempre que isso lhe for solicitado, por exemplo, ao necessitar desse conceito, pode lembrar-se de uma expressão, gráfico ou um problema que resolveu por meio dele.

A imagem do conceito é exclusiva de cada indivíduo por estar relacionado com as experiências desse com o ambiente em que vive, uma vez que se trata de impressões e representações visuais que ele tem ao entrar em contato com tal conceito. Já a imagem, apenas, por sua vez, é constantemente alterada por ter relação com as experiências vivenciadas pelo sujeito, isto é, como o cérebro reage de maneira diferente aos diversos estímulos recebidos, logo não há uma única imagem do conceito.

A imagem conceitual precede a formalização do conceito. O conceito é construído passo a passo e por meio de diferentes estratégias. A definição do conceito é uma etapa formal que envolve um conjunto de palavras e estruturas matemáticas. Se de fato a compreensão do conceito existe, a sua definição formal é uma etapa que possui significado.

De acordo com Giraldo, Carvalho e Tall (2002, p.2),

> [...] uma imagem conceitual pode ainda estar associada a uma sentença usada para especificar o conceito em questão, denominada definição conceitual que,

> por sua vez, pode ou não ser coerente com a definição matemática correta, isto é, aquela aceita pela comunidade matemática.

Portanto, a definição de conceito é entendida como uma forma que um determinado indivíduo tem para expressar um conceito por meio de palavras, segundo Tall e Vinner (1981). Essa definição pode estar ou não relacionada com a imagem do conceito, e pode, também, estar ou não em consonância com a definição formal do conceito, ou seja, aquela que se encontra na maioria dos livros didáticos. Os autores afirmam, ainda, que é possível aprender um novo conceito a partir de experiências e usá-los de forma adequada em diferentes contextos sem saber uma definição formal, em outras palavras, a ausência de uma definição formal não impede a aprendizagem de um novo conceito.

Além disso, os autores argumentam que definições formais podem ser memorizadas pelos alunos, mas sem significado, e, no momento em que em uma nova situação o conceito aparece, o indivíduo não consegue evocá-lo. Esta é a base de uma pedagogia de sala de aula cujos resultados são os altos índices de evasão e reprovação e o fraco desempenho nos exames nacionais e internacionais dos alunos da educação básica e superior, conforme dados do IDEB, ENEM, PISA e ENADE.

A pedagogia de sala de aula, tendo como base a lógica da estruturação do pensamento matemático formal, pouco tem contribuído para o desenvolvimento de competências nos alunos, pois o conhecimento é construído passo a passo, levando em consideração as experiências dos mesmos que são ricas de significado para a construção de imagens conceituais.

De acordo com Tall e Vinner (1981), no trabalho de sala de aula, partir-se da ideia de construção de imagens conceituais pode contribuir, de modo significativo, no desenvolvimento de competências matemáticas, especialmente àquelas descritas no documento emitido pelo *National Council of Teacher of Mathematics* (NCTM, 2000). Neste documento, destaca-se a seguinte competência matemática: saber elaborar perguntas e respostas sobre questões matemáticas. Isto, por sua vez, exige uma série de outras competências a ela relacionadas, tais como: raciocínio lógico na abordagem de resolução de problemas e de modelagem matemática; lidar com a linguagem matemática por meio de diferentes representações: simbolismo matemático, formalismo, comunicação oral e escrita; ter conhecimentos básicos sobre conteúdos de álgebra, geometria, números e operações, medidas, análise de dados e probabilidade; e estabelecer conexões entre os mesmos.

Para o desenvolvimento dessas competências, é necessário prever práticas de sala de aula que, de fato, possam ajudar os alunos a compreenderem os conceitos matemáticos e, para isso, é necessário levar em consideração os processos cognitivos de aquisição de conhecimentos e não seguir, apenas, a sequência do pensamento matemático formalmente estruturado.

A aquisição de competências matemáticas para De Corte (2007) e Kilpatrick (2002, apud ALEJO; ESCALANTE, 2012) é possível a partir da criação de um ambiente de sala de aula em que os alunos tenham a oportunidade de aprender Matemática como uma disciplina dinâmica e em constante evolução e não uma ciência ser reduzida à memorização e procedimentos desconectos da realidade.

3 Aportes metodológicos

Para realização deste estudo, optou-se por uma abordagem quanti-qualitativa. Os instrumentos de pesquisa foram enquetes organizadas utilizando-se uma escala do tipo Likert. Ao responderem a estes instrumentos, os sujeitos participantes, num total de 20, especificaram o seu nível de concordância com cada afirmação, optando entre uma série graduada de respostas de uma a dez. A construção das enquetes foi via Grupo Focal, tendo por base Pichon-Rivière (2002).

Para este artigo, selecionaram-se as competências pessoais derivadas do Mestrado, bem como aquelas oriundas das Diretrizes Curriculares Nacionais (DCN) da Licenciatura em Matemática.

Quanto às competências relativas ao Mestrado foram selecionadas as seguintes:

Código	Variavéis
V27_1	Ser capaz de produzir conhecimentos na área do ensino de Matemática
V27_2	Apresentar consistência pedagógica no ensino de Matemática
V27_3	Ser capaz de realizar estudos e pesquisas a partir de conteúdos curriculares de Matemática
V27_4	Saber fazer a integração entre os conteúdos curriculares
V27_5	Saber fazer uso das Tecnologias de Informação e Comunicação (TIC) e dos ambientes virtuais para o ensino de Matemática
V27_6	Conhecer as diversas concepções e modalidades de avaliação da aprendizagem
V27_7	Utilizar técnicas e métodos de pesquisa apropriados ao ensino de Matemática
V27_8	Desenvolver um trabalho colaborativo e interdisciplinar entre a Matemática e as demais ciências

Quadro 1. Domínio pessoal das competências derivadas do Mestrado Profissionalizante em ensino de Matemática. [vector-variavel V27 = (V27_1,..., V27_8)].

Das competências relativas à Licenciatura em Matemática foram selecionadas:

Código	Variavéis
V28_1	Elaborar propostas de ensino-aprendizagem de Matemática para a educação básica
V28_2	Analisar, selecionar e produzir materiais didáticos
V28_3	Analisar criticamente propostas curriculares de Matemática para a educação básica
V28_4	Desenvolver estratégias de ensino que favoreçam a criatividade, a autonomia e a flexibilidade do pensamento matemático dos educandos, buscando trabalhar com mais ênfase nos conceitos do que nas técnicas, fórmulas e algoritmos
V28_5	Perceber a prática docente de Matemática como um processo dinâmico, carregado de incertezas e conflitos, um espaço de criação e reflexão, onde novos conhecimentos são gerados e modificados continuamente
V28_6	Contribuir para a realização de projetos coletivos dentro da escola básica
V28_7	Utilizar o computador como instrumento de trabalho, incentivando-se sua utilização para o ensino de Matemática, em especial para a formulação e solução de problemas
V28_8	Utilizar vários recursos mediante as TIC, que possam contribuir para o ensino de Matemática

Quadro 2. Domínio pessoal das competências derivadas da Licenciatura em Matemática [vector-variável V28=_1,...V28_8)].

Os dados foram analisados no quadro teórico da Análise Estatística Implicativa ASI. Atualmente, a ASI designa um campo teórico centrado sobre o conceito de implicação

estatística ou mais precisamente sobre o conceito de quase implicação para distinguir este da implicação lógica dos domínios da lógica e da matemática. A modelização e o estudo deste conceito de quase implicação enquanto objeto matemático, no campo das probabilidades e da estatística, conduz a construir as ferramentas teóricas instrumentado um novo método de análise de dados. O software designado pelo acrônimo CHIC (Classificação Hierárquica, Implicativa e Coesiva) cuja construção foi iniciada por Régis Gras em 1985, foi aperfeiçoado em seguida nas teses de Saddo Ag Almouloud (1992) e de Harrison Ratsimba Rajohn (1992). Sua manutenção e sua atualização são garantidas por Raphaël Couturier (2008) em resposta às demandas de pesquisadores e aos desenvolvimentos teóricos e das expectativas dos utilizadores.

4 Análise e discussão dos resultados

Os resultados serão apresentados em dois momentos. No primeiro, faremos uma análise geral do nível de concordância dos participantes a respeito do seu processo formativo; posteriormente, as análises fundamentadas no quadro da análise estatística implicativa - ASI. Antes de discutir os resultados, discorremos brevemente sobre o perfil dos sujeitos.

Como já destacado, os sujeitos participantes foram (20) envolvendo egressos que se formaram desde a primeira turma que se iniciou em 2004 e estudantes que estão ainda desenvolvendo suas atividades no curso até o ano de 2011.

No que se refere ao gênero, observamos que 70% dos participantes são do gênero feminino e apenas 30% masculino, o que confirma a tese de feminização do magistério.

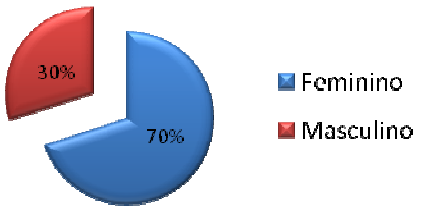


Gráfico 1 – Sexo. (Fonte: elaboração dos autores)

No que tange a idade dos participantes, grande parte (35%) disseram ter entre 29 e 33 anos, 30% entre 36 e 44 anos, 20% entre 47 a 55 anos e 15% 23 a 27.

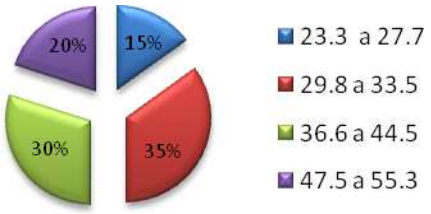


Gráfico 2 – Idade. **(**Fonte: elaboração dos autores)

Questionados sobre o interesse de mudar de nível de ensino depois de ter concluído o mestrado, a maioria (60%) disseram que sim e 40% afirmaram que não. Dessa forma, temos um número significativo de professores da matemática, possivelmente, insatisfeitos com a modalidade de ensino em que estão inseridos.

Gráfico 3 – Intenção de mudar de nível de ensino. (Fonte: elaboração dos autores)

4.1 Competências derivadas do Mestrado Profissionalizante e da Licenciatura em Matemática

Constatou-se que os sujeitos optaram por escolher parâmetros elevados da escala de aferição situando-se, preferencialmente, entre os graus 8, 9 em relação às competências apresentadas. Com relação às competências do mestrado, as alternativas (a) e (c) (d) (g) evidenciam o entendimento dos sujeitos quanto à importância da área de conhecimento específico. Cabe dizer que não é possível pensar o processo de ensinar e aprender desvinculado do domínio do campo específico dos conteúdos a serem desenvolvidos, conjuntamente, por professores e alunos. Corroboram com esta percepção as competências (III) e (V) selecionadas em relação à Licenciatura em Matemática. As alternativas escolhidas permitem inferir a preocupação do curso em possibilitar instrumental de pesquisa, tendo por mote conteúdos curriculares, o que é complementado pela possibilidade do estudante saber fazer a integração entre estes conteúdos. Tal competência se torna possível uma vez que os estudantes podem desenvolver seu espírito investigativo.

Em termos de conteúdo específico, é possível salientar que, muito mais do que os próprios conteúdos, é fundamental que os docentes respeitem, conheçam e compreendam o caminho lógico de sua construção. Assim, a aula poderá deixar de ser um espaço apenas de transmissão mecânica e fragmentada de conhecimentos específicos (acadêmicos e escolares) para instaurar-se como um lugar que possibilita ao aluno uma compreensão genuína, que o tornará capaz de aplicá-los a novas situações. Importante, ainda, é que os professores tenham a consciência de que é possível aprender um novo conceito a partir de experiências e usá-lo de forma adequada em diferentes contextos sem saber uma definição formal, uma vez que a ausência de uma definição formal não impede a aprendizagem de um novo conceito.

Em relação às competências do mestrado, a alternativa (h) mostra a importância do trabalho colaborativo e interdisciplinar entre a Matemática e as demais ciências. É importante salientar que a escola caminha na direção da cooperação profissional em virtude de novas demandas relacionadas com as dificuldades dos alunos, sejam elas comportamentais ou cognitivas que requerem, para superação, o trabalho integrado de todos os setores, desde a direção pedagógica e administrativa com a participação de psicólogos, professores, assistentes sociais, pais etc.; e a divisão do trabalho pedagógico

com a intervenção de profissionais de diferentes áreas do saber, para dar sentido aos conteúdos específicos que são trabalhados na sala de aula.

Um determinado conteúdo específico, por si só, em geral, não tem sentido para o aluno, pois, muitas vezes, a gênese desse conteúdo não está na matemática, mas em outras ciências. Dessa forma, é preciso explicitar a relação entre os saberes com o intuito de suscitar a curiosidade, o desejo de aprender e, portanto, nessa óptica, o trabalho interdisciplinar é de fundamental importância, uma vez que demanda uma equipe que planeja e executa um projeto com objetivos comuns.

Nesta direção, encontramos a mesma preocupação em relação aos licenciandos em Matemática, competência (VI). O trabalho colaborativo e interdisciplinar, apesar de ser uma competência a ser desenvolvida, pouco é enfatizada, pois, geralmente, o que predomina, no trabalho junto aos cursos de formação de professores, é o individualismo profissional. Contudo, contrariamente desta realidade egocêntrica, o curso de mestrado tem possibilitado a integração de conteúdos com outras ciências e um trabalho colaborativo em equipes, envolvendo diferentes metodologias de ensino e aprendizagem. Assim, dinâmica de trabalho implementada no curso de mestrado tem permitido ao grupo de alunos realizar seus projetos de forma colaborativa e interdisciplinar. As atividades curriculares desenvolvidas no curso têm contribuído, de fato, com o aperfeiçoamento do trabalho de sala de aula integrando conteúdos da área da Física, da Matemática e das demais ciências sem abdicar do domínio pedagógico e profissional.

As competências de domínio pedagógico explicitadas nas alternativas (b), (d), (e), (f) (g), (h); (I), (II), (IV), (VI) (VII), (VIII) indicam que as atividades curriculares do mestrado levam em conta as DNC de ensino de Matemática (2001). Esta constatação referenda o caráter profissionalizante do curso, voltado para a formação de professores. Tanto as competências do Curso quanto a DCN, indicam o grau de importância que atribuem às TIC e aos ambientes virtuais para o ensino de Matemática, complementando a visão pedagógica inovadora do mestrado. Portanto, o curso, ao instrumentalizar seus estudantes na dimensão pedagógica, cumpre a função de um mestrado profissional.

As competências de ser professor evidenciam a dimensão pedagógica da docência. Nesse sentido, tais competências ratificam sua finalidade compreendendo as formas de conceber e desenvolver a docência, bem como a organização de estratégias pedagógicas, levando em conta a transposição dos conteúdos específicos de um domínio para sua efetiva compreensão. Dessa forma, por consequência, haverá a aplicação desses conhecimentos por parte dos alunos, que transformarão em instrumentos internos capazes de mediar a construção de seu processo formativo. Juntamente a essas competências, tem-se a dimensão profissional da docência que abarca um repertório de conhecimentos, saberes e fazeres advindos das áreas específicas de conhecimento, englobando o acadêmico e o escolar, da área pedagógica e da área de experiência docente. Este conjunto de interações na área da docência encontra, no mestrado, um espaço formativo privilegiado, pois permite que discentes e docentes interatuem na busca de um processo educativo mais qualificado.

5 Reflexões

O cruzamento entre as competências desenvolvidas pelo mestrado e as indicadas pelas DCN de Matemática demonstra que os sujeitos participantes foram coerentes em assinalar o grau de importância de cada uma, evidenciando correlação entre elas.

Em termos avaliativos iniciais, é possível depreender que as atividades curriculares desenvolvidas no mestrado vêm contribuindo com o aperfeiçoamento do domínio pedagógico e profissional de seus estudantes sem esquecer a importância do conhecimento do domínio específico para o trabalho pedagógico a ser desenvolvido na sala de aula.

Nesse sentido, o processo formativo continuado tem sido incrementado ao longo do mestrado, o que pode ser evidenciado nos produtos pedagógicos desenvolvidos pelos mestrandos e que se encontra na página do curso.

Em relação à competência do domínio do saber específico, lançamos um olhar sobre a construção de imagens conceituais para chegar à definição do conceito, conforme as ideias preconizadas por Tall e Vinner (1981). É importante também destacar que a imagem conceitual é individual e modifica-se com o tempo. Assim, a forma como cada aluno constrói uma imagem para um dado conceito é particular e está sempre impregnada de valores, de referências e de competências que são pessoais e variam de indivíduo a indivíduo. Por outro lado, concordamos com os autores quando salientam a importância de considerar os aspectos procedimentais no trabalho de sala de aula, uma vez que, em suas pesquisas, eles constataram que os alunos demonstraram um desempenho bem mais satisfatório e bem diferente daquele quando consideraram apenas os aspectos conceituais formais.

Com a realização da pesquisa, observou-se também que, embora haja uma recomendação explícita sobre o uso das TIC nas DCN e no projeto pedagógico do curso de Mestrado (competências VIII,), os professores se sentem inseguros frente a tantas novidades, para as quais não estão preparados. Muitos percebem o esforço do curso de Mestrado para integrar as TIC em suas disciplinas e reconhecem que o uso das tecnologias na sala de aula pode representar um auxílio para a melhoria e qualificação de sua atividade profissional.

Sendo uma das finalidades do ensino da Matemática contribuir para o desenvolvimento de indivíduos participantes e cidadãos ativos na vida social, os alunos devem ter experiências e saber lidar com situações reais usando conceitos matemáticos na interpretação e modelagem de situações reais. Isso se dá, de acordo com Kilpatrick (2002, apud ALEJO; ESCALANTE, 2012, p. 51), à medida que tenham a competência de transferir para uma nova situação os conceitos aprendidos. Por outro lado, essas experiências propiciam a criação de imagens conceituais que favorecem a compreensão dos conceitos matemáticos.

Tendo em vista as competências escolhidas pelos participantes em grau mais elevado da escala de importância, compreende-se que o ensinar não pode ser confundido com o repasse simples de conteúdos prontos. Precisa envolver um processo intencional e sistematizado de organizar os conhecimentos, saberes e fazeres próprios à determinada área de conhecimento, e de oferecer ajudas capazes de auxiliar os alunos a construírem suas próprias estratégias de apropriação, em direção à sua autonomia formativa.

Em um tempo em que a Educação Básica e os cursos de licenciatura estão em crise, necessitando ser reconsiderados, entende-se que pesquisas desta natureza podem contribuir para repensar e trazer novas estratégias educativas à formação de professores.

5.1 Resultados a partir do tratamento oriundo do software CHIC

A tabela explorada pelo softawe CHIC foi estruturada por 20 linhas (que retratam o número de sujeitos participantes do estudo) e 24 colunas (que mostram as variáveis). Consideramos os componentes dos vetores V27, V28 como variáveis modais, ou seja, tomando os seus valores no conjunto {0,0.1, 0.2, ...,1}. São as variáveis principais. Sendo consideradas variáveis suplementares, os componentes dos vetores Sexo; Idade; Pretensão de mudança de nível de ensino depois de ter concluído o mestrado.

Vetor-Variável	Tipo ASI
Domínio pessoal das competências derivadas do Mestrado Profissionalizante em ensino de Matemática V27=(V27_1; V27_2; V27_3; V27_4; V27_5; V27_6; V27_7; V27_8)	Modais
Domínio pessoal das competências derivadas da Licenciatura em Matemática V28=(V28_1; V28_2; V28_3; V28_4; V28_5; V28_6; V28_7; V28_8)	Modais
Sexo = (SexoF; SexoM)	Binárias
Idade = (Idade 1_ 23.3 - 27.7; Idade 2_ 29.8 - 33.5; Idade 3_ 36.6 – 44.5; Idade 4_ 47.5 - 55.3).	Binárias
Pretensão de mudar de nível de ensino depois de ter concluído o mestrado = (SimMudar; NãoMudar)	Binárias

Quadro 3. Variáveis analisadas. (Fonte: elaboração dos autores)

No primeiro momento analisaremos os dados com base na árvore de similaridade; no segundo, na árvore coesitiva; e, no terceiro, no gráfico implicativo. O modelo matemático escolhido no quadro dos tratamentos ASI com o software CHIC é o modelo baseado na lei de probabilidade binominal.

5.1.1 Análises do gráfico de similaridade

Conforme pode ser observado no gráfico 4, a árvore de similaridade é formada por duas grandes classes estruturadas pelas variáveis componentes dos vetores V27 e V28. **A classe 1** se refere ao Mestrado Profissional em ensino de Matemática; a **classe 2** à Licenciatura em Matemática.

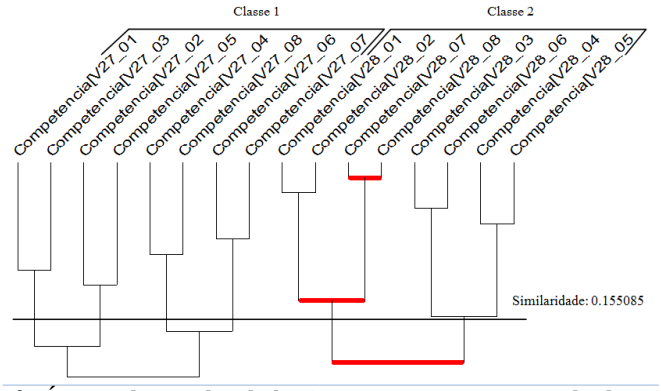
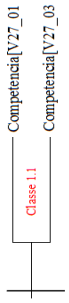
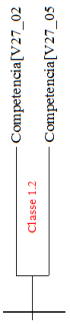


Gráfico 4: Grafo Árvore de Similaridade. (Fonte: tratamento oriundo do software CHIC)

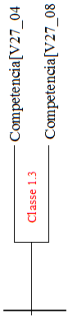
Como mostra o grafo, cada uma das classes é estruturada por subclasses de variáveis. Na **classe 1** podemos localizar 4 (quatro) subclasses com similaridade (0155085) entre 8 variáveis:



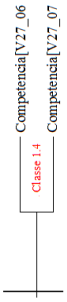
Na **classe 1.1** observamos uma relação de similaridade entre as competências "ser capaz de produzir conhecimentos na área do ensino de Matemática" (V27_01) e "ser capaz de realizar estudos e pesquisas a partir de conteúdos curriculares de Matemática" (V27_03). Sobre isso, destaca-se a coerência na avaliação do domínio das referidas competências, uma vez que estas são complementares. Para a produção do conhecimento, faz-se fundamental o ato de estudar e/ou pesquisar.



Evidenciando novamente a coerência das percepções dos participantes, na **classe 1.2** surge uma relação de similaridade entre as competências "apresentar consistência pedagógica no ensino de Matemática" (V27_02) e "saber fazer uso das Tecnologias de Informação e Comunicação (TIC) e dos ambientes virtuais para o ensino de Matemática" (V27_05).



Em relação à **classe 1.3** podemos verificar similaridade no grau de domínio pessoal das competências "saber fazer a integração entre os conteúdos curriculares" (V27_04) e "desenvolver um trabalho colaborativo e interdisciplinar entre a Matemática e as demais Ciências" (V27_08). Outra vez, observamos que os participantes foram coerentes em suas avaliações.



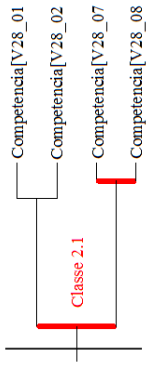
Na **classe 1.4** a similaridade ocorre entre as competências "conhecer as diversas concepções e modalidades de avaliação da aprendizagem" (V27_06) e "utilizar técnicas e métodos de pesquisa apropriados ao ensino de Matemática" (V27_07).

Em relação às contribuições dos sujeitos (através das variáveis suplementares), observamos que a variável que mais contribui a esta classe é **Idade1** (de 23 a 27 anos – grupo de participantes mais jovens), com um risco de: 0. Em segundo lugar, a variável **Idade 3** (de 36 a 44 anos), com risco de 0.118.

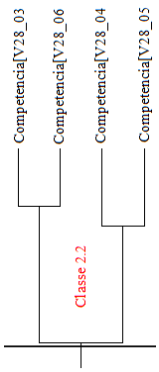
A variável Idade1 contribui a esta classe com um risco de : 0 A variável Idade2 contribui a esta classe com um risco de : 0.874 A variável Idade3 contribui a esta classe com um risco de : 0.118 A variável Idade4 contribui a esta classe com um risco de : 0.24 A variável V19_SimMudar contribui a esta classe com um risco de : 0.493 A variável V19_NaoNR_Mudar contribui a esta classe com um risco de : 0.255 A variável SexoF contribui a esta classe com um risco de : 0.355 A variável SexoM contribui a esta classe com um risco de : 0.42 A variável que contribui mais a esta classe é Idade1 com um risco de : 0

Quadro 4: Contribuição à classe 1. Fonte: tratamento oriundo do software CHIC

A segunda classe (competências da Licenciatura em Matemática) é formada por apenas 2 (duas) subclasses:



Observando o gráfico 4, a **subclasse 2.1** é a que apresenta maior índice de similaridade, uma vez que apresenta nós significativos. No primeiro nível, observamos uma aproximação entre as competências "utilizar o computador como instrumento de trabalho, incentivando-se sua utilização para o ensino de Matemática, em especial para a formulação e solução de problemas" (V28_07) e "utilizar vários recursos mediante as TIC, que possam contribuir para o ensino de Matemática" (V28_08). A relação de similaridade ainda pode ser observada entre as competências "elaborar propostas de ensino-aprendizagem de Matemática para a educação básica" (V8_01) e "analisar, selecionar e produzir materiais didáticos" (V8_02).



Na **subclasse 2.2** surge uma aproximação entre as competências "analisar criticamente propostas curriculares de Matemática para a educação básica" (V28_03) e "contribuir para a realização de projetos coletivos dentro da escola básica" (V28_06); assim como, entre "desenvolver estratégias de ensino que favoreçam a criatividade, a autonomia e a flexibilidade do pensamento matemático dos educandos, buscando trabalhar com mais ênfase nos conceitos do que nas técnicas, fórmulas e algoritmos" (V28_04) e "perceber a prática docente de Matemática como um processo dinâmico, carregado de incertezas e conflitos, um espaço de criação e reflexão, onde novos conhecimentos são gerados e modificados continuamente"(V28_05).

A variável que mais contribuiu para a formação da classe 2 é **Idade 1** (23 a 27 anos), assim como **a Idade 2** (29 a 33 anos), ambas com risco de: 0. Nesse sentido, essa classe foi construída sob maior influência dos participantes mais jovens.

A variável Idade1 contribui a esta classe com um risco de : 0 A variável Idade2 contribui a esta classe com um risco de : 0 A variável Idade3 contribui a esta classe com um risco de : 0.531 A variável Idade4 contribui a esta classe com um risco de : 0.656 A variável V19_SimMudar contribui a esta classe com um risco de : 0.282 A variável V19_NaoNR_Mudar contribui a esta classe com um risco de : 0.43 A variável SexoF contribui a esta classe com um risco de : 0.229 A variável SexoM contribui a esta classe com um risco de : 0.531 A variável que contribui mais a esta classe é Idade1 com um risco de : 0

Quadro 5: Contribuição à classe 2. Fonte: tratamento oriundo do software CHIC

Diante do que foi apresentado, a árvore de similaridade confirma os achados de que os participantes foram coerentes em assinalar o nível de importância e apropriação das competências. Todavia, apesar desse resultado positivo, a árvore aponta uma separação entre às competências decorrentes da Licenciatura e do Mestrado em Matemática. Sobre isso, entendemos que se trata de formações com perspectivas diferentes, uma vez que a primeira objetiva tratar da preparação de base do profissional, enquanto que a segunda o aprofundamento e aperfeiçoamento dos conhecimentos já adquiridos pelo professor. Porém, trata-se de formações complementares e/ou interligadas, sendo que ambas buscam a preparação dos profissionais para o ensino da Matemática.

Outra análise importante é a respeito da contribuição das variáveis suplementares. Como vimos ambas as classes foram construídas sob a maior influência da variável idade (sobretudo participantes mais jovens). As demais (gênero; pretensão de mudar de nível de ensino) não apresentaram interferências significativas.

5.1.2 Análises do gráfico coesitivo

A seguir apresentaremos o gráfico 5 (árvore hierárquica coesitiva), que trata do índice de coesão entre variáveis.

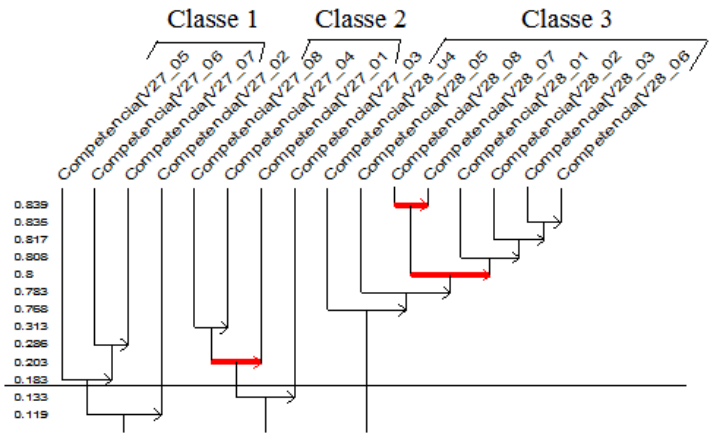
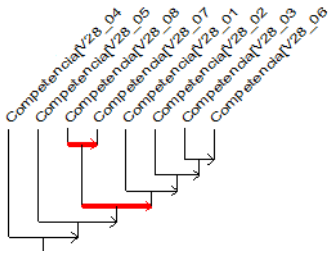


Gráfico 5: Grafo Árvore Coesitiva. Fonte: tratamento oriundo do software CHIC

Com similaridade de 0,183, a árvore coesitiva é estruturada em três classes. Apresentaremos cada uma delas observando a ordem hierárquica de coesão. Assim, começaremos pela classe 3, posteriormente classe 2 e classe 1.



Como podemos observamos, as variáveis componentes do vetor 28 (competências da Licenciatura em Matemática) formou uma única classe com relação estatística bastante significativa. As competências do lado direito da árvore são as menos dominadas e do lado esquerdo as mais dominadas. Diante disso, podemos assinalar que os participantes acreditam que a realização de uma prática docente focada mais nos conceitos da matemática do que nas fórmulas; desenvolver estratégias um ensino que favoreçam a criatividade, a autonomia e a flexibilidade do pensamento matemático do educando (V28_04),

envolve, primeiramente, o domínio de determinadas competências como: a percepção da prática docente como um processo dinâmico, onde novos conhecimentos são construídos e reconstruídos (V28_05); e, utilizar vários recursos mediante as TIC que possam contribuir para o ensino da Matemática (V28_8).

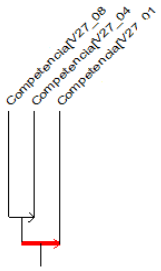
Neste contexto, concordamos com os participantes, uma vez que é importante o professor compreender que o espaço da sala de aula não deverá ser restringido ao ato de ensinar algo, mas, sobretudo de desconstrução, reconstrução e construção do conhecimento. Quando o docente tem esse entendimento, o desenvolvimento de outras competências ocorre de forma mais efetiva.

A partir desse gráfico, é possível observar ainda que as competências dos vetores V28_01, V28_02 e V7_03 apresentam um forte índice de coesão, indicando que a elaboração de uma proposta de ensino de Matemática envolve saber analisar criticamente as propostas curriculares de Matemática; analisar, selecionar e também produzir materiais didáticos.

A variável que mais contribuiu para a formação dessa classe é **Idade 1** (23 a 27 anos), com risco de: 0

A variável Idade1 contribui a esta classe com um risco de : 0
A variável Idade2 contribui a esta classe com um risco de : 0.321
A variável Idade3 contribui a esta classe com um risco de : 0.377
A variável Idade4 contribui a esta classe com um risco de : 0.522
A variável V19_SimMudar contribui a esta classe com um risco de : 0.443
A variável V19_NaoNR_Mudar contribui a esta classe com um risco de : 0.272
A variável SexoF contribui a esta classe com um risco de : 0.357
A variável SexoM contribui a esta classe com um risco de : 0.377
A variável que contribui mais a esta classe é Idade1 com um risco de : 0

Quadro 6: Contribuição à classe 3. Fonte: tratamento oriundo do software CHIC



As variáveis dos componentes do vetor 27 (competências do Mestrado) formaram duas classes. A classe 2 (de maior índice de coesão neste vetor) mostra que os sujeitos deste estudo compreendem que o desenvolvimento de um trabalho interdisciplinar entre a Matemática e as demais ciências (V27_8), envolve a principio, o saber fazer a integração dos conteúdos curriculares (V27_04); e, ser capaz de produzir conhecimentos na área de Matemática (V27_01).

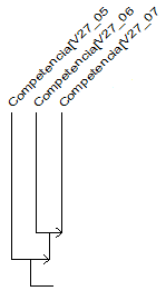
Nesse sentido, os participantes novamente apresentaram uma percepção coerente. A realização de um trabalho interdisciplinar entre Matemática e demais disciplinas exige do professor o conhecimento dos princípios teóricos e metodológicos que envolvem a atividade de integração entre os conteúdos.

A variável que contribui mais a esta classe é **Idade 4** (47 a 55 anos), com risco de: 0. Partindo do pressuposto de que a compreensão dos desafios que envolvem a interdisciplinaridade no ensino é adquirida, especialmente por meio da prática e/ou da experiência no espaço escolar, a participação de professores com maior faixa etária

nesta classe pode ser justificada, haja vista que possivelmente esses professores possuem maior tempo de exercício da profissão.

A variável Idade1 contribui a esta classe com um risco de : 0.896 A variável Idade2 contribui a esta classe com um risco de : 0.21 A variável Idade3 contribui a esta classe com um risco de : 0.262 A variável Idade4 contribui a esta classe com um risco de : 0 A variável V19_SimMudar contribui a esta classe com um risco de : 0.558 A variável V19_NaoNR_Mudar contribui a esta classe com um risco de : 0.168 A variável SexoF contribui a esta classe com um risco de : 0.198 A variável SexoM contribui a esta classe com um risco de : 0.655 A variável que contribui mais a esta classe é Idade4 com um risco de : 0

Quadro 7: Contribuição à classe 2. Fonte: tratamento oriundo do software CHIC



Na classe 1 a relação coesitiva entre as variáveis componentes dos vetores V7_05, V27_06 e V27_07 mostra que o saber fazer uso das TIC e dos ambientes virtuais para o ensino da Matemática (V27_05) envolve o conhecimento das diversas concepções e modalidades de avaliação da aprendizagem (V27_06); e, sobretudo o domínio da competência utilizar técnicas e métodos de pesquisa apropriados ao ensino de Matemática (V27_07). A variável que contribui mais a esta classe é **Idade1** (23 a 27anos), **Idade 3** (36 a 44 anos) e **Idade 4** (47 a 55 anos), ambas com um risco de : 0.

A variável Idade1 contribui a esta classe com um risco de : 0 A variável Idade2 contribui a esta classe com um risco de : 0.996 A variável Idade3 contribui a esta classe com um risco de : 0 A variável Idade4 contribui a esta classe com um risco de : 0 A variável V19_SimMudar contribui a esta classe com um risco de : 0.493 A variável V19_NaoNR_Mudar contribui a esta classe com um risco de : 0.255 A variável SexoF contribui a esta classe com um risco de : 0.355 A variável SexoM contribui a esta classe com um risco de : 0.42 A variável que contribui mais a esta classe é Idade1 com um risco de : 0

Quadro 8: Contribuição à classe 1. Fonte: tratamento oriundo do software CHIC

5.1.3 Análises do gráfico implicativo

A seguir, apresentamos o grafo implicativo (gráfico 6) que se refere ao conjunto das variáveis componentes do vetor 27 (Mestrado Profissional em ensino de Matemática), com nível de confiança 0.60.

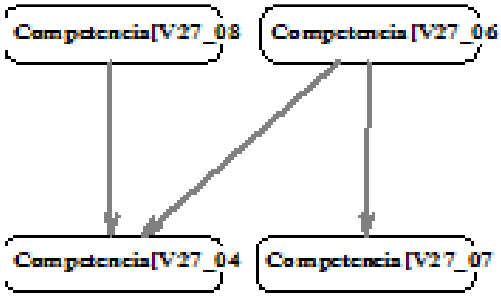


Gráfico 6 – Grafo implicativo. Fonte: tratamento oriundo do software CHIC

Tendo como objetivo observar as relações de quase implicação, nota-se que das 8 variáveis que formam o vetor V27, 4 (quatro) ficaram retidas. O grafo é constituído de três caminhos, sendo que as variáveis **V27_04** e **V27_07** correspondem às competências mais dominadas pelos participantes, entre os quatros em destaque.

Caminho 1: [V7_08] ⟶ [V27_04] sinaliza que o ato de "desenvolver um trabalho colaborativo e interdisciplinar entre Matemática e demais ciências" é consequência do alcance da competência "saber fazer a integração entre os conteúdos curriculares". Sobre isso, vale destacar que está mesma análise foi também realizada por meio do grafo coesitivo. A variável mais típica a este caminho é **Idade 3** (36-44), com um risco de: 0.0277.

Caminho 2: [V7_06] ⟶ [V27_04] mostra que o domínio da competência "saber integrar a Matemática com as demais ciências" é também resultado da apropriação da competência "conhecer as diversas concepções e modalidades de avaliação da aprendizagem". A variável mais típica a este caminho é também **Idade 3** (36-44), com risco de: 0.

Caminho 3: [V7_06] ⟶ [V27_07] aponta que "utilizar técnicas e métodos de pesquisa apropriados ao ensino de Matemática" está ligado ao saber "conhecer as diversas concepções e modalidades de avaliação da aprendizagem". A variável mais típica a este caminho é **Idade 3** (36 – 44 anos), com risco de: 0.

A variável Idade1 contribui a este caminho com um risco de : 0.575 A variável Idade2 contribui a este caminho com um risco de : 0.608 A variável Idade3 contribui a este caminho com um risco de : 0.0277 A variável Idade4 contribui a este caminho com um risco de : 0.391 A variável V19_SimMudar contribui a este caminho com um risco de : 0.304 A variável V19_NaoNR_Mudar contribui a este caminho com um risco de : 0.477 A variável SexoF contribui a este caminho com um risco de : 0.337 A variável SexoM contribui a este caminho com um risco de : 0.442 A variável que contribui mais a este caminho é Idade3 com um risco de : 0.0277

Quadro 9: Contribuições dos sujeitos à construção do Caminho 1, vetor V27. Fonte: tratamento oriundo do software CHIC

A variável Idade1 contribui a este caminho com um risco de : 0.992
A variável Idade2 contribui a este caminho com um risco de : 0.21
A variável Idade3 contribui a este caminho com um risco de : 0
A variável Idade4 contribui a este caminho com um risco de : 0
A variável V19_SimMudar contribui a este caminho com um risco de : 0.795
A variável V19_NaoNR_Mudar contribui a este caminho com um risco de : 0
A variável SexoF contribui a este caminho com um risco de : 0.448
A variável SexoM contribui a este caminho com um risco de : 0.262

A variável que contribui mais a este caminho é Idade3 com um risco de : 0

Quadro 10: Contribuições dos sujeitos à construção do Caminho 2, vetor V27. Fonte: tratamento oriundo do software CHIC.

A variável Idade1 contribui a este caminho com um risco de : 0.343
A variável Idade2 contribui a este caminho com um risco de : 0.971
A variável Idade3 contribui a este caminho com um risco de : 0
A variável Idade4 contribui a este caminho com um risco de : 0
A variável V19_SimMudar contribui a este caminho com um risco de : 0.724
A variável V19_NaoNR_Mudar contribui a este caminho com um risco de : 0.0576
A variável SexoF contribui a este caminho com um risco de : 0.161
A variável SexoM contribui a este caminho com um risco de : 0.744

A variável que contribui mais a este caminho é Idade3 com um risco de : 0

Quadro 11: Contribuições dos sujeitos à construção do Caminho 3, vetor V27. Fonte: tratamento oriundo do software CHIC

Em um segundo tratamento foi obtido o grafo implicativo que é estruturado pelas variáveis do vetor 28 (Licenciatura em Matemática), com nível de confiança 0.80. Como pode ser observado no gráfico 7, nenhuma das 8 variáveis ficaram retidas. Logo abaixo do citado gráfico apresentamos a descrição das competências.

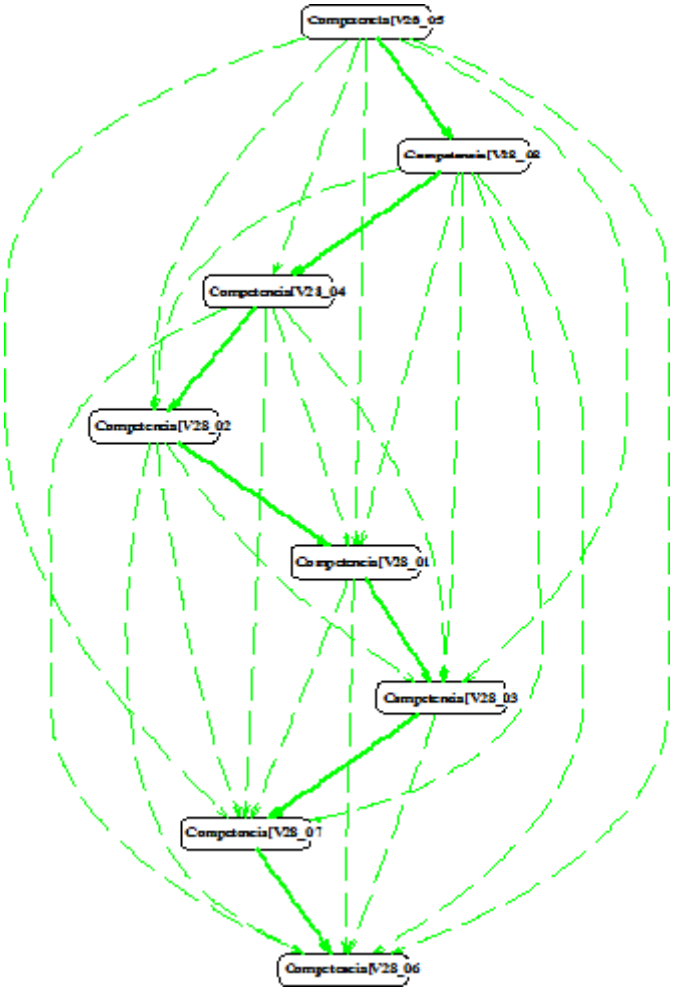


Gráfico 7 – Grafo das propensões com os fechamentos transitivos. Fonte: tratamento oriundo do software CHIC

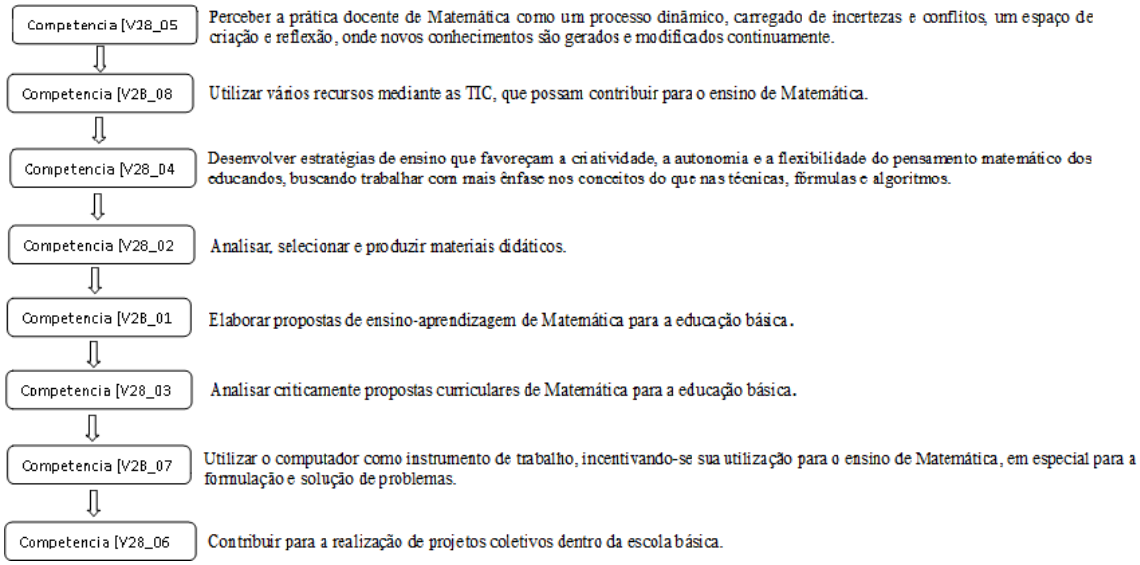


Figura 1: Conteúdo do vetor V28

Como podemos notar, as relações são transitivamente fechadas. O subconjunto, formado pelos 8 variáveis é totalmente ordenado pela relação de propensão. Nas duas

extremidades do grafo temos as competências de maior e menor domínio. A competência mais dominada (parte de baixo do grafo) é “Contribuir para a realização de projetos coletivos dentro da escola básica” (V28_06); e, a menos dominada (parte de cima do grafo) é “Perceber a prática docente de Matemática como um processo dinâmico, carregado de incertezas e conflitos, um espaço de criação e reflexão, onde novos conhecimentos são gerados e modificados continuamente” (V28_05). Nesse sentido, é importante ressaltar a importância dessa última competência, haja vista que o ensino da Matemática não pode ser pautado na transferência de conteúdos, mas na construção e recriação do conhecimento.

Entre as competências de alto e baixo grau de domínio pessoal, encontramos as competências [V28_08, [V28_04, [V28_02, [V28_01, [V28_03, [V28_07 que contemplam o uso de instrumentos didático - pedagógicos, bem como a necessidade de uma prática de ensino crítica da Matemática.

Vale destacar que a variável mais típica a este caminho é **Idade1,** 23 a 27 anos**,** assim como **Idade 2,** ambas com um risco: 0

A variável Idade1 é típica a este caminho com um risco de : 0
A variável Idade2 é típica a este caminho com um risco de : 0
A variável Idade3 é típica a este caminho com um risco de : 0.531
A variável Idade4 é típica a este caminho com um risco de : 0.656
A variável V19_SimMudar é típica a este caminho com um risco de : 0.282
A variável V19_NaoNR_Mudar é típica a este caminho com um risco de : 0.43
A variável SexoF é típica a este caminho com um risco de : 0.229
A variável SexoM é típica a este caminho com um risco de : 0.531

A variável mais típica a este caminho é Idade1 com um risco de : 0

Quadro 12: Contribuições dos sujeitos à construção do Caminho do vetor V28. Fonte: tratamento oriundo do software CHIC

6 Considerações finais

Observamos que os participantes selecionaram parâmetros elevados da escala de aferição de domínio das competências relativas tanto ao Mestrado quanto à Licenciatura em Matemática, na maioria das vezes, entre os graus 8 e 9. Tal resultado é positivo, uma vez que revela que o curso de mestrado tem cumprido com seus objetivos. Ao realizar um cruzamento das competências do Mestrado e da Licenciatura, observamos parâmetros de avaliação mais elevados em relação às competências da primeira. Tendo em vista que os participantes são estudantes ou egresso do Mestrado, possivelmente estes têm maiores lembranças das competências trabalhadas nessa formação. Em relação aos grafos (similaridade, coesitiva e de implicação), notamos a coerência das percepções dos participantes, porém revelamos dados que mostram uma separação das competências oriundos do Mestrado e da Licenciatura em Matemática. Tal resultado merece atenção, uma vez que ambas as formações se complementam no processo de preparação do profissional da Matemática; a Licenciatura que cuida da preparação de base do professor; e, o Mestrado profissional com o aprofundamento e aperfeiçoamento dos conhecimentos.

Références

[1] ABRAHAN, A. (org.)(2000). *El enseñante es también una persona*. Barcelona: Gedisa.

[2] ALEJO, V.V., ESCALANTE, C.C. (2012). Developing Mathematical Competences, Learning Linear Equations, Functions and the relation among these Concepts. *Journal of Mathematical Modelling and Application*, 1 (7), 50-57.

[3] BRASIL (2001). *Diretrizes Curriculares Nacionais para os Cursos de Matemática, Bacharelado e Licenciatura*. Conselho Nacional de Educação / Câmara de Educação Superior. Distrito Federal.

[4] ESTEVES, M (2009). Construção e desenvolvimento das competências profissionais dos professores. In: Sísifo, *Revista de ciências da educação*, n.º 8, pp. 37-48.

[5] GAMBOA, S (2010). Práxis Educativa, *Ponta Grossa*, v.4, n.1, p.9-19, jan.-jun, 2009. Disponível em <http://www.periodicos. uepg.br>. Acesso em: jun.

[6] GIRALDO, V.; CARVALHO, L. M. & Tall, D. O (2002). Theoretical-Computational Conflicts and the Concept Image of Derivative. *Proceedings of the BSRLM Conference*. Nottingham, England, 22 (3), 37–42.

[7] IMBERNÓN, F (2006). *Formação docente e profissional. Formar-se para a mudança*. São Paulo: Cortez.

[8] IMBERNÓN, F (2009). Formación e innovación em La universidad del siglo XXI. In: LAMARRA, N. F. *Universidad, sociedad e innovación*. Buenos Aires: EDUNTREF, p. 179-190.

[9] IMBERNÓN, F (2010). *Formação continuada de professores*. Porto Alegre: Artmed.

[10] IMBERNÓN, F (2007). Aprendizagem docente como articuladora da formação e do desenvolvimento profissional dos professores da Educação Superior. IN: ENGERS, M. E.; MOROSINI, M. (orgs.). *Pedagogia Universitária e Aprendizagem.* Porto Alegre: EDIPUCRS, v. 2, p. 153-165.

[11] ISAIA, S. e BOLZAN, D (2009). Trajetórias da docência: articulando estudos sobre os processos formativos e a aprendizagem de ser professor. IN: ISAIA, S. e BOLZAN, D. P. V. (orgs.) *Pedagogia Universitária e Desenvolvimento Profissional, Docente*. Porto Alegre: EDIPUCRS, p. 121-143.

[12] MARCELO GARCIA, C (1999). Formación del professorado para el cambio educativo. Barcelona: EUB.

[13] MARCELO GARCIA, C 2013). *Desarrollo professional docente. Como se aprende a enseñar?* Madrid: NARCEA.

[14] NÓVOA, A (2009). *Professores imagens do futuro presente*. Lisboa: Educa.

[15] NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

[16] PERRENOUD, Philippe (1999). *Construir competências desde a escola.* Porto Alegre: Artes Médicas Sul.

[17] PERRENOUD, Philippe (2000). *Dez novas competências para ensinar*. Porto Alegre: Artes Médicas Sul.

[18] PICHON-RIVIÈRE, E. (1998). *O processo grupal*. São Paulo: Martins Fontes.

[19] PIMENTA, S. G.; ANASTASIOU, L. das G. C. *Docência no Ensino Superior*. Volume I. São Paulo: Cortez, 2002.

[20] RATSIMBA_RAJOHN, H (1992). *Contribution à l'étude de la hiérarchie implicative. Application à l'analyse de la gestion didactique des phénomènes d'ostension et de contradiction*, Thèse de doctorat de l'Université de Rennes 1.

[21] SHULMAN, L (1968). Those who understand: knowledgegrowth in teaching. *Educational Researcher*, v.15, n.2., Washington DC, AERA.

[22] SHULMAN, L (1987). Knowledge and teaching: the foundation of the new reform. *Harvard Educational Review*, v.57, n.1, p.1-21.

[23] SHULMAN, L (1989). Paradigmas y programas de investigaciónenelstudio de laenseñanza: una perspective contemporanea. In: WITTROCK, M. *La investigación de la enseñanza. Enfoques, teorias y métodos*. Barcelona: PAIDÓS, p.9-91.

[24] TALL, D.; Vinner, S (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity, *Educational Studies in Mathematics*, Dordrecht, vol. 3, n. 12, p. 151-169.

[25] TARDIF, M, LESSART, C (2005). *O trabalho docente. Elementos para uma teoria da docência como profissão de interações humanas*. Petrópolis, RJ: Vozes.

[26] VINNER, S (1991). The Role of Definitions in the Teaching and Learning of Mathematics. In: TALL, D. *Advanced Mathematical Thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers, p. 65-81.

[27] ZABALA, A.; ARNAU, L (2007). *11 Ideias clave. Como aprender y ensñar competencias*. Barcelona: Graó.

[28] ZABALZA, M.(2004). *O ensino universitário. Seu cenário e seus protagonistas*. Porto Alegre: Artmed.

ANEXO 1

	SexoF s	SexoM s	Idade1 s	Idade2 s	Idade3 s	Idade4 s	V19_SimMudar s	V19_NaoNR_Mudar s	Competencia[V27_01	Competencia[V27_02	Competencia[V27_03
Suj001	1	0	0	0	1	0	1	0	0.9	0.8	0.9
Suj002	1	0	0	0	0	1	1	0	0.9	0.9	0.9
Suj003	1	0	0	0	0	1	0	1	0.8	0.9	0.8
Suj004	0	1	0	0	1	0	1	0	0.9	0.8	0.8
Suj005	1	0	0	1	0	0	0	1	0.8	0.9	0.8
Suj006	1	0	0	0	0	1	0	1	0.9	0.8	0.9
Suj007	1	0	0	0	1	0	0	1	0.9	0.9	0.9
Suj008	0	1	0	0	1	0	0	1	0.7	0.9	0.9
Suj009	1	0	0	1	0	0	1	0	0.9	0.9	1
Suj010	1	0	1	0	0	0	1	0	0.9	1	1
Suj011	1	0	0	1	0	0	1	0	0.8	0.9	0.7
Suj012	1	0	0	1	0	0	0	1	0.8	1	1
Suj013	0	1	0	0	0	1	1	0	0.9	0.9	0.9
Suj014	1	0	1	0	0	0	1	0	1	0.8	0.9
Suj015	0	1	1	0	0	0	1	0	0.6	0.9	0.9
Suj016	0	1	0	1	0	0	1	0	1	1	1
Suj017	0	1	0	1	0	0	1	0	0.9	0.8	0.8
Suj018	1	0	0	0	1	0	0	1	0.7	0.8	0.8
Suj019	1	0	0	1	0	0	1	0	0.6	0.8	0.8
Suj020	1	0	0	0	1	0	0	1	0.8	0.9	0.9

	Competencia[V27_04	Competencia[V27_05	Competencia[V27_06	Competencia[V27_07	Competencia[V27_08	Competencia[V28_01	Competencia[V28_02	Competencia[V28_03	Competencia[V28_04	Competencia[V28_05	Competencia[V28_06	Competencia[V28_07	Competencia[V28_08
Suj001	1	0.8	0.9	0.9	1	1	1	1	1	0.9	1	0.8	0.9
Suj002	0.9	0.8	0.9	0.9	0.7	0.7	0.7	0.8	0.7	0.7	0.9	0.7	0.7
Suj003	0.9	1	0.9	0.9	0.6	0.8	0.9	0.9	0.8	0.8	0.9	1	1
Suj004	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.9	0.9	0.9
Suj005	0.9	0.9	0.8	0.7	0.9	0.9	1	0.9	0.8	0.8	1	0.9	0.8
Suj006	0.8	0.7	0.8	0.8	0.7	0.7	0.7	0.9	0.7	0.7	1	0.7	0.7
Suj007	0.9	0.9	0.9	0.9	0.9	0.1	0.1	0.1	0.2	0.2	0.2	0	0
Suj008	0.9	0.9	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Suj009	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1	1	1
Suj010	0.9	0.9	1	1	0.4	1	1	0.9	0.9	0.9	1	1	0.9
Suj011	0.9	0.9	0.8	0.7	0.7	0.9	0.8	0.8	0.9	0.9	0.8	0.9	0.9
Suj012	0.8	1	0.8	1	0.7	0.8	0.8	1	0.8	0.8	1	1	1
Suj013	0.9	0.9	0.9	0.9	0.9	0	0	0	0	0	0	0	0
Suj014	0.8	0.7	0.9	0.9	0.8	0.9	0.8	0.8	0.8	0.9	0.9	0.9	0.7
Suj015	0.7	0.9	0.9	0.8	0.8	0.8	0.8	0.7	0.9	0.6	0.7	0.9	0.9
Suj016	0.9	0.5	0.6	0.8	1	1	1	1	0.5	0.6	1	1	0.7
Suj017	0.8	0.6	0.8	0.7	0.6	0.9	0.8	0.8	1	0.9	0.8	0.7	0.7
Suj018	0.8	0.7	0.8	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.7
Suj019	0.5	0.8	0.4	0.5	0.4	0.6	0.6	0.6	0.6	0.6	0.5	0.8	0.8
Suj020	1	0.9	0.9	0.9	0.9	1	1	0.9	0.9	0.9	1	0.9	0.9

LE TRAITEMENT STATISTIQUE DE DONNEES ET LA COMPLEXITE DE L'EXPRESSION ET DU DEVELOPPEMENT DIFFERENTIELS DU SUJET PSYCHOLOGIQUE

Christian PELLOIS[1]

THE STATISTICAL TREATMENT OF DATA AND THE COMPLEXITY OF DIFFERENTIAL DEVELOPMENT AND EXPRESSION OF THE PSYCHOLOGICAL PERSON

RÉSUMÉ

Par rapport à la conception épistémique du développement de la pensée, l'approche différentielle des sujets psychologiques amène à concevoir l'utilisation de l'analyse implicative dans une démarche méthodologique de recherche plus large. Celle-ci permet, de traiter des complexités de l'expression et du développement différentiels de ces sujets psychologiques, notamment sous l'angle de la vicariance des voies de développement empruntées. Cette approche différentielle amène à suggérer au travers d'un exemple d'application de CHIC sur données formelles, des pistes de recherche possibles portant sur l'élargissement des perspectives de l'analyse implicative à d'autres considérations concernant l'analyse statistique de données dans son ensemble.

Mots-clés : *sujet, différence, psychologique, voies, expressions, analyse, statistique, implicative.*

ABSTRACT

Compared to the epistemic conception of the development of thought, the differential approach of psychological persons led him to design the use of implicative analysis in a methodological approach of wider research. This allows to process incremental the complexities of the differential expression and development of these psychological persons, including the issue of vicarious borrowed development ways. This differential approach leads to suggest through an example application of CHIC on formal data, possible research tracks on the expansion prospects of the implicative analysis to other considerations for the statistical analysis of data as a whole.

Keywords: *person, difference, psychological, ways, expressions, analysis, statistical, implicative.*

1 Introduction

Le contexte de la psychologie, et en particulier celui la psychologie différentielle, amène à concevoir le développement et l'expression humaine du *sujet*, les *organisations* produites qui semblent les sous-tendent, sous l'angle : soit de conceptions homogènes, structurées selon une modélisation à visée, sinon universelle tout au moins générale, mais tolérant une modulation différentielle des capacités, notamment cognitives, acquises, soit de conceptions renvoyant à des approches plus locales et relevant davantage d'une hétérogénéité plus radicale des domaines de conduites de la personne. Les premières conceptions réfèrent, tel le modèle piagétien, tout au moins dans son

[1] 7 rue des pins, 14470, Courseulles sur mer, pellois.christian@wanadoo.fr

approche désormais classique, à un *sujet épistémique*, plutôt conçu selon une organisation « abstraite » plus ou moins logique et dont le développement serait, dans ses grandes étapes, univoque. Les secondes conceptions relèvent des différentes approches du *sujet psychologique* pris dans ses différents contextes d'expression et de développement. Ces secondes conceptions amènent à considérer cette fois-ci davantage *des sujets*, non plus observés sous l'angle d'une conception générale, mais sous *des conceptions sujets/contextes* variés et variables, sans doute plus *locales*, en fonction des sujets et des situations considérées.

Cette deuxième approche davantage plurielle a des conséquences sur la recherche en ce domaine et, notamment, sur le recueil, et l'utilisation des données d'observation. Elle a des conséquences, en particulier, sur l'utilisation, méthodique et circonstanciée, qui peut être faite des outils de traitement statistique de ces données.

Les développements qui suivent rappellent quelques aspects conceptuels importants à prendre en compte dans le domaine ci-dessus précisé. Il est présenté, ensuite, une illustration formelle d'une utilisation méthodologique possible adaptée au contexte particulier d'une psychologie différentielle qui inscrit, cependant, le développement pluriel de sujets dans une modélisation qui reste, au sens large, généraliste. Cette illustration permet de poser, alors de multiples questions et des perspectives de recherches qui en découlent au regard des modèles statistiques du traitement de données et en particulier au regard de celui de l'analyse implicative. Ces perspectives permettent d'entrevoir des chantiers possibles d'évolution du logiciel, voire de développement de modules complémentaires de CHIC et/ou qui lui seraient « articulés ».

2 Le cadre conceptuel considéré : brefs rappel

La présentation de ce cadre conceptuel, plus différentiel et pluraliste, au regard d'un cadre initial plus classique et davantage généraliste a donné lieu, déjà depuis quelques décennies, à de nombreux travaux et conceptions, tels que présentés, par exemple, dans Bideaud *et al.* (1993). Il a également fait l'objet d'un développement plus complet dans un autre texte (Pellois (2015)) et, en outre, a été repris et complété au point 4 en vue d'introduire la réflexion finale du présent texte. Aussi il ne sera rappelé ici que les grandes lignes de ce cadre, ceci afin d'introduire, au mieux : aux préoccupations sous-jacentes à la simulation donnant lieu à application du logiciel CHIC et présentée ensuite, ainsi qu'aux perspectives de développement de celui-ci inscrites également dans la réflexion finale de ce texte.

Rappelons, premièrement, que le modèle piagétien présente, de l'enfance à l'adolescence, le développement progressif puis l'expression à l'âge adulte de la structure, supposée acquise, de la pensée opératoire permettant notamment le raisonnement hypothético-déductif, soit une modélisation formelle qui renvoie davantage à un *sujet épistémique* qu'à un sujet réel particulier, le *sujet psychologique*.

Or, une fois la structure ou les sous-structures opératoires acquises, dans son développement aussi bien que dans sont expression il est rare que le *sujet psychologique* puisse activer celles-ci sur l'ensemble des registres et contenus abordés. L'expression d'une pensée opératoire efficace se fait, si elle se fait, pour beaucoup de *sujets psychologiques*, de manière souvent plus « locale » en fonction des situations et contextes rencontrés. Le développement lui-même chemine par des « phases » (sous-

stades ?) variées (cf., notamment, Longeot (1978)). Des *sujets psychologiques* passent par des sous-étapes du développement selon des successions différentes, aboutissant plus ou moins aux mêmes grandes étapes de ce développement. Il s'en déduit que des cheminements menant à l'organisation opératoire formelle à l'adolescence peuvent être alternatifs et donc vicariants.

Une part de ces différences tient au fait que ce qui permet le développement opératoire piagétien dépendrait d'habiletés cognitives dont les performances évoluent elles-mêmes dans le temps, certaines pouvant être communes d'autres plus ou moins différentes d'un *sujet psychologique* à un autre, voire, pour le même sujet, d'une situation à une autre. Les variations, alors, tant intra-individuelles qu'interindividuelles dans le développement, comme dans l'expression opératoire, traduiraient des modalités, davantage du domaine fonctionnel, de styles (Huteau (1980)) ou de types d'expression cognitive de *sujets psychologiques* plus ou moins différents et, et, pour un même sujet, plus ou moins aisément mobilisables (Lautrey *et al.* (1995)).

Deuxièmement, il apparaît que le sujet psychologique développe dans d'autres domaines d'expression, d'autres conduites, y compris cognitives, différentes de celles relevant du raisonnement de forme hypothético-déductive. Il en découle la conséquence suivante. Bien que des études aient été faites en ce sens y compris dans le cadre des conceptions piagétiennes, il n'est pas sûr, au jour d'aujourd'hui, que l'on puisse ramener l'ensemble des conduites et comportements des *sujets psychologiques* à un développement se faisant essentiellement sur le mode piagétien, voire néo-piagétien, même pris au sens large. Ces différents comportements et conduites autres, comme, par exemple, l'expression créative, la compréhension de l'ordre du sensible, de l'émotionnel, le jugement esthétique, l'empathie, renvoient à des dimensions de la personne humaine sans doute, pour partie au moins et dans leur genèse initiale, très différentes de celles constitutives de la pensée opératoire au sens piagétien du terme, voire, potentiellement, plus ou moins compatibles avec celle-ci.

Enfin, troisièmement, les conduites opératoires du sujet au sens piagétien du terme, leur expression efficace en situation, et surtout d'autres formes complémentaires sophistiquées de l'expression du raisonnement, semblent, elles-mêmes, dépendre, non seulement de fonctions cognitives sous-jacentes, comme la mémoire, l'attention, l'espace mental à disposition (Pascual Leone, 1982, Rideaupierre, 1980), les capacités spatiales/représentatives ou verbales/numériques, analogiques, propositionnelles (Lautrey et Chartier, 1987), mais aussi, pour partie tout au moins, des autres registres d'expression, tel qu'évoqués ci-dessus, donc d'autres aspects du développement du *sujet psychologique* (Richard, 1990, Bideaud *et al.*,1993, Damasio, 1995).

Existe-il alors un ensemble global d'expressions acquises du sujet comme *produit* de son développement, ensemble donnant lieu à une forme organisée, incluant plus ou moins la dimension opératoire, et dont il conviendrait de décrire l'organisation ? Ou différents *sujets psychologiques* se développeraient-ils, puis s'exprimeraient-ils selon différentes formes organisées plus ou moins communes, plus ou moins complémentaires, voire plus ou moins incompatibles, c'est différentes formes donnant lieu à des organisations ou configurations qualitativement différentes ? Au jour d'aujourd'hui le débat sur ce point ne semble pas encore réellement tranché.

3 Une simulation et des calculs avec CHIC pour voir

3.1 Rappel des raisons à l'origine de cette étude en termes de simulation ; précisions concernant ce contexte, et point de vue complémentaire

Les développements qui précèdent amènent bien à une série de questionnements majeurs quant aux manières d'exprimer une certaines efficacité, ici cognitive, du sujet psychologique et plus particulièrement sous ses aspects différentiels. Si les voies de l'expression ou du développement sont plus ou moins hétérogènes, lorsque ces voies ne sont pas équivalentes, qu'est-ce qui fait obstacle au développement considéré et/ou à une polyvalence de mise en œuvre des organisations cognitives atteintes ? Ou bien, sous une forme plus clairement différentielle, et compte-tenu des configurations de sujets qui seraient différentes, s'il existe plus ou moins de difficulté, suivant les sujets considérés, à maîtriser une compétence[2], en fonction de la configuration des habiletés de chaque personne, comment, par une voie qui serait propre à cette configuration, permettre, malgré tout et sans dommage par ailleurs[3], l'accès[4] à cette compétence ? Il s'agirait, ici, de la pensée opératoire de sujets non spécialement préparés, de par leur orientation « dispositionnelle » au sens de Bruchon-Schweitzer (2002), à une expression pleine et entière, plus ou moins « spontanée » et/ou acquise, en ce domaine.

Dans le prolongement de réflexions antérieures (Pellois, 2007, 2012, 2013a, 2013b), les développements qui précèdent et les questions associées conduisent à s'intéresser au logiciel CHIC et à ses applications d'une certaine manière qui peut être présentée sous la forme des développements suivants.

Pour le *sujet psychologique* dont les conduites adaptatives, ou simplement l'ensemble des registres d'expression, sont plus larges que celles qui sont inscrites simplement dans l'organisation de la pensée logicomathématique hypothéticodéductive, il convient ici de rappeler, de redire, que, l'expression cognitive formelle dans le domaine opératoire du raisonnement[5] dépend[6], de multiples dimensions endogènes[7] et exogènes à la personne[8], dimensions autres que ce qui relève strictement du domaine des opérations logiques. De plus, elle ne constitue pas nécessairement, cette expression cognitive, la seule capacité utile au développement et à l'expression des conduites adaptatives. Elle ne s'applique que dans des contextes de l'expression du sujet particulier renvoyant à la pensée convergente (logique dite du « nécessaire ») et, d'une manière qui lui est propre[9], aux situations conditionnelles plus ou moins contraintes.

En conséquence, les conditions de l'utilisation de l'analyse implicative (Gras, 2005, Gras et Couturier, 2010) font que celle-ci ne peut se faire que dans un cadre strict, bien clarifié, qu'il convient de cerner puis localiser précisément, des modalités de recherche

[2] Ici une capacité, celle du raisonnement opératoire piagétien.

[3] Effets d'incompatibilité.

[4] Objectif éminemment éducatif, perspective d'un enseignement développemental cognitif, problématique pédagogique particulièrement pertinente…

[5] Sous son organisation conceptualisée, le développement de l'enfant, qui permet l'acquisition de cette capacité formelle.

[6] Dans sa construction comme dans son utilisation.

[7] Tel qu'évoqué précédemment, non seulement cognitif, mais référant à la personnalité, au comportement social, à l'affectif, à l'émotionnel, etc.

[8] Environnement physique, social, relationnel, culturel, etc.

[9] Emboîtement du raisonnement, loi de non contradiction interne et avec les faits observés, etc.

mises en œuvre concernant le *sujet psychologique*. Sa pertinence dépend donc des conditions dûment et rigoureusement précisées du contexte dans lequel son utilisation peut-être faite (Pellois, 2007, 2012, 2013a) et ceci plutôt sous la forme confirmatoire d'un modèle implicatif conçu pour traiter des questions que se pose la recherche considérée.

Mais, à côté de cette position restrictive, la reconnaissance par CHIC d'une conception non pas strictement implicative mais quasi-implicative, amène à des ouvertures. Ces ouvertures conduisent à une utilisation du logiciel féconde, sous la forme de méthodes particulières de traitement, organisées de manière successive et s'articulant autour de la question de l'homogénéité ou de l'hétérogénéité des populations et sous-populations considérées. Ces méthodes appliquées à des modélisations formelles, telles que celles qui sont évoquées dans une autre contribution (Pellois, 2015), permettent de dépasser une position implicative strictement (abusivement ?) généraliste et de prendre une position davantage différentielle plus adaptée, en partie tout au moins, à des conceptions plus récentes portant sur le développement, notamment cognitif.

3.2 En quoi une utilisation particulière de CHIC peut-elle faire avancer sur ces questions portant sur l'homogénéité/hétérogénéité d'un domaine considéré ?

Une simulation et des calculs avec CHIC ont été faits pour voir en quoi le logiciel permet ou pas de traiter de la question posée et, en particulier, en quoi l'approche différentielle et l'utilisation de l'indice de typicalité permet, ou ne permet pas, de rendre compte de cette question de l'homogénéité ou de l'hétérogénéité en fonction des caractéristiques de sous-groupes issus d'une population plus large.

Cette simulation et ces calculs partent d'une population de sujets, retenue ici comme population de référence, et dont les soixante six sujets sont répartis sur une grille croisant les performances supposées être réalisées sur deux dimensions (X et Y[10]) selon une distribution qui renvoie à une quasi-implication entre les deux dimensions considérées.

Les chiffres ici portés dans chaque case présentant cette distribution (voir ci-dessous à la section 3.3.1.) correspondent non pas **à un nombre de sujets mais aux numéros affectés à chaque sujet considéré** dans cette simulation. Ces numéros « étiquettes » vont de 1 à 66. Un sujet correspond donc à chaque case portant un numéro. Les questions posées à propos de cette distribution, et au regard des développements précédents, sont les suivantes.

[10] Ces épreuves susceptibles de valider après analyse statistique de possibles phénomènes de vicariance, ou pas, selon les différents cas de figure traités ici, pourrait relever pour l'une de niveaux d'exercices de combinatoire et, pour l'autre, de logique propositionnelle (groupe INRC) (cf. Longeot (1978)), ou d'épreuves portant sur des caractéristiques du développement opératoire sous contenus favorables à un traitement selon le style *« dépendant du champ »* et selon le style *« indépendant du champ »* (cf. Huteau (1980)), ou des contenus traité sur un plan analogique et propositionnel (cf. Lautrey, in Reuchlin *et al.* (1990), notamment, p. 202 et suivantes), ou de compréhension de syllogismes comportant ou non un contenu sémantique, ou des processus logico-linguistiques et spatiaux (Houdé, in Houdé *et al.* (2002), p.556 et suivantes), ou de contenus de difficulté progressive nécessitant des traitements successifs et simultanés etc.

La tendance générale de quasi-implication, aux erreurs de mesures près, représente-t-elle l'expression d'une *contrainte* générale agissant sur l'ensemble de la population considérée amenant à ce que le seul chemin possible entre les variables X et Y soit la nécessité de maîtriser X pour maîtriser Y (Y implique X quasi-systématiquement) ?

De manière plus générale, la population considérée dans son ensemble est-elle composée d'une tendance —attestée par les calculs faits— effectivement homogène aux sous-populations choisies par le chercheur en fonction de ses hypothèses sur cette question, renvoyant donc bien au même constat d'homogénéité de la tendance pour chacune d'elles ? Ou bien ces sous-populations caractérisent-elles des relations entre les variables spécifiques à ces sous-populations et, donc, des tendances hétérogènes entre elles (à côté de Y implique X pour des sujets, X implique Y, pour une part minoritaire de sujets, ou il y a absence d'implication pour une autre part minoritaire de sujet, etc.) ?

Par ailleurs, la quasi-implication peut être étudiée sur la base d'un continuum[11] de performances (une échelle de performances standardisées allant de 0.0 à 1.0 avec des pas successifs à intervalles standards de 0.1[12]) sur deux dimensions, ou sur le mode d'une répartition dichotomique (maîtrise/non maîtrise, oui/non, 0 ou 1). Passer de l'une à l'autre modalité de distribution modifie-t-il quelque peu la forme du raisonnement implicatif classique tenu pour les deux dimensions considérées ? Et, si tel est le cas, cela renvoie-t-il alors, ici, à des résultats différents dont pourrait rendre compte CHIC ?

3.3 La simulation proprement dite, soit :

3.3.1 La population {3a} considérée (groupe de référence X3a/Y3a)

y3a/x3a	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,9	1	2	3	4	5	6	7	8	9	10
0,8	11	12	13	14	15	16	17	18	19	
0,7	20	21	22	23	24	25	26	27		65
0,6	28	29	30	31	32	33	34		63	
0,5	35	36	37	38	39	40	41			
0,4	42	43	44	45	46		48		55	
0,3	49	50	51	52	53	54				
0,2	56	57	58					64		
0,1	59	60		62		47				
0,0	61	66								

Tableau 1 – la population {3a} considérée

Le tableau des données correspondant à cette population.

[11] Relevant d'un traitement possible en termes de logique floue ?

[12] Marquant, par exemple, l'évolution progressive : du degré d'expression d'une qualité ou de maîtrise d'une compétence, du score portant sur un acquis de connaissance, du niveau de performance dans une capacité donnée, etc.

	x3a	y3a		x3a	y3a		x3a	y3a
sujet 1	0	0,9	sujet 23	0,3	0,7	sujet 45	0,3	0,4
sujet 2	0,1	0,9	sujet 24	0,4	0,7	sujet 46	0,4	0,4
sujet 3	0,2	0,9	sujet 25	0,5	0,7	sujet 47	0,5	0,1
sujet 4	0,3	0,9	sujet 26	0,6	0,7	sujet 48	0,6	0,4
sujet 5	0,4	0,9	sujet 27	0,7	0,7	sujet 49	0	0,3
sujet 6	0,5	0,9	sujet 28	0	0,6	sujet 50	0,1	0,3
sujet 7	0,6	0,9	sujet 29	0,1	0,6	sujet 51	0,2	0,3
sujet 8	0,7	0,9	sujet 30	0,2	0,6	sujet 52	0,3	0,3
sujet 9	0,8	0,9	sujet 31	0,3	0,6	sujet 53	0,4	0,3
sujet 10	0,9	0,9	sujet 32	0,4	0,6	sujet 54	0,5	0,3
sujet 11	0	0,8	sujet 33	0,5	0,6	sujet 55	0,8	0,4
sujet 12	0,1	0,8	sujet 34	0,6	0,6	sujet 56	0	0,2
sujet 13	0,2	0,8	sujet 35	0	0,5	sujet 57	0,1	0,2
sujet 14	0,3	0,8	sujet 36	0,1	0,5	sujet 58	0,2	0,2
sujet 15	0,4	0,8	sujet 37	0,2	0,5	sujet 59	0	0,1
sujet 16	0,5	0,8	sujet 38	0,3	0,5	sujet 60	0,1	0,1
sujet 17	0,6	0,8	sujet 39	0,4	0,5	sujet 61	0	0
sujet 18	0,7	0,8	sujet 40	0,5	0,5	sujet 62	0,3	0,1
sujet 19	0,8	0,8	sujet 41	0,6	0,5	sujet 63	0,8	0,6
sujet 20	0	0,7	sujet 42	0	0,4	sujet 64	0,7	0,2
sujet 21	0,1	0,7	sujet 43	0,1	0,4	sujet 65	0,9	0,7
sujet 22	0,2	0,7	sujet 44	0,2	0,4	sujet 66	0,1	0

Tableau 2 – Tableau des données de la population {3a}

3.3.2 Une répartition en sous-groupes d'une population {3a1} dont la répartition globale des données est identique à {3a}

Cette population présente une répartition de données identique à la population du tableau ci-dessus (le groupe de référence X3a/Y3a devient le groupe X3a1/Y3a1) et, donc, la même tendance globale *quasi-implicative*. Mais, sa répartition en trois sous-groupes selon de trois critères (croisement X3a1/Y3a1, critères A, B, C), judicieusement choisis par le chercheur, atteste de trois tendances affirmées, et *quasi- homogènes* pour chacun des sous-groupes considérés, dont les caractéristiques de rapport logique sont différentes les unes des autres (implication, corrélation, implication inverse). Ce constat amène alors à reconnaître une *hétérogénéité* entre ces sous-groupes (et donc interne au groupe {3a1}), et à rejeter l'hypothèse *d'homogénéité* interne du groupe dans son ensemble, ce que rend formellement possible le concept de *quasi-implication* dans son caractère plus ou moins accentué.

La population {3a1} considérée (groupe de référence X3a1/Y3a1) et les sous-populations correspondant aux trois critères : A (cases rayées), B (cases blanches) et C (cases gris foncé).

y3a1/x3a1 ↓ →	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,9	1	2	3	4	5	6	7	8	9	10
0,8	11	12	13	14	15	16	17	18	19	
0,7	20	21	22	23	24	25	26	27		65
0,6	28	29	30	31	32	33	34		63	
0,5	35	36	37	38	39	40	41			
0,4	42	43	44	45	46		48		55	
0,3	49	50	51	52	53	54				
0,2	56	57	58					64		
0,1	59	60		62		47				
0,0	61	66								

Tableau 3 – la population {3a1} considérée

Le tableau des données correspondant à cette population.

	x3a1	y3a1	A s	B s	C s		x3a1	A s	B s	C s	y3a1		x3a1	y3a1	A s	B s	C s
sujet 1	0	0,9	1	0	0	sujet 23	0,3	1	0	0	0,7	sujet 45	0,3	0,4	0	1	0
sujet 2	0,1	0,9	1	0	0	sujet 24	0,4	1	0	0	0,7	sujet 46	0,4	0,4	0	1	0
sujet 3	0,2	0,9	1	0	0	sujet 25	0,5	1	0	0	0,7	sujet 47	0,5	0,1	0	0	1
sujet 4	0,3	0,9	1	0	0	sujet 26	0,6	0	0	1	0,7	sujet 48	0,6	0,4	0	0	1
sujet 5	0,4	0,9	1	0	0	sujet 27	0,7	0	1	0	0,7	sujet 49	0	0,3	1	0	0
sujet 6	0,5	0,9	1	0	0	sujet 28	0	1	0	0	0,6	sujet 50	0,1	0,3	1	0	0
sujet 7	0,6	0,9	1	0	0	sujet 29	0,1	1	0	0	0,6	sujet 51	0,2	0,3	0	0	1
sujet 8	0,7	0,9	1	0	0	sujet 30	0,2	1	0	0	0,6	sujet 52	0,3	0,3	0	1	0
sujet 9	0,8	0,9	0	0	1	sujet 31	0,3	1	0	0	0,6	sujet 53	0,4	0,3	0	1	0
sujet 10	0,9	0,9	0	1	0	sujet 32	0,4	1	0	0	0,6	sujet 54	0,5	0,3	0	0	1
sujet 11	0	0,8	1	0	0	sujet 33	0,5	0	1	0	0,6	sujet 55	0,8	0,4	0	0	1
sujet 12	0,1	0,8	1	0	0	sujet 34	0,6	0	1	0	0,6	sujet 56	0	0,2	1	0	0
sujet 13	0,2	0,8	1	0	0	sujet 35	0	1	0	0	0,5	sujet 57	0,1	0,2	0	0	1
sujet 14	0,3	0,8	1	0	0	sujet 36	0,1	1	0	0	0,5	sujet 58	0,2	0,2	0	1	0
sujet 15	0,4	0,8	1	0	0	sujet 37	0,2	1	0	0	0,5	sujet 59	0	0,1	0	0	1
sujet 16	0,5	0,8	1	0	0	sujet 38	0,3	1	0	0	0,5	sujet 60	0,1	0,1	0	1	0
sujet 17	0,6	0,8	1	0	0	sujet 39	0,4	1	0	0	0,5	sujet 61	0	0	0	1	0
sujet 18	0,7	0,8	0	0	1	sujet 40	0,5	0	1	0	0,5	sujet 62	0,3	0,1	0	0	1
sujet 19	0,8	0,8	0	1	0	sujet 41	0,6	0	1	0	0,5	sujet 63	0,8	0,6	0	0	1
sujet 20	0	0,7	1	0	0	sujet 42	0	1	0	0	0,4	sujet 64	0,7	0,2	0	0	1
sujet 21	0,1	0,7	1	0	0	sujet 43	0,1	1	0	0	0,4	sujet 65	0,9	0,7	0	0	1
sujet 22	0,2	0,7	1	0	0	sujet 44	0,2	1	0	0	0,4	sujet 66	0,1	0	0	0	1

Tableau 4 – Tableau des données de la population {3a1}

3.3.3 Une nouvelle répartition en sous-groupes d'une population {3a2} dont la répartition globale des données est identique à {3a}

Cette répartition en sous-groupes, selon les mêmes critères que précédemment (croisement X3a2/Y3a2, critères A, B, C), présente, cette fois-ci, des caractéristiques *assez semblables* (plutôt *quasi-implicatives*) de distribution des sujets à l'interne de chaque sous-groupe, donc sans tendances caractéristiques nettement différentes entre eux, mais dans une répartition *plutôt hétérogène à l'interne de chaque sous-groupe*. Il s'agit toujours, par ailleurs, de la même répartition globale du groupe d'ensemble, répartition *quasi-implicative* conforme à celle de la même population de référence {3a} que précédemment.

y3a2/x3a2	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,9	1	2	3	4	5	6	7	8	9	10
0,8	11	12	13	14	15	16	17	18	19	
0,7	20	21	22	23	24	25	26	27		65
0,6	28	29	30	31	32	33	34		63	
0,5	35	36	37	38	39	40	41			
0,4	42	43	44	45	46		48		55	
0,3	49	50	51	52	53	54				
0,2	56	57	58					64		
0,1	59	60		62		47				
0,0	61	66								

Tableau 5 – la population {3a2} considérée

Le tableau des données correspondant à cette population.

	x3a2	y3a2	A s	B s	C s		x3a2	A s	B s	C s	y3a2		x3a2	y3a2	A s	B s	C s
sujet 1	0	0,9	1	0	0	sujet 23	0,3	0	1	0	0,7	sujet 45	0,3	0,4	0	1	0
sujet 2	0,1	0,9	0	1	0	sujet 24	0,4	0	0	1	0,7	sujet 46	0,4	0,4	0	0	1
sujet 3	0,2	0,9	0	0	1	sujet 25	0,5	1	0	0	0,7	sujet 47	0,5	0,1	1	0	0
sujet 4	0,3	0,9	1	0	0	sujet 26	0,6	0	1	0	0,7	sujet 48	0,6	0,4	1	0	0
sujet 5	0,4	0,9	0	1	0	sujet 27	0,7	0	0	1	0,7	sujet 49	0	0,3	0	0	1
sujet 6	0,5	0,9	0	0	1	sujet 28	0	0	1	0	0,6	sujet 50	0,1	0,3	1	0	0
sujet 7	0,6	0,9	1	0	0	sujet 29	0,1	0	0	1	0,6	sujet 51	0,2	0,3	0	1	0
sujet 8	0,7	0,9	0	1	0	sujet 30	0,2	1	0	0	0,6	sujet 52	0,3	0,3	0	0	1
sujet 9	0,8	0,9	0	0	1	sujet 31	0,3	0	1	0	0,6	sujet 53	0,4	0,3	1	0	0
sujet 10	0,9	0,9	1	0	0	sujet 32	0,4	0	0	1	0,6	sujet 54	0,5	0,3	0	1	0
sujet 11	0	0,8	0	1	0	sujet 33	0,5	1	0	0	0,6	sujet 55	0,8	0,4	0	1	0
sujet 12	0,1	0,8	0	0	1	sujet 34	0,6	0	1	0	0,6	sujet 56	0	0,2	0	0	1
sujet 13	0,2	0,8	1	0	0	sujet 35	0	1	0	0	0,5	sujet 57	0,1	0,2	1	0	0
sujet 14	0,3	0,8	0	1	0	sujet 36	0,1	0	1	0	0,5	sujet 58	0,2	0,2	0	1	0
sujet 15	0,4	0,8	0	0	1	sujet 37	0,2	0	0	1	0,5	sujet 59	0	0,1	1	0	0
sujet 16	0,5	0,8	1	0	0	sujet 38	0,3	1	0	0	0,5	sujet 60	0,1	0,1	0	1	0
sujet 17	0,6	0,8	0	1	0	sujet 39	0,4	0	1	0	0,5	sujet 61	0	0	0	1	0
sujet 18	0,7	0,8	0	0	1	sujet 40	0,5	0	0	1	0,5	sujet 62	0,3	0,1	0	0	1
sujet 19	0,8	0,8	1	0	0	sujet 41	0,6	1	0	0	0,5	sujet 63	0,8	0,6	0	0	1
sujet 20	0	0,7	0	1	0	sujet 42	0	0	1	0	0,4	sujet 64	0,7	0,2	0	0	1
sujet 21	0,1	0,7	0	0	1	sujet 43	0,1	0	0	1	0,4	sujet 65	0,9	0,7	1	0	0
sujet 22	0,2	0,7	1	0	0	sujet 44	0,2	1	0	0	0,4	sujet 66	0,1	0	0	0	1

Tableau 6 – Tableau des données de la population {3a2}

3.3.4 Une troisième répartition en sous-groupes d'une population {3a3} dont la répartition globale des données est identique à {3a}

Cette fois-ci la répartition en sous-groupes, toujours selon les mêmes critères, permet d'observer des tendances comparables entre les sous-groupes (croisement X3a3/Y3a3, critères A, B, C), donc, une homogénéité de tendance entre eux. Ces tendances sont également homogènes à la tendance observée pour toute la population considérée {3a3}.

y3a3/x3a3	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,9	1	2	3	4	5	6	7	8	9	10
0,8	11	12	13	14	15	16	17	18	19	
0,7	20	21	22	23	24	25	26	27		65
0,6	28	29	30	31	32	33	34		63	
0,5	35	36	37	38	39	40	41			
0,4	42	43	44	45	46		48		55	
0,3	49	50	51	52	53	54				
0,2	56	57	58					64		
0,1	59	60		62		47				
0,0	61	66								

Tableau 7 – la population {3a3} considérée

Les tableaux des données correspondant à cette population.

	x3a3	y3a3	A s	B s	C s		x3a3	A s	B s	C s	y3a3		x3a3	y3a3	A s	B s	C s
sujet 1	0	0,9	1	0	0	sujet 23	0,3	0	0	1	0,7	sujet 45	0,3	0,4	0	0	1
sujet 2	0,1	0,9	1	0	0	sujet 24	0,4	0	0	1	0,7	sujet 46	0,4	0,4	1	0	0
sujet 3	0,2	0,9	1	0	0	sujet 25	0,5	0	0	1	0,7	sujet 47	0,5	0,1	0	1	0
sujet 4	0,3	0,9	1	0	0	sujet 26	0,6	0	0	1	0,7	sujet 48	0,6	0,4	0	0	1
sujet 5	0,4	0,9	1	0	0	sujet 27	0,7	0	0	1	0,7	sujet 49	0	0,3	1	0	0
sujet 6	0,5	0,9	1	0	0	sujet 28	0	1	0	0	0,6	sujet 50	0,1	0,3	0	1	0
sujet 7	0,6	0,9	1	0	0	sujet 29	0,1	0	1	0	0,6	sujet 51	0,2	0,3	0	0	1
sujet 8	0,7	0,9	1	0	0	sujet 30	0,2	0	0	1	0,6	sujet 52	0,3	0,3	0	0	1
sujet 9	0,8	0,9	1	0	0	sujet 31	0,3	0	0	1	0,6	sujet 53	0,4	0,3	0	1	0
sujet 10	0,9	0,9	0	1	0	sujet 32	0,4	0	0	1	0,6	sujet 54	0,5	0,3	1	0	0
sujet 11	0	0,8	1	0	0	sujet 33	0,5	0	0	1	0,6	sujet 55	0,8	0,4	1	0	0
sujet 12	0,1	0,8	0	1	0	sujet 34	0,6	0	0	1	0,6	sujet 56	0	0,2	1	0	0
sujet 13	0,2	0,8	0	1	0	sujet 35	0	1	0	0	0,5	sujet 57	0,1	0,2	0	1	0
sujet 14	0,3	0,8	0	1	0	sujet 36	0,1	0	1	0	0,5	sujet 58	0,2	0,2	0	0	1
sujet 15	0,4	0,8	0	1	0	sujet 37	0,2	0	0	1	0,5	sujet 59	0	0,1	1	0	0
sujet 16	0,5	0,8	0	1	0	sujet 38	0,3	0	0	1	0,5	sujet 60	0,1	0,1	0	1	0
sujet 17	0,6	0,8	0	1	0	sujet 39	0,4	0	0	1	0,5	sujet 61	0	0	1	0	0
sujet 18	0,7	0,8	0	1	0	sujet 40	0,5	0	1	0	0,5	sujet 62	0,3	0,1	0	1	0
sujet 19	0,8	0,8	0	1	0	sujet 41	0,6	1	0	0	0,5	sujet 63	0,8	0,6	0	0	1
sujet 20	0	0,7	1	0	0	sujet 42	0	1	0	0	0,4	sujet 64	0,7	0,2	0	0	1
sujet 21	0,1	0,7	0	1	0	sujet 43	0,1	0	1	0	0,4	sujet 65	0,9	0,7	0	0	1
sujet 22	0,2	0,7	0	1	0	sujet 44	0,2	0	0	1	0,4	sujet 66	0,1	0	0	1	0

Tableau 8 – Tableau des données de la population {3a3}

3.4 Les résultats obtenus après traitements statistiques par CHIC

3.4.1 Tableau synthétisant l'ensemble des résultats des calculs effectués par CHIC

a) *Résultats obtenus avec des performances en continuum (échelle standard de 0.0 à 1.0) sur les deux dimensions considérées*

Tableau avec performances progressives												
	Population globale	Sous-populations			Population globale	Sous-populations			Population globale	Sous-populations		
	3a1 (X/Y)	Critère A	Critère B	Critère C	**3a2 (X/Y)**	Critère A	Critère B	Critère C	**3a3 (X/Y)**	Critère A	Critère B	Critère C
Correlation	**.32**	.47	.98	.73	**.32**	.35	.25	.35	**.32**	.32	.44	.35
Similarité	**.66/.66**	.61/.61	.69/.69	.69/.69	**.66/.66**	.60/.60	.57/.57	.60/.60	**.66/.66**	.63/.63	.64/.64	.55/.55
Implication	**67/79**	60/82	78/78	80/73	**67/79**	62/69	57/60	61/69	**67/79**	61/75	65/75	57/59
Cohésion	**.41/.67**	.23/.74	.65/.65	.70/.54	**.41/.67**	.28/.46	.17/.35	.27/.45	**.41/.67**	.27/.58	.35/.59	.17/.22
	A .008	A .43	A 0	A 0	**A .545**	A .42	A 0	A 0	**A .716**	A .42	A 0	A 0
Typicalité	**B .393**	B 0	B .40	B 0	**B .359**	B 0	B .42	B 0	**B .545**	B 0	B .42	B 0
	C .997	C 0	C 0	C .40	**C .359**	C 0	C 0	C .42	**C .085**	C 0	C 0	C .42
Orientation de la tendance	X --> Y	X --> Y	*X et Y*	**Y --> X**	X --> Y	X --> Y	X --> Y	X --> Y	X --> Y	X --> Y	X --> Y	X --> Y
Ef. pop. t./s.-p.	**66**	38	14	15	**66**	22	22	22	**66**	22	22	22

Tableau 9 – Résultats obtenus par CHIC avec des données quantitatives en continu

b) *Résultats obtenus avec transformation des données en une répartition dichotomique (maîtrise/non maîtrise : 0, 1) de part et d'autre de la limite de classe correspondant au score théorique moyen (0.5), pour chacune des deux dimensions retenues*

Tableau avec performances ramenées à o (absence, score < ou = à 0,4) ou 1 (présence, score > ou = à 0,5)												
	Population globale	Sous-populations			Population globale	Sous-populations			Population globale	Sous-populations		
	3a1 (X/Y)	A	B	C	**3a2 (X/Y)**	A	B	C	**3a3 (X/Y)**	A	B	C
Correlation	**.20**	.19	1.00	.50	**.20**	.23	.04	.31	**.20**	.11	.31	.17
Similarité	**.78/.78**	.67/.67	.97/.97	.82/.82	**.78/.78**	.67/.67	.54/.54	.77/.77	**.78/.78**	.60/.60	.77/.77	64/.64
Implication	**73/83**	58/63	98/98	83/72	**73/83**	61/64	44/37	67/74	**73/83**	50/48	67/74	56/56
Cohésion	**.55/.75**	.18/.30	.99/.99	.75/.53	**.55/.75**	.27/.32	.0/.0	.40/.56	**.55/.75**	.0/.0	.40/.56	.14/.14
	A .003	A .43	Absence	A 0	**A .307**	A .42	Absence	A 0	**A .480**	A .40	A 0	A 0
Typicalité	**B .820**	B 0	de	B 0	**B .480**	B 0	de	B 0	**B .480**	B 0	B .40	B 0
	C .988	C 0	résultats	C .40	**C .480**	C 0	résultats	C .40	**C .307**	C 0	C 0	C .42
Orientation de la tendance	X --> Y	X --> Y	*X et Y*	**Y --> X**	X --> Y	X --> Y	X ? Y	X --> Y	X --> Y	X ? Y	X --> Y	X ? Y
Ef. pop. t./s.-p.	**66**	38	14	15	**66**	22	22	22	**66**	22	22	22

Tableau 10 – Résultats obtenus par CHIC avec des données quantitatives dichotomiques

3.4.2 Bref commentaire synthétique portant sur les tableaux et les résultats obtenus

En ce qui concerne le tableau portant sur les données présentant un continuum des performances, les résultats obtenus par le groupe dans son entier et ceux qui sont obtenus par les différentes répartitions des sous-groupes sont conformes aux résultats attendus. Ils caractérisent systématiquement la tendance implicative de type X→Y, excepté pour la répartition en sous-groupes hétérogènes entre eux (groupe {3a1}) et dont la répartition choisie amenait à s'attendre à la même tendance implicative pour un groupe, la tendance inverse pour un autre, enfin une relation de type corrélatif seulement pour le troisième groupe. C'est bien ce que nous confirment les résultats obtenus soit :

- critère A indices d'implication 60/82 (coefficient de corrélation .47, similarité .61/.61),
- critère C indices d'implication 80/73 (coefficient de corrélation .73, similarité .69/.69),
- critère B coefficient de corrélation .98, indices de similarité .69/.69 et indices d'implication .78/.78.

Le traitement statistique, en termes de *typicalité,* des critères caractéristiques des sous-groupes conduisant à différentes répartitions internes aux groupes d'ensemble considérés ({3a1}, {3a2}, {3a3}), et dont les répartitions globales de données sont identiques à celle de la population de référence ({3a}), donne des indices de typicalité sans valeur significative. Il y a toutefois deux exceptions, l'une concernant l'indice portant sur la sous-population, seule représentative (population {3a1} critère A) de la tendance générale du groupe initial d'ensemble pour la première répartition des sous-groupes, mais aussi, ce qui est peut-être moins attendu, celle concernant un des sous-groupes des répartitions homogènes(population {3a3}, critère C). Le traitement de la population en termes dichotomiques ne change pas franchement les résultats, mais donne des résultats accentués pour certaines tendances attendues et des résultats non exploitables pour d'autres.

4 A l'issue de cette approche exploratoire, que conclure ?

4.1 La réinscription du sens à l'intersection des différents contextes abordés

Le calcul de typicalité, correspondant à une option de calcul du logiciel CHIC, permettrait bien, dans une certaine mesure, de cerner globalement, et indirectement, en fonction de critères retenus, la question de l'hétérogénéité/homogénéité interne à une distribution de données (population {3a1}, notamment). Mais, ce n'est d'ailleurs pas son objet, il ne permet pas de traiter des caractéristiques détaillées de l'hétérogénéité interne observée pour une distribution donnée. Par contre, une méthode différentielle, traitant de sous-populations observées sous des critères donnés, telle que présentée à la section 3, permet d'en rendre compte, mais de manière simplement descriptive. C'est en ce sens que l'utilisation du logiciel, en l'état de son développement, associée à la

méthode différentielle, peut déjà fournir des résultats féconds dans une conception complexe et complète de l'expression psychologique du sujet.

Plus largement, et de manière très générale est posé ici le problème du contexte des différents domaines de recherche dans lequel sont traitées les données statistiques. Comme cela a déjà été évoqué, le contexte strict de la psychologie cognitive pris au sens précis de la pensée opératoire piagétienne et sous l'angle de la logique implicative, même sous l'angle très spécifié d'un sujet formel, nécessite que soient distingués, et confrontés deux aspects conceptuels articulés en un modèle plus large. Il convient d'y distinguer ce qui relève du développement ou du fonctionnement *psychologique* du sujet réel, de la dynamique d'expression en cours d'élaboration du raisonnement de ce sujet *réel* (Inhelder et Cellerier, 1992), d'une part, ***et***, d'autre part, ce qui relève du *sujet épistémique*, plus « abstrait », renvoyant au « produit » spécifique « construit », une « structure », celle de la pensée formelle telle que le logicien peut la décrire (Piaget *et al.*, 1972).

Mais, même sous ce cadre élargi de l'analyse des conduites, la psychologie ne peut limiter ses investigations du développement et de la mise en œuvre de l'expression du sujet aux conceptions relevant uniquement de ce cadre. Elle doit le faire dans une perspective encore plus large. Cette perspective est celle d'un sujet s'exprimant et agissant dans toute la plénitude des caractéristiques dynamiques qu'il peut mettre en œuvre, caractéristiques qui sont susceptibles d'interagir, positivement ou négativement avec le raisonnement, son développement. L'ensemble de ces différentes modalités de l'expression du sujet va jusqu'à intégrer des dimensions, déjà évoquées, relevant de différents aspects de la mémoire et de l'attention, des modalités ou styles d'expression préférentiels, et des contenus traités, mais également de l'émotionnel, de l'affectivité, de la personnalité (types ou styles), et même des croyances, des modes de fonctionnement plus ou moins spontanés, rationnels, des résistances qui opèrent, de l'éthique, des intérêts, de la motivation et même du désir, des processus implicites plus ou moins conscients, plus ou moins logiques, qui s'expriment, etc. Tout cela amène nécessairement à se confronter à la complexité dans l'approche de la compréhension des situations considérées, des rapports classiques ou moins classiques[13] entre causes et effets, comme de la question de l'objectif et du subjectif au regard de cheminements multiples et variés, voire singulier, et donc à la complexité dans la mise en œuvre des traitements statistiques de données.

Ce contexte du traitement statistique, dans sa version étroite, peut renvoyer, comme dans le cadre de la conception de la pensée opératoire piagétienne, à une forme d'expression psychologique/épistémique relevant de « la logique du nécessaire ». Il invite, donc, ce contexte, à la simple prise en compte de *contraintes* dans le raisonnement conduisant au développement de la pensée convergente pour laquelle il existerait une solution au problème posé et, dans les cas les plus *contraints*, une seule. Mais le cadre général des formes, y compris efficaces, des conduites du sujet n'est pas nécessairement restreint aux *contraintes* qui s'appliqueraient à un raisonnement, et l'expression du sujet est bien davantage multiple en fonction de situations variées rencontrées, situations qui ne relèvent pas ou pas strictement de la simple logique du « nécessaire ».

[13] Effets placebo, nocebo, Hawthorne, effet « expérience », etc.

Les travaux de psychologie différentielle notamment Lautrey (1995) et Longeot, (1978), montrent déjà que, même dans ce cadre restreint, la voie de résolution des problèmes posés n'est pas univoque et qu'il existe parfois, et peut-être même souvent, plusieurs manières[14] d'aboutir à la solution d'un problème considéré, ou à un état d'organisation d'un système donné. Plus largement, la situation du sujet en activité cognitive, dans un cadre plus ou moins riche, plus ou moins ouvert, et de *contraintes* permettant plus ou moins de marges de liberté (Pellois, 2010), peut relever de la possibilité — beaucoup plus complexe à cerner — d'envisager plusieurs, voire de multiples solutions aux problèmes posés, solutions qui ne sont pas nécessairement équivalentes mais simplement différentes. Ces solutions peuvent avoir des avantages et des inconvénients sans qu'il soit possible d'établir d'une manière extrinsèque, plus ou moins universelle, quelle est la solution la meilleure concernant le problème considéré[15]. Chaque solution envisageable peut alors relever du dilemme, c'est-à-dire à la fois répondant par certains aspects au problème à traiter, mais aussi, sous d'autres aspects, posant problème. Le classement hiérarchique ne peut alors se faire, parfois, que, si l'on établit, par exemple à l'extérieur du raisonnement lui-même, un ou des critères, retenus comme prioritaires, et permettant de trancher sur la valeur, ainsi hiérarchisable, de chacune des solutions envisageables.

Et puis il y a d'autres situations de résolution de « problèmes » où une certaine combinatoire complexe, comme c'est le cas dans la créativité, l'inventivité, fait appel d'abord et essentiellement à des capacités de fluidité, de flexibilité mentale, même si, in fine, une certaine logique « du nécessaire »[16] est susceptible d'être mise en œuvre[17]. Ensuite il y a des situations qui renvoient à des dimensions esthétiques ou éthiques[18], pour lesquelles les recherches de « solutions » peuvent passer par des aspects relevant du raisonnement, mais aussi bien d'autres considérations.

Il peut y avoir également, et y compris en prise avec le raisonnement, la nécessité, pour cerner des solutions pertinentes, dans des situations complexes, de mettre en œuvre des cadres plus ouverts d'organisation de la pensée. Ces cadres incluent l'activation de fonctions multiples et variées[19] qui peuvent nécessiter, d'une manière différente, mais aussi complémentaire, des évocations qui ont été faites, précédemment, à ce sujet, comme une certaine forme d'association du cognitif et de l'émotionnel telle qu'envisagée par Damasio (1995). Enfin il y a des situations, ou des moments dans une situation, dans lesquels des capacités totalement spontanées, dites « intuitives », ou relevant d'un fonctionnement par analogie, sont à l'œuvre à côté ou même pendant un raisonnement logique strict. Il peut être pertinent, également, dans certaines conduites, de fonctionner par ressenti sensible des situations vécues[20], perception empathique, imitation (Winnykamen, 1990), etc.

Tout cela peut apporter, directement ou en retour, de l'efficacité dans la compréhension et la résolution de problèmes pris dans un sens très large, mais ces

[14] Repérage du principe d'équifinalité ou observation de la vicariance.

[15] C'est le cas, par exemple du projet architectural.

[16] Confrontation et traitement inévitable, au moins *« localement »*, en termes de contraintes.

[17] Par exemple, encore, en ce qui concerne l'architecture…

[18] L'art, les questions morales, les valeurs, les règles de vie dans une société donnée, la déontologie professionnelle, etc.

[19] Intuition probabiliste, modalités d'orientation de la pensée, etc.

[20] L'écrivain, tel M. Proust, percevant et décrivant une ambiance.

conduites peuvent servir aussi dans des applications et des élaborations mathématiques, dans l'analyse, voire la conceptualisation logique (Richard, 1990, Damasio, 1995). Elles servent, par ailleurs, en sciences humaines comme dans des activités qui mettent en rapport des personnes et dans lesquelles la compréhension de ce qui se passe, en interaction avec d'autres, relève de comportements ou de manières d'être « efficaces » qui renvoient, à côté du raisonnement, à des habiletés d'accès plus direct plus immédiat aux personnes et aux questions, aux « problèmes », qui se posent. Ces habiletés font appel à des notions encore difficiles à conceptualiser, déjà évoquées, de l'ordre du « sensible », de « l'empathie » pour lesquelles ont été développés des recherches anciennes ou plus récentes (par exemple, sur la question des *« neurones miroirs »*, Rizzolatti, Sinigaglia, 2008). Non seulement toutes ces approches du fonctionnement psychologique du sujet peuvent être convoquées par le chercheur de manière indépendante comme outils pertinents suivant les situations considérées, mais elles peuvent aussi s'inscrire dans des équilibres composés entre elles et avec le raisonnement, y compris de type hypothético-déductif. Les équilibres complexes qui en résultent sont à l'œuvre dans les découvertes comme dans les grandes avancées culturelles, scientifiques, esthétiques, éthiques, les évolutions politiques, sociétales[21]. Au point d'intégration de toutes ces dimensions, peut-il exister un nouvel équilibre[22], pour *le sujet épistémique*, comme pour *les sujets psychologiques*, ces derniers s'exprimant, alors, de nouveau[23], de manière plus ou moins partielle ou plus ou moins complète par rapport à ce nouvel équilibre d'ensemble ? Cet ensemble est-il susceptible d'être modélisé, ou pas, à sont tour dans une « logique » en termes de stades d'un développement à concevoir[24] dans une perspective plus vaste que celle relevant d'une certaine forme de la logique ? Et, notamment dans sa conception unitaire globale, potentiellement généraliste, plus particulièrement dans sa logique ordonnée, voire emboîtée, qu'elle sous-tendrait, est-ce la bonne perspective de compréhension des *sujets psychologiques*, pris dans la réalité complexe des situations rencontrées, et s'exprimant de manière plus ou moins pertinente, mais riche et variée, et, parfois même, au regard des modèles existants, de manière totalement *inattendue* (Pellois, 2005) ?

La réalité couramment observée est que les *sujets psychologiques* suivant leurs contextes propres s'expriment préférentiellement dans un registre de conduites ou dans un autre. Ils ne semblent disposer que très rarement, voire pas vraiment, de tous les registres. Et surtout ils ne sembleraient pas en disposer aisément de manière conjointe et équivalente. De ce fait, l'expression *des sujets psychologiques* n'est-elle pas, plutôt différentielle/plurielle que générale/univoque ? L'opération logique, le raisonnement, interviennent-ils, alors, dans toutes les formes de pensées, dans toutes celles sensées traiter efficacement de toutes les situations existantes que connaît l'humain, insérées dans toutes formes de contextes possibles ? Sont-ils incontournables quelle que soit la situation considérée ?

Le problème du *cadre conceptuel* potentiellement assimilateur, du *contexte de référence*, dans lequel est mise en œuvre l'expression cognitive, de son caractère étendu ou plus restreint[25], devient alors un élément central du choix entre option différentielle

[21] Par exemple, par le passé, la place et le rôle de la pensée philosophique au *« siècle des lumières »* ?
[22] Ou équilibration, ceci pour traiter sous l'angle de la problématique générale piagétienne.
[23] Mais est-ce l'apparence ou la (une ?) *« réalité »* des choses ?
[24] Serait-ce, alors, sur le mode piagétien ?
[25] Sur ce plan, par exemple, en mathématiques, qu'advient-il de la conception des nombres premiers,

radicale et différentielle hiérarchique. Plus le contexte est étroit et restreint, ou plus les règles de base qui régissent la structure du contexte conceptuel sont homogènes, ou plus ce contexte est strictement organisé selon un ensemble limité de règles de base[26], plus il paraît aisé de définir des performances permettant un classement hiérarchique[27]. A contrario, plus l'éventail des « habiletés » à maîtriser est envisagé largement, plus il est difficile de traiter des différences en termes hiérarchiques susceptibles de donner lieu à des mesures classables sur une échelle quantitative. Il peut être nécessaire, alors, comme évoqué dans les propos initiaux, de recourir à une description davantage qualitative, par exemple : en termes de profils, de « figures » ou de configurations[28] différentes, sans qu'un quelconque raisonnement puisse trancher sur la meilleure composition, en termes de qualité globale, valide pour tous les contextes.

En ce qui concerne les conceptions piagétiennes elles-mêmes, il y a, derrière cette réflexion, référant à des conceptions théoriques (Lautrey 1995) potentiellement novatrices, et les multiples constats, étayés par des faits, à comprendre, a minima, ce qui s'exprime dans la différence entre :

- l'application potentielle, plus ou moins générale, en fonction des sujets considérés, des contenus, des situations, des contextes, d'une opération logique, d'un raisonnement,
- et la généralisation, supposée, à tort, d'emblée acquise par la maîtrise formelle de l'opération considérée, de l'utilisation aisée de celle-ci, par tous les sujets y accédant, à toutes les situations rencontrées où elle s'appliquerait.

Certes, il peut exister des personnes à l'aise sur un ensemble assez vaste et très hétérogène d'expressions du *sujet psychologique*. Dans une perspective d'intégration selon le modèle piagétien[29], une conception homogène d'ensemble amènerait à penser qu'il pourrait y avoir un « stade » ultime de maîtrise à caractère universel, et donc des personnes susceptibles d'y accéder. Mais, si tel est le cas, celles-ci ne s'exprimeraient-elles pas, très probablement, tantôt dans tels ou tels domaines, tantôt dans tels autres, mais pas dans tous à la fois, et, si c'était le cas, pas avec le même niveau d'aisance dans tous les domaines considérés ? Ceci pose de nouveau la question du fonctionnel et de sa dynamique au regard du structural et celle de la compatibilité/non compatibilité, dans le développement, de dimensions de l'expression humaine au regard d'autres. Sur ce point, par exemple : le développement, voire l'apprentissage, de la pensée opératoire[30] et son exercice n'interfère-t-ils pas de manière négative, ou au moins sous une forme d'incompatibilité partielle, avec le développement, par exemple, des capacités d'empathie, et, vice-versa ? Tout comme, dans le domaine sportif, l'apprentissage du ping-pong peut être contrarié, par l'entraînement à des postures et des mouvements contradictoires ou opposés, tels que ceux relevant de l'apprentissage du tennis. Une

---

défini comme seulement décomposable que par lui-même et l'unité, dans un contexte où seul le concept de l'addition existe ?

[26] La logique propositionnelle binaire, un domaine particulier des mathématiques ?

[27] Celui qui maîtrise le mieux le raisonnement, la pensée opératoire piagétienne ou de manière plus ciblée encore l'opération d'implication (?), etc.

[28] Dans une approche multifactorielle, par exemple.

[29] (cf., par exemple, Piaget (1932), Piaget (1961)).

[30] De la pensée convergente, pour d'autres courants théoriques plus ou moins compatibles avec les conceptions piagétiennes.

bonne maîtrise des deux, relevant d'un équilibre instable, fragile et sophistiqué, ne permettrait alors pas ou pas nécessairement, une performance optimale pour chacune des deux activités sportives considérées…

Les conceptions parmi les plus récentes du développement (les développements ?) du *sujet psychologique* pencheraient davantage vers la conjonction plus ou moins étendue et plus ou moins harmonisée, d'habiletés « premières »[31], ou de compétences ou de capacités cognitives acquises, dont les performances se développent conjointement, dans l'entraînement sur contenus variés. Cette *conjonction* d'évolutions de ces habiletés, continues et progressives serait alors susceptible de permettre que se développe, de manière plus ou moins discontinue, des structures d'organisation de la pensée[32], connotées culturellement, comme l'expression de « produits » conceptuels. Ce raisonnement n'exclut pas, bien évidemment, de possibles effets *aller-retour*.

Cependant, vu sous cet angle même, l'outil CHIC est un outil d'analyse de données qui présente un progrès incontestable au regard du traitement statistique habituellement utilisé dans la recherche en psychologie, ceci à condition de l'utiliser dans un cadre méthodologique plus vaste suffisamment rigoureux et suffisamment approfondi sur le plan de la réflexion épistémologique, afin de donner toute la valeur, la validité interprétative scientifique, aux résultats obtenus par les calculs dans CHIC. Ce cadre méthodologique pourrait s'élargir, de manière pertinente, en passant de l'analyse quasi implicative, à caractère plutôt généraliste, à une analyse résolument différentielle telle que présentée dans l'exemple formel développé au point précédent, et qui traiterait, a minima des alternatives possibles dans des cheminements multiples et variés. Elle pourrait aussi s'articuler, avec d'autres modalités de calculs s'attelant aux traitements de formes *d'incompatibilité*, plus ou moins radicales, qui pourraient apparaître entre des voies parallèles. Ces voies parallèles pourraient, néanmoins, déboucher sur des capacités conjointes, mais, dissociées dans leurs modes de développement, multiples, plus ou moins compatibles, et, cependant, pouvant être complémentaires, dans des situations particulières, pour un même sujet, voire entre sujets.

Mais, dans ce cadre du traitement des données, l'interrogation subsiste sur le choix à faire entre une conception généraliste et des conceptions radicalement différentielles[33], appliquées au *sujet psychologique*. Et, si une conception « généraliste » était possible, de quel ordre serait-elle ?

Dans une vision qui serait, en première approximation, et peut-être seulement dans un premier temps, radicalement différentielle, il peut y avoir, autant de schémas-cadres, de contextes pour le chercheur, que les conceptions théoriques en développent, et qui permettent à la fois l'élaboration progressive et/puis l'utilisation plus ou moins générale c'est-à-dire polyvalente à différents contenus, à différents problèmes posés. Ces schémas-cadres, ces contextes, peuvent être, à leur tour, en articulation mais aussi en

[31] Mémoires, attention, espace mental à disposition, etc. Voir, par exemple, à ce sujet, Bideaud *et al.* (1993), et, de manière générale, les conceptions néopigétiennes, l'approche cognitiviste.

[32] Cf. par exemple, dans le modèle de Pascual Leone, l'effet du développement de l'espace mental (EM) à disposition sur le développement opératoire au sens piagétien du terme (Ribeaupierre (1980)).

[33] Inscrites non nécessairement dans la compétition d'un modèle plus pertinent qu'un autre, ces modèles n'étant pas nécessairement complémentaires, mais simplement différents voire contradictoires (incompatibles ?) entre eux… Pour l'aspect différentiel davantage configurationnel voir, par exemple, encore, Gardner (1996) Sternberg (1994).

opposition, compétition (etc.), entre eux dans *l'expression psychologique du sujet*, à la fois courante ou non courante, plus ou moins complexe et complète.

Dans ce contexte élargi, l'utilisation de CHIC nécessite alors de définir un (des) champ(s) particulier(s) qui se prête(nt) (plus ou moins ?) à un traitement en termes d'implications et d'implications de systèmes implicatifs. Par exemple, cela ne passe-t-il pas par la nécessité de poser un modèle théorique fonctionnel hypothétique alors nécessairement réduit ? Celui-ci serait conçu selon une logique implicative ou quasi-implicative dont il conviendrait, dans une réalité plus large, de vérifier la validité statistique[34] des résultats obtenus par observation et/ou expérience mise en œuvre au regard du modèle hypothétique initial (cf., par exemple, à ce sujet, Chamsine (2012)) incluant la dimension implicative.

En ce sens, l'interprétation de chemins implicatifs ou quasi-implicatifs comme de structures de relations entre variables et de structures de structures de relations (relations *cohésitives*) ne rend compte que de la configuration d'un système référant à une modélisation dans un champ d'observation compatible avec les conceptions de relations implicatives ou quasi-implicatives au sens du logiciel CHIC. Mais, même dans ce champ réduit, et pour une population considérée, le modèle sous-jacent et plus complexe retenu peut renvoyer tout autant à une situation hétérogène qu'à une situation homogène. Dans le cas de figure d'une situation hétérogène, des sous-groupes peuvent témoigner de « modélisations » possibles différentes entre elles et différentes du modèle reflétant la tendance générale du groupe dans sa globalité[35].

Au regard de ces différences une *réalité* nouvelle, y compris cognitive, est alors susceptible d'émerger au plan conceptuel. Cette *réalité* n'est pas véritablement assimilable aux conceptions formelles et épistémiques piagétiennes, mais elle les inclut en un ensemble dont la conception est encore à comprendre dans les différentes formes d'expression des *sujets psychologiques* en activité et se développant.

Ceci amène, de nouveau, à l'hypothèse qui suit. Il se pourrait que la réalité, vue dans un ensemble très large, de chaque *sujet psychologique*, qui s'exprime alors dans sa complétude, ne relève pas d'un développement progressif univoque et formalisé, à étapes organisationnelles structurées et emboîtées les unes par rapport aux autres, d'un *sujet épistémique* à caractère universel tout au moins au sens classique des positions généralistes habituelles. Ou, a minima, que cette « réalité » éventuelle, d'un nouvel universel ne relève pas *uniquement* ou spécialement d'un développement se faisant au travers de l'élaboration formelle de la logique propositionnelle aboutissant à une structure logicomathématique permettant un raisonnement hypothéticodéductif efficace sous certaines *contraintes*. Cependant, la conception initiale, limitée, liant *contraintes* et « sujet », plutôt de l'ordre d'une épistémique piagétienne réduite à une forme de la logique, pourrait bien relever effectivement, peut-être, d'une modélisation « implicative/cohésitive », par exemple, de type intelligence artificielle à l'œuvre dans le domaine de la robotique, de systèmes automatisés hautement complexes ou de systèmes experts. Il n'en demeurerait pas moins que la réalité du *sujet psychologique*,

[34] C'est-à-dire l'ordonnancement des données observées conformes à l'ordonnancement attendu.

[35] Chamsine (2012), en ce qui concerne les différences de résultats entre filles et garçons ou entre les élèves présentant une bonne ou une faible réussite scolaire.

pris cette fois-ci au sens générique, qui vit et qui s'exprime dans les complexités de la vie pourrait, elle, du fait d'un principe de vicariance plus général, ne pas relever de cela.

Relèverait-elle, pour autant, d'une conception bio/neuro/psychologique des mécanismes d'inhibition/facilitation[36], y compris dans l'imitation (Winnykamen, 1990), permettant une organisation choisie et progressive de la pensée volontaire comme outil des conduites adaptatives du sujet ? Cette conception renvoie à un sujet qui s'élaborerait progressivement à partir de l'émergence spontanée de schèmes multiples s'exprimant, en dehors de la volonté du sujet et sans organisation particulière, d'une manière d'emblée variée, donc dans des formes, y compris sophistiquées, « déjà là ». Elle se rapporterait alors, cette conception, à une modélisation en termes de « dressage éducatif », sous différentes formes, des conduites volontaires d'un tel sujet, ce qui peut renvoyer aux positions néo piagétiennes, potentiellement susceptibles de relever d'une certaine forme « d'éducabilité », et, de nouveau, à Houdé (1995).

Ou bien relèverait-elle des mécanismes de prise de conscience (Piaget, 1974a, 1974b) de schèmes plus ou moins sophistiqués apparaissant progressivement en fonction des maturations dans le développement biologique du sujet, maturations lui permettant, par l'organisation *réflexive* en retour, l'expression culturelle de la pensée opératoire, ceci dans une modélisation de l'élaboration éducative consciente progressive, d'un ensemble plus large ?

Ou bien relèverait-elle encore, cette *réalité* conceptuelle, d'émergences de « formes » telles qu'évoquées, par exemple, par Delacour (1998) ? sur la base d'une modélisation de type, tout à la fois, biopsychologique, systémique, dynamique constructiviste, a-représentatif et a-conscient[37], de l'ordre du sensible, et/ou cognitiviste représentatif et conscient dans un monde largement ouvert, permettant cette émergence, disons « tout azimut », de ces formes, y compris dans des réalités culturelles sophistiquées[38] ?

Aurait-elle à voir, plus largement avec les mécanismes efficaces d'orientation de la pensée inscrits dans des expressions cognitive et émotionnelle associées, ou plus largement encore, avec l'émergence spontanée complète de formes complexes… ? Tout cela en sachant, en arrière plan que le sujet se pensant lui-même pourrait bien relever du principe « d'incomplétude » (Gödel, 1971) ?…

Et pourquoi, in fine, toutes ces réalités du sujet psychologique ne fonctionneraient-elles pas conjointement ? A partir de cette conception élargie, qu'en est-il d'une vicariance, inscrite, à la fois dans ces *réalités* hétérogènes[39] et des processus d'homogénéisation, qui restent souvent locaux et partiels[40] ? Quelle part est faite, dans l'interprétation des calculs dans CHIC, non pas seulement en termes implicatifs, mais plutôt, en termes d'évolutions conjointes pas nécessairement interactives, aux

[36] Au plan bio-psycho-physiologique, Laborit (1980), et au plan de la psychologie du développement cognitif, Houdé (1995).

[37] Connexionnisme, système auto-poïétique, etc.

[38] Relevant aussi du prolongement de la psychologie de la forme, de conceptions philosophiques telles que la phénoménologie ?

[39] Les faits de cette réalité.

[40] Dans le raisonnement hypothético-déductif mis en œuvre au regard de l'ensemble des seize opérations de la logique propositionnelle binaire.

corrélations —c'est-à-dire à la covariation— non nécessairement interprétables au sens de l'interaction (circularité), ou sous la forme de *jeu(x) implicatif(s) associé(s)* ?

Et, donc, quelle place est faite, à côté du fonctionnement séquentiel, aux fonctionnements parallèles, éventuellement simultanés, et ce qui les met en œuvre ? Quelle place est faite, à côté de l'expression, bien sûr[41], de l'implication, à sa contraposée, mais aussi aux jeux d'équivalence, de complémentarité (le « ou » non disjonctif), mais également de compétition, d'incompatibilité (le « ou » disjonctif), entre différentes expressions et organisations psychiques du sujet, sujet abordé sous l'angle de la personne toute entière ?

Quelle part est faite aux jeux de convergence/divergence entre différentes organisations, d'harmonisation mais aussi d'accentuation des différences, voire de différenciation dynamique, donc, l'expression, l'observation, de l'accentuation d'une certaine indépendance évolutive, et potentiellement riche d'autres évolutions possibles[42] ?

A partir de cette ouverture très large, à un *sujet psychologique* aux multiples expressions en interactions, et conçu comme un tout, et le jeu de questions qui en découle ci-dessus, quel sens, alors, peut être donné à la fonction *cohésitive* du logiciel CHIC ? Celle-ci appliquée à 4 dimensions a, b, c, d, organisées en structure *cohésitive*, peut-elle correspondre, par exemple, à l'observation statistique suivante ? Une règle d'implication est attestée entre a et b et une règle d'implication est attestée entre c et d. Il est attesté une règle d'implication au second niveau traitant de l'implication entre les règles d'implication concernant a, b *et* c, d.

Si c'est le cas, cette structure de jeux de règles qui peut être ultérieurement de 3ème, 4ème niveau, incluant d'autres dimensions (e, f, i, etc.), serait compatible avec les conceptions d'organisation qui se structurent d'une certaine manière pour traiter de *contraintes* rencontrées dans un contexte donné plus ou moins fermé. Elle pourrait rendre-compte de modes adaptatifs d'un système à ces types de contextes. Il s'agit bien, ici, de l'examen, *sous contexte particulier*, de la réponse susceptible d'être adaptée *aux contraintes particulières posées*.

Mais, au regard de ce qui a été dit précédemment, cette *superstructure* peut exister à côté d'ouvertures fussent-elles marginales[43], mais rendues possibles par un *contexte* qui ne serait pas fermé au point de ne permettre qu'une voie adaptative et une seule. Les cas non conformes ne relèveraient pas alors d'erreurs de fluctuation de la mesure, seulement par le fait que ce qui vaudrait essentiellement, sur le plan des résultats du calcul dans CHIC, c'est, dans une optique potentiellement réductrice, la voie majoritaire, mais, plus ou moins, de relations différentielles *cachées*.

Il a bien été abordé, ici, dans ce texte, qu'il peut exister, pour les *sujets psychologiques*, d'autres contextes que ceux, particulièrement fermés, et qui subissent des contraintes maximales. Il peut exister des ouvertures possibles à d'autres *contextes* différents traitant des mêmes variables dans d'autres *contraintes* plus ou moins strictes,

[41] Mais de façon plus ou moins complète ou partielle (quasi-implication).

[42] Par ailleurs, dans le domaine des sciences de la vie, l'exemple de l'évolution de la nature dans ses mécanismes de diversification/adaptation aux différents milieux conjoints ou successifs.

[43] Du fait d'une certaine part d'observations minoritaires tolérée, au regard des observations majoritaires, par le principe même de la *quasi-implication*.

ou de ces mêmes variables dans des ouvertures non limitées à l'approche implicative/cohésitive. La littérature nous dit bien qu'il y a, dans le domaine de l'expression psychologique, des formes d'expression parallèles, complémentaires alternées, sinon alternatives, voire compensatrices (rappel, pour mémoire, des positions de Longeot (1978), Lautrey (1995)) des formes s'exprimant par inhibition/facilitation, plutôt univoques et abusivement généralistes, dans les conduites apprises, et plus ou moins volontaires, du sujet[44].

4.2 Vers l'expression d'un jeu de problématiques ?

In fine, dans une perspective d'ouverture plus complexe mais plus complète, **en fonction de contextes, donc de spécificités, de formes et de niveaux de contraintes** variées et variables, ne s'agirait-il pas, ici, rien moins que de l'évocation d'une forme de recherche, de conceptualisation à caractère épistémique, d'un objet de recherche pour des chercheurs en mathématiques associés aux spécialistes de questions et de *contextes* non strictement mathématiques ? Ne s'agit-il pas, a minima, de perspectives d'évolution de CHIC, incluant une démarche différentielle ?

En première approximation, la problématique formelle qui résulterait de cette perspective, initiée par les développements des différentes sections du texte qui précèdent, pourrait être posée de la manière suivante :

- Soit une « configuration[45] » caractérisée par une organisation de liaisons structurelles observées de manière quasi-implicative entre différentes variables, et caractérisant elle-même l'expression d'une qualité atteinte ou de la genèse de ce qui permet cette qualité réputée pertinente au traitement de « problèmes[46] » considérés, dans un *contexte* donné, et pour un ensemble de sujets donnés[47].
- Cette configuration renvoie à une confirmation statistique dans la majorité des cas amenant à observer **globalement** la tendance implicative dans le contexte donné, c'est-à-dire les *contraintes* propres à ce contexte.
- A contrario, quel sens donner à la minorité des cas ne confirmant pas l'hypothèse globale d'une structure implicative posée et majoritairement observée ?
- S'agit-il de parts d'approximation et d'erreurs de mesure, ou de fluctuations fortuites de l'expression de performances des sujets considérés dans l'expérience, relevant, donc, également, d'une forme d'erreurs d'appréciation de la performance « vraie » du sujet ?
- Ou s'agit-il de l'expression de voies s'exprimant de <u>manière différentielle</u> et caractérisant, pour une minorité de sujet, des possibilités d'organisation

[44] Par exemple : pour les néo-piagétiens, les perturbations de l'incompatibilité dépassées par inhibition/facilitation (Pascual Leone (1982), Houdé (1995)).

[45] Un « profil », un « type », un « style ».

[46] Terme pris, ici, au sens large.

[47] En prenant la métaphore ou l'image *« illustrative »* du saut en hauteur, une configuration *« type »* de l'expression efficace d'un sauteur en hauteur dans laquelle joue préférentiellement celle de la détente de la jambe et du pied, à côté d'autres considérées comme moins fondamentales pour cette configuration (?), tel que : *« vélocité »*, vitesse de course, coordination des gestes et de leurs enchaînement, qualité de la *« figure »* de saut, etc. (?).

vicariante, donc, renvoyant à une certaine pertinence plus ou moins accentuée dans ce *contexte*, mais sous une autre *organisation*[48] que celle s'exprimant majoritairement[49] ? Subsidiairement, cette *organisation* ou « *configuration* autre » pourrait être davantage pertinente dans un autre *contexte* donné[50], voire, d'autres formes ou niveaux de *contraintes*, pour un autre (ou le même) ensemble de sujets donnés[51].

– Il résulterait du point précédent qu'il peut exister des qualités de *configurations* plutôt non-pertinentes dans un *contexte*[52] —un ensemble de *contraintes* données— mais tout à fait pertinentes dans d'autres *contraintes*, un autre *contexte* ; des aspects de la *configuration* « type », pertinents dans le premier, ne l'étant pas dans le second[53]. Il pourrait même exister des *configurations* pertinentes dans des *contextes* faisant « obstacle » à l'expression de *configurations* pertinentes dans d'autres *contextes*, ce qui se traduirait par des jeux d'incompatibilités, susceptibles d'apparaître et de devoir être pris en compte, dans un système de *sujet épistémique* plus complexe, mais potentiellement apte à mieux approcher les réalités du *sujet psychologique* aux prises avec la complexité de ses *contextes*.

– Les propositions qui précèdent amènent à considérer que le traitement[54] de ces différentes *configurations*, envisagées, ici, dans une perspective large, et dans leurs rapports à l'expression et au développement[55] du *sujet psychologique*, peut déjà relever d'une approche désormais classique, et des méthodes statistiques qui la permettent (dont CHIC), à savoir le raisonnement hypothético-déductif à deux propositions tel que le conçoit, de manière très complète, mais dans un *contexte* limité, J. Piaget[56]. Mais, dans une observation très ouverte de ses conduites et comportements —y compris cognitifs donc—, si les configurations que met en œuvre le *sujet psychologique*, peuvent renvoyer à des formes de rapports relevant de l'implication ou de sa contraposée, elles peuvent aussi renvoyer à de la covariation conjointe, au « ou » disjonctif, ou au « ou » non disjonctif[57], voire à l'affirmation complète[58] dans le cas d'expressions statistiquement équilibrées de voies vicariantes.

---

[48] Voire même selon d'autres rapports qu'implicatifs.

[49] Dans Le même contexte *illustratif* du saut en hauteur, cela pourrait-il être une autre configuration « *type* » de l'expression efficace du saut, celle qui fait, par ailleurs, la performance d'un sauteur en longueur incluant l'accentuation particulière plus prononcée (jouant donc préférentiellement, mais aussi partiellement) d'une autre capacité, par exemple, celle de la vitesse de course ?

[50] Le saut en longueur, justement.

[51] Celui des coureurs de 100 mètres, par exemple ?

[52] Pour en revenir à l'analogie sportive déjà évoquée, le contexte du saut en hauteur, la configuration « *type* » du lanceur de poids… ?

[53] Soit, par exemple, et pour rester dans l'analogie sportive, du sauteur en hauteur au lanceur de poids, des aspects relevant de la morphologie, du poids, de formes de coordination des gestes et de séquences de gestes, etc.

[54] Méthodologique, en termes de recherches, de restitution de résultats, de connaissances, etc.

[55] Configurations considérées, dans les développements mêmes de ce texte, comme complexes autant dans leur conception que dans leurs rapports entre-elles, et susceptibles de caractériser, dans leurs rapports, des styles d'expressions et de développements variés.

[56] F. Longeot, à propos de son épreuve des T. O. F.

[57] Non compatibilité, dilemme ?

Par ailleurs, dans une approche très ouverte, le traitement de ces différentes *configurations* ne pourrait-il pas relever, aussi, d'autres formes logiques[59] ? Dans ce cas de figure, les conceptions ne référaient pas, ou pas uniquement à la notion du « nécessaire » et pourrait renvoyer à l'expression de conduites émergentes dans des espaces de liberté relevant de *contextes* conçus différemment[60] ? Tout ceci ne constituerait-il pas, alors, autant de défis, à inscrire dans la réflexion ultérieure portant sur la dynamique évolutive des conceptions autour et à propos de l'ASI ?

4.3 Pour conclure : le modèle piagétien, retour à un concept fondamental ?

Le psychologue, qu'il soit dans une logique *constructiviste* ou pas, observe et constate couramment que le *sujet psychologique* peut élaborer, formaliser, codifier, et même concevoir, puis utiliser les « produits » de son intelligence. Ces *produits* sont : les organisations formelles, structurées en langages abstraits complexes, opérantes et hautement sophistiquées, telles les mathématiques, la logique, mais aussi la grammaire, et, plus largement, « l'orchestration » des différentes voies de la communication humaine (Winkin, 1989), les règles de l'art, le langage cinématographique, etc. Mais il reste encore beaucoup à dire[61] sur ce qui a permis de développer ces *organisations* et ce qui permet de les mettre en œuvre, même si le psychologue repère des étapes partielles des « constructions » plus ou moins complètes, donc plus ou moins *ouvertes* ou *fermées*, et plus ou moins *locales* ainsi réalisées. Le *sujet psychologique* est, bien évidemment, en quelque sorte, la source de cette expression. Mais une question reste posée de manière persistante malgré les avancées déjà réalisées. Des *organisations* biologiques au *sujet psychologique*, en passant par les différents *contextes* qui ont rendu les choses possibles, quelles dimensions biologiques puis psychologiques[62] plus simples, ou moins simples, sont à l'œuvre et de quelle façon tout cela peut-il s'élaborer et s'exprimer ? In fine, ce qui a été dit ici permet, sans doute, de mieux saisir, dans une perspective d'ouverture, et face à toute conception *assimilatrice* trop hâtivement développée et donc potentiellement réductrice, en quoi la fonction *accommodatrice*, concept clef du modèle piagétien, censée s'appliquer à tous, et, donc, aux concepteurs de modèles eux-mêmes, a encore de belles perspectives devant elle !

Références

[1] Bideaud, J., O. Houdé, et J.-L. Pedinielli (1993), *L'homme en développement*, PUF, Paris.

[2] Bruchon-Schweitzer, M. (2002), *Psychologie de la santé. Modèles, concepts et méthodes*, DUNOD, Paris.

---

[58] Cas de contraintes faibles ou non existantes ?

[59] La logique ternaire, une forme possible, déjà mise en œuvre étant la logique floue, des logiques modales, etc.

[60] Fluctuations, turbulences, phénomènes d'émergences « spontanées » d'organisations dans des systèmes non linéaires dynamiques, chaos *« déterministes »*, rôle de *« l'intention »* du sujet, des sentiments et des principes qui l'animent, dimension auto-poïétique, etc.

[61] Même si la maîtrise de tels objets complexes, devenant des éléments contextuels pour l'esprit, peut permettre l'amélioration des capacités à s'en servir, comme des qualités qui les ont construites, et des évolutions vers de nouveaux objets encore plus complexes, ceci vers un nouveau développement du sujet psychologique...

[62] Comme mémoires, attention, représentations ?

[3] Chamsine, D. (2012), *Situation d'échec scolaire et sentiment d'échec scolaire de l'élève. Etude des interactions de divers facteurs qui entre en jeu*, Thèse de doctorat, Université de Caen.

[4] Damasio, A. R. (1995), *L'erreur de Descartes, la raison des émotions*, Odile Jacob, Paris.

[5] Delacour, J. (1998), *Une introduction aux neurosciences cognitives*, de Boeck Université, Paris.

[6] Gardner, H. (1996), *Les intelligences multiples. Pour changer l'école : la prise en compte des différentes formes d'intelligence*, Retz, Paris.

[7] Gödel, K. (1971), Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I., traduit en anglais par van Heijenoort. In *From Frege to Gödel*. Harvard University Press, 596-616.

[8] Gras, R. (2005), Panorama du développement de l'A.S.I. à partir de situations fondatrices, *Actes des Troisièmes Rencontres Internationales – A.S.I. Analyse Statistique Implicative*- Palermo, 6-9 octobre, 9-33.

[9] Gras, R. et R. Couturier (2010), Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association, *« Quaterni di Ricerca in Didattica (Mathématics) », Supplemento n.1, n.20, 2010*, G.R.I.M. (Department of Mathematics, University of Palermo, Italy) A.S.I. 5 Proceedings 5-7- November 2010, 175-200.

[10] Houdé, O. (1995), *Rationalité développement et inhibition. Un nouveau cadre d'analyse*, PUF, Paris.

[11] Houdé, O., B. Mazoyer, et N. Tzourio-Mazoyer (2002), *Cerveau et psychologie. Introduction à l'imagerie cérébrale anatomique et fonctionnelle*, PUF, Paris.

[12] Huteau, M. (1980), Dépendance indépendance à l'égard du champ et développement de la pensée opératoire, *Archives de psychologie*, XLVIII (184), 1-40.

[13] Inhelder, B., G. Cellérier, G., et Coll. (1992), *Le cheminement des découvertes de l'enfant*, Delachaux & Niestlé, Neuchâtel.

[14] Laborit, H. (1986), *L'inhibition de l'action biologie comportementale et physiologie pathologie*, Masson, Paris.

[15] Lautrey, J. et D. Chartier (1987), Images mentales de transformations et opérations cognitives: une revue critique des études développementales, *L'Année Psychologique*, 87, 581-602.

[16] Lautrey, J. (sous la direction de) (1995), *Universel et différentiel en psychologie*, PUF, Paris.

[17] Longeot, F. (1978), *Les stades opératoires de Piaget et les facteurs de l'intelligence*, PUG, Paris.

[18] Pascual Leone, J. (1982) Growing into human maturity towards a meta subjective theorie of adulte hood stages, *Report York University*, Department of psychology, 120.

[19] Pellois, C. (2005), Complexité et développement : reconnaître l'émergence de qualités nouvelles dans le domaine des conduites adaptatives. Difficultés et intérêts des nouvelles perspectives, *Cahiers de la MRSH*, n° spécial, université Caen Basse-Normandie, avril, 45-68.

[20] Pellois, C. (2007), L'utilisation des modèles structuraux afin d'aborder la complexité du réel en sciences humaines, *Mathématiques et Sciences Humaines*, 45 (177), 53-85.

[21] Pellois, C. (2010), Sens et incertitude, une forme de complexité en psychologie : des contraintes aux parts de liberté, le développement et ses contextes. In *Traitement de la complexité dans les sciences humaines*, Cadet, C., et G. Chasseigne, Editions Publibook Université, 177-207.
[22] Pellois, C. (2012), De l'observation de données à leur traitement par l'analyse statistique implicative : quelques constats, In *L'analyse statistique implicative : de l'exploratoire au confirmatoire*, Régnier, J.-C., Bailleul, M., Gras, R. (Eds), Actes du colloque, Caen, 7-12 novembre 2012, 87-117.
[23] Pellois, C. (2013a), Distributions plus ou moins complexes de données : contextes, et sens. In *L'analyse statistique implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.* Gras, R. (dir.), Gras, R., Régnier, J.-C., Marinica, C., Guillet, F (Eds), Cépaduès Editions, Toulouse.
[24] Pellois, C. (2013b), La psychologie et l'usage du traitement mathématique des données statistiques. Nouvelles perspectives conceptuelles, *Revista Brasiliera de Ensino de Cienca e Technologia*, 6 (1), 230-259.
[25] Pellois, C. (2015), Le sujet psychologique : la complexité différentielle de son expression et de son développement au regard du traitement statistique de données, *Actes de A.S.I. 8*, Radès (Tunisie).
[26] Piaget, J. (1932), *Le jugement moral chez l'enfant.* Delachaux et Niestlé, Neuchâtel.
[27] Piaget, J. (1961), *Les mécanismes perceptifs : modèles probabilistes, analyse génétique, relations avec l'intelligence*, PUF, Paris.
[28] Piaget, J. (1972), *Essai de logique opératoire*, 2e Ed. révisée par Jean-Blaise Grize, Dunod, Paris.
[29] Piaget, J. (1974a), *La prise de conscience*, PUF, Paris.
[30] Piaget, J. (1974b), *Réussir et comprendre*, PUF, Paris.
[31] Reuchlin, M., J. Lautrey, C. Marendaz, et T. Ohlmann (sous la direction de) (1990), *Cognition : l'individuel et l'universel*, PUF, Paris.
[32] Richard, J.-F. (1990), *Les activités mentales, comprendre, raisonner, trouver des solutions*, Armand Colin, Paris.
[33] Ribeaupierre, A. de (1980), Application d'un modèle néo-piagétien à l'étude du stade des opérations formelles, *Bulletin de psychologie*, XXXIII (345), 699-709.
[34] Rizzolatti, G. et C. Sinigaglia (2008), *Les neurones miroirs*, Odile Jacob, Paris.
[35] Sternberg, R. J. (1994), La théorie triarchique de l'intelligence, *l'Orientation Scolaire et Professionnelle*, 23 (1), 119-136.
[36] Winkin, Y. (1989), *La nouvelle communication*, Éd. du SEUIL, Paris.
[37] Winnykamen, F. (1990), *Apprendre en imitant ?*, PUF, Paris.

LE SUJET PSYCHOLOGIQUE : LA COMPLEXITE DIFFERENTIELLE DE SON EXPRESSION ET DE SON DEVELOPPEMENT AU REGARD DU TRAITEMENT STATISTIQUE DE DONNEES

Christian PELLOIS[1]

PSYCHOLOGICAL PERSON: DIFFERENTIAL COMPLEXITY OF ITS EXPRESSION AND DEVELOPMENT IN THE LIGHT OF STATISTICAL DATA PROCESSING

RÉSUMÉ

Au regard du modèle épistémique, l'approche différentielle des sujets psychologiques amène à concevoir l'utilisation de l'analyse implicative dans une démarche méthodologique de recherche plus large permettant de traiter des complexités de l'expression et du développement différentiels de ces sujets psychologiques, notamment, mais pas seulement, sous l'angle de voies de développement et d'expression alternatives. Cette méthodologie peut se traduire, sous une forme différentielle restreinte, par une modélisation formelle du développement, associant structure et fonctions mises en œuvre. Cette modélisation est présentée ici à titre d'exemple. Elle articule logique implicative, conception généraliste, et approche différentielle vicariante.

Mots-clés : *sujet, différence, psychologique, voies, développement, expressions, analyse, statistique, implicative.*

ABSTRACT

Compared with .the epistemic pattern, the differential approach of psychological persons induce to imagine the use of the implicative analysis in a methodological approach to broader research to address the complexities of expression and differential development of these psychological persons, including, but not only, in terms of alternative ways of development and expressions. This methodology can lead, in a restricted differential form, by a formal modeling of development, combining structure and functions implemented. This modeling is presented here as an example. It articulates implicative logic, general design, and vicarious differential approach.

Keywords*: person, difference, psychological, ways, development, expressions, analysis, statistical, implicative.*

1 Introduction

Traiter de l'analyse implicative en tant que tel, sur le plan formel, est une chose. Traiter de son application à la psychologie en particulier celle du développement, notamment sous l'angle différentiel au regard des conceptions déjà anciennes ou plus récentes, en est une autre. Ceci nécessite au préalable une réflexion approfondie et des modalités d'application du logiciel CHIC s'inscrivant dans une réflexion méthodologique de qualité qui passe, tout d'abord, par une connaissance plus précise

[1] 7 rue des pins, 14470, Courseulles sur mer, pellois.christian@wanadoo.fr

des problématiques du domaine considéré, et ensuite par la mise en œuvre de procédures d'application de CHIC tenant compte de ses conceptions formelles et des réalités multiples et complexes du domaine d'investigation considéré, ici, la psychologie dans ses multiples dimensions.

Dans le domaine du développement psychologique de l'enfant, il y a eu l'émergence, à côté d'autres modèles du développement, des conceptions piagétiennes. Il y a eu, ensuite, ou parallèlement, les recherches et conceptions multiples et variées de la psychologie post-piagétienne, néo-piagétienne, cognitive au sens large, psycho dynamique telles que exposées dans Bideaud *et al.* (1993), ou autres, en particulier les conceptions différentielles, notamment dans leurs développements déjà anciens ou plus récents. Ces différentes approches soulèvent toute une série de question et en particulier la question suivante. Au regard de ces différentes conceptions, le modèle de l'expression du développement humain est-il simplement hiérarchique, général et emboîté, ou est-il à envisager dans une forme plus complexe, au moins vicariante, et donc différentielle (voir, par exemple, à ce sujet, Pellois (2002), p. 214 et suivantes) ? Répondre à cette question passe par une discussion de l'approche généraliste, également évoquée par ailleurs (Gras et Acioly-Régnier (2015)), dans ses fondements conceptuels, tel que conçu, dans ses formes initiales, par la théorie piagétienne, mais aussi, au regard d'une psychologie différentielle plus ou moins articulée à l'approche généraliste. Cette seconde perspective fait l'objet du texte qui suit. Ce texte plus théorique et conceptuel se prolonge par un autre texte (Pellois (2015)) développant la réflexion sur la base plus concrète d'un exemple d'utilisation de CHIC tenant compte des complexités différentielles du domaine considéré.

Pour en revenir au plan théorique, tenter de répondre à la question différentielle ci-dessus peut conduire à une réflexion portant sur des schémas conceptuels différents de ceux du cadre piagétien initial et, donc, par le biais de cette réflexion, à une autre *« manière »* d'utiliser CHIC, dans une méthodologie de recherche qui concevrait son utilisation plutôt comme *« outil »* que comme *« modèle »*. Mais dans quel cadre conceptuel peut s'inscrire de telles perspectives ?

2 Un nouveau cadre conceptuel ?

2.1 La vicariance : des variations plus ou moins compatibles avec une même forme organisée : structure, « configuration » qualitative ?

Ce « nouveau cadre » conceptuel, déjà en partie connu et exploré (voir, notamment, à ce sujet, Lautrey, (1995)), amène à postuler, au départ, que la *forme* constituée par un ensemble d'observations, la *« réalité »* observée sur une population de sujet, ne renvoie pas nécessairement, et aux erreurs près, à *une organisation*, ici cognitive, univoque générale et commune à tous les sujets considérés, mais peut renvoyer à un ensemble *d'organisations* plus « locales », temporaires ou non, et plus ou moins comparables, dans le développement et/ou l'expression cognitive. Ces *organisations* relèveraient de « configurations » sujet/situation, peut-être, au sens, par exemple, de Leenhardt (2005), p. 63, ou celui de Perrot *et al.* (2006), plus ou moins assimilables à un modèle général, ou radicalement différentes (voir, par exemple, sur ce point, Sternberg (1994), ou, sous une approche encore plus radicale, Gardner (1996)). Au regard de l'expression cognitive ou de son développement, ces « configurations » cognitives « locales », plus

ou moins partagées, par des sujets, sont mobilisées —l'une ou l'autre pour le même sujet, par un sujet ou par un autre— dans des situations de contraintes générales internes/externes plus ou moins différentes. Il en résulte des capacités adaptatives intra et/ou interindividuelles plus ou moins efficaces, de manière complexe. Elles le sont en fonction *de ces formes organisationnelles* mais aussi de leurs *modalités d'usage* (modulation, rigidité/flexibilité, etc.) par des sujets dans les contextes variés auxquels ces configurations « sujet/contexte » sont appliquées, et dont les conceptions qui tentent d'en rendre compte peuvent relever tant du structural que du procédural et du fonctionnel, comme cela est présenté dans Lautrey (1995), voire de *systèmes dynamiques*, et du non-linéaire, comme l'envisage Juhel (2005). Il se déduit de cette position très générale différentes perspectives.

Il en est ainsi de la *vicariance*, sous des formes alternatives et potentiellement complémentaires. Cette position tolère que, dans un contexte donné, et des contraintes propres à ce contexte, si une tendance des ***évènements*** observés traduit le caractère similaire et comparable des cheminements ou expressions de sujets, il peut exister des expressions ou des cheminements plus ou moins différents relevant pour autant du même schéma *organisationnel* général. Il en serait ainsi, par exemple, du développement puis de l'expression de la pensée hypothético-déductive, référant à une *organisation* formelle ouverte du modèle Piagétien. En effet, cette première approche différentielle, telle que défendue par Longeot (1978), est compatible avec l'observation d'une tendance schématique générale résumant les grandes étapes des cheminements du développement, ou de l'expression, sous une forme, ici implicative, *tout en rendant compte d'une succession d'observations différentielles*. Celles-ci ne relèveraient pas, toujours selon Longeot, ou, tout au moins, transitoirement pas, en fonction de l'expression « locale » sujet/contexte, d'un sous-schéma univoque de développement (évolution structurale différentielle).

2.2 De la forme organisée aux performances qui la permettent : la question de la moyenne des performances quantitative observées et de leurs variations.

Il en est ainsi, également, à un niveau de développement atteint, de l'observation d'évolutions de *performances* et non plus *d'organisations*, que ce soit dans le domaine cognitif, adaptatif au sens large, sportif, esthétique, social, éthique, ou autre. L'attention serait portée, ici, sur la question de la *moyenne d'évolutions de performances individuelle*s, mettant en œuvre des fonctions (mémoire, attention, etc.) antérieurement acquises et qui voient leur efficacité progressivement s'améliorer en termes : d'aisance d'utilisation, de rapidité d'exécution, ou de généralisation élargie à des contextes variés (évolution fonctionnelle).

Il en est de même pour *l'augmentation plus ou moins brusque de la moyenne des évolutions* de ces mêmes *performances* individuelles, à partir d'une nouvelle étape[2] d'une *organisation*, par exemple, cognitive, acquise sur un mode différentiel, vertical comme horizontal, du développement tel que cela est déjà envisagé dans la conception piagétienne. Cette nouvelle étape, verrait s'améliorer progressivement les adaptations antérieures à des situations cognitives plus variées[3], donc, sous d'autres formes, les

[2] Pouvant amener, par ailleurs, à un nouveau cadre de raisonnement, à de nouvelles compétences.
[3] Donc à d'autres contenus, d'autres situations.

performances, par l'aisance de l'expression, l'étendue d'application, la rapidité d'utilisation (etc.) accrues, de certaines fonctions (A) opératoires piagétiennes ou de type néo piagétien, post-piagétien, tel que présenté, par exemple, par Houdé (1992), chapitre II, Bideaud , *et al.* (1993), notamment pages 93 à 127, 321 et suivantes, 379 et suivantes, 492 à 497. Cette nouvelle étape (stade du développement), mieux adaptée à la complexité grandissante des situations/contextes, ou à l'apparition de nouvelles situations/contextes non abordables par les processus antérieurement acquis, permettraient, l'ajustement *qualitativement* « pertinent » à des situations cognitives radicalement nouvelles. Elle pourrait permettre alors l'expression de procédures différentielles, elles-mêmes, pour une part au moins, nouvelles, mais plus ou moins adaptées, suivant les contextes. Celles-ci pourraient, également, améliorer, plus ou moins, au regard des contenus traités, le score des *performances* mesurées des fonctions et procédures (B), avec possible effet retour sur le développement (évolution développementale structuro/procéduro/fonctionnelle).

Ces deux conceptions schématiques, hypothétiques complémentaires (A) et (B), acceptent parfaitement, au plan théorique : des cheminements individuels qualitativement différents, mais caractérisant les mêmes évolutions quantitatives positives de *performances* que l'évolution *moyenne* observée pour un groupe considéré, ou des évolutions quantitatives de performances individuelles différentes, ramenées à la même *moyenne* d'évolution de groupe, ces différences individuelles étant liées ou pas aux mêmes schémas qualitatifs de l'expression ! Ne sommes-nous pas, ici, déjà bien loin de cette représentation simplifiée, mais courante, d'évolutions de grandeurs ou de mesures d'évolutions supposées plutôt univoques et positives (progressions), et cependant variées, mais considérées comme sensiblement comparables, aux erreurs de mesure près, parce que s'inscrivant dans un même cadre généraliste ?

Cependant, à ce stade de la réflexion, il reste encore un a priori qui peut être considéré comme abusivement généraliste, y compris dans ces deux dernières perspectives élargies. Les évolutions plus ou moins importantes, mais comparables, *sur la base d'une norme moyenne*, réfèreraient toujours, ici, à une même *configuration générique* globale, associant structure et sous-structures, procédures et fonctions dans le développement et/ou l'expression « cognitive ». Cette *configuration* prise dans sa conception la plus large, irait jusqu'à inclure les dimensions de la personnalité, des sensations, des émotions, de l'orientation de l'action, du jugement, cette liste n'étant pas limitative. Pourtant ne peut être exclue l'hypothèse selon laquelle les évolutions observées pour chaque sujet présentent des formes variées plus radicalement différentielles, c'est-à-dire relevant de *configurations différentes*.

2.3 Une forme organisée, globale, ou plusieurs, plus « locales » ? Homogénéité ou hétérogénéité la question du cadre conceptuel de traitement de données

En fonction des développements ci-dessus, si observer, derrière une tendance d'évolution positive des cheminements individuels présentant des modes de progression différents[4], peut amener à considérer qu'il existe une majorité de variabilités individuelles relevant de la même dynamique *configurationnelle*, cela peut aussi amener à considérer, au regard d'observations particulières (progressions accentuées, ou très

[4] Très faibles et en continu, linéaire, exponentielle, à pente variable, plus ou moins ondulatoire, fluctuante, voire chaotique.

réduites, stagnations, régressions), que les configurations sous-jacentes considérées *ne sont pas nécessairement les mêmes*. Ce dernier point de vue consisterait, alors, ni plus ni moins, à substituer un cadre conceptuel hypothétique d'hétérogénéité latente à celui d'une homogénéité latente postulée. Au bout de ce raisonnement, et dans ce cas de figure, chaque réalité d'évolution différente des autres ne peut-elle aller jusqu'à constituer un *évènement* singulier en soi relevant d'une configuration sujet/personne particulière ? Cette position, davantage inscrite dans des approches de psychologie clinique, dynamique, est, ainsi, bien radicalement opposée à la tendance globale constitutive d'une norme statistique ou même conceptuelle à laquelle seraient systématiquement ramenés un ensemble d'*évènements*, plus ou moins radicalement différents.

Dans cette perspective différentialiste radicale[5], une progression, pour un sujet donné, plus rapide ou plus lente que la moyenne, soudaine pour un autre[6], une stagnation pour un troisième, une involution de performance pour un quatrième constitueraient autant d'*évènements* singuliers non assimilables à une évolution *« moyenne »* et relèveraient d'évolutions de configurations *sujet/contexte* différentes…

Cependant, malgré la variabilité interindividuelle toujours constatée, rien ne justifie non plus, a priori, que le chercheur doive adopter, de manière systématique, le cadre différentiel et l'hypothèse d'hétérogénéité. En effet, la tendance générale *moyenne* peut, témoigner de l'*évènement* d'une évolution générale, commune, à tous les sujets, et dite *« normale »* au sens statistique du terme. Cette évolution constatée correspondrait à l'expression et au développement d'une configuration observée chez tous les sujets[7], certes modulée différemment, par les uns et par les autres, mais sensiblement identique, donc, compatible avec l'hypothèse d'homogénéité.

Mais, pour autant, le constat de la tendance générale, considérée, par exemple, celle d'une *moyenne de performances*, si elle peut amener à affirmer ce qui précède, au nom d'une considération conceptuelle univoque, ne saurait conduire à assimiler aux mêmes bases d'analyse évènementielle la variabilité des performances et de leurs évolutions observées au regard de modélisations alternatives (vicariance) ou radicalement différentes.

Il s'en déduit que ***sans maîtrise de la totalité du contexte interne/externe*** latent, sur le plan conceptuel et méthodologique, ***il convient formellement de supposer*** qu'aucune position, qu'elle soit généraliste, modulatrice, vicariante ou radicalement différentielle ne peut être prise avec certitude.

L'interprétation, selon l'une des options évoquées ci-dessus, par exemple de *chaque évolution* originale contribuant à la *pente moyenne des évolutions* peut donc renvoyer à une prise de risque conceptuel en termes d'appréciation de la situation considérée.

La position différentielle peut empêcher le repérage et la compréhension conceptuelle d'une grande tendance générale sous-jacente aux évolutions communes. L'acceptation d'une même configuration fonctionnelle hypothétique, aux irrégularités

[5] Multiplicité des configurations des sujets considérés.

[6] Rupture considérable de pente de l'évolution.

[7] Donc elle aussi *« normale »*, pour une population considérée, au sens d'une norme statistique d'évolution.

d'expression près[8], peut, elle, faire *obstacle*, pour un observateur donné, ne connaissant pas totalement le système configurationnel *contexte/sujet* mis en jeu, à la reconnaissance, dans des fluctuations interindividuelles de *performances*, ou *d'évolutions de performances*, de « signes » particuliers alertant sur des fonctionnements originaux, « configurationnellement » différents ou sur des dysfonctionnements, dans le domaine de la pathologie par exemple.

Il conviendrait, alors, de renvoyer les observations faites de ces différences marginales d'évolutions, plutôt, à : des *évènements* internes/externes[9] singuliers, remarquables, et donc inattendus[10], des émergences *extraordinaires*, témoignant de quelque chose de différent, renvoyant à des configurations différentes, voire nouvelles.

Dans l'analyse implicative et par les résultats fournis par CHIC, c'est ce que permet aussi de cerner le phénomène[11] de la plus ou moins grande quasi-implication, attirant l'attention tout autant ***sur des situations marginales, non nécessairement dues aux erreurs de mesure***, que sur la tendance générale[12] observée pour le groupe considéré dans son ensemble.

La singularité « vraie » d'une observation, que ce soi une *performance* plus ou moins différente ou l'émergence d'un *évènement* par exemple une réalité logique observée différente, peut donc être la bonne hypothèse du chercheur. Elle *peut l'être*, mais ne l'est pas nécessairement. C'est toute la difficulté de l'interprétation des résultats de recherche.

C'est aussi tout le problème de *la moyenne des évolutions de performances*, validant une certaine forme « type » *d'évolution moyenne*, au regard de chacune des évolutions, toutes différentes. Tel est le cas, dans le domaine de la croissance physique, des *performances* sportives, ou celles relevant de l'évolution cognitive. Il en est ainsi des phénomènes de ***rattrapage*** après des évolutions moindres, voire des stagnations momentanées, ou des régressions. Il en va de même pour des *évènements* tels qu'un rapport logique, par exemple entre deux variables, majoritairement observé, à côté d'autres rapports logiques minoritaires variés.

Lorsque l'interprétation se fait au départ sur le mode d'un schéma *configurationnel*, c'est l'analyse des fonctionnalités externes/internes, dans une conception connue mais pouvant être élargie, à terme, à une élaboration théorique ouverte différente, aboutissant à une meilleure maîtrise conceptuelle contexte/sujet qui permettrait de trancher entre :

a) ce qui relèverait d'un fonctionnement évolutif variable mais lié à une forme changeante de rythme ou de forme dans l'expression de la même configuration, du même évènement, changement de rythme qui ne serait pas *alarmant* (le signe de quelque chose) c'est-à-dire ne constituerait pas une *alerte* pour le chercheur,

[8] *« Irrégularités »* qu'il conviendrait de comprendre dans le sens d'une part de liberté *ordinaire* du système considéré.

[9] Et/ou des configurations de sujets, ou de contextes, ou mixtes, sujet(s)/contexte(s).

[10] Dans le sens de la *sérendipité* inhérente à des résultats de recherche.

[11] L'observation indirectement réalisée par traitement statistique interposé ? La réalité effectivement observée ?

[12] *« Moyenne »* selon le raisonnement analogique portant sur la performance et ci-dessus envisagé pour mieux se faire comprendre ?

b) et ce qui relèverait d'un fonctionnement original, ou d'un dysfonctionnement, qui attirerait l'attention sur un aspect susceptible d'entraîner justement *alerte* et *alarme*, au regard de l'expression d'une autre configuration, d'un autre évènement.

Cette position interpelle bien la persistance trop systématique à renvoyer une accélération brutale d'une évolution[13], au-delà du jeu des fluctuations normales connues[14] d'une *« organisation »*[15], à des fluctuations supposées *« normales »* de cette *« organisation »* (assimilation), et non pas à la reconnaissance d'un *signe* de quelque chose, au plan d'une *« organisation »*, de différent, voire, « du nouveau » qui émerge (accommodation). Vu sous l'angle des deux options *a* et *b* évoquées ci-dessus, cette persistance traduirait davantage des réticences, des résistances aux changements, ou l'ignorance, voire le refus d'admettre la complexité inhérente à un *système*, plus vaste, et non connu sous toutes ses formes. Cela revient à se placer, au plan méthodologique, dans la position, a priori, de simplification abusive, de réduction hypothétique d'un système, ***alors que ne sont pas connus précisément dans le détail, et dans toute son étendue, le fonctionnement de toutes les réalités possibles*** du *système* considéré.

Au plan formel de la méthodologie, dans le cas de la position *de fluctuations « normales »* acceptables, il ne peut y avoir simplement que la reconnaissance, ***jusqu'à preuve du contraire***, d'une *non-connaissance* d'autres configurations possibles entrant en jeu.

Dans le cas de *l'émergence originale*, reconnue par le chercheur, la décision serait prise, a contrario, d'opter pour cette position, en recherchant les raisons précises, novatrices mais de manière certes hypothétique, explicitables, raisons qui alerteraient sur quelque chose qu'il conviendrait d'approfondir.

Tout cela relève bien, en dehors de l'observation statistique, du *contexte d'interprétation* et des *connaissances*, ou parts de *non-connaissances*, que l'on a ou pas, du *système* dans son ensemble et dans le détail de ses fonctionnements. C'est donc le chercheur qui choisit le sens du *« signe »* qu'il reçoit. Ce qui l'amène à considérer qu'un *« évènement »*, une occurrence, une cooccurrence, une *« exception »* dans l'observation, est fortuite ou bien qu'un évènement qui émerge est consécutif à l'expression d'une variable latente nouvelle, ou de relations d'un nouveau type, entre éléments du *système* considéré. Ce *« non connu »* peut amener à modifier la compréhension que l'on a de ce *système* par l'expression d'une conceptualisation originale, donc nouvelle, d'une chaîne d'occurrences emboîtées ou organisées les unes par rapport aux autres, d'une manière particulière, différente de ce qui était considéré antérieurement.

2.4 Complexité de l'approche du développement conçu de manière hétérogène, voire sous des aspects plus ou moins compatibles : du *sujet épistémique* au *sujet psychologique* pris sous toutes les formes organisées dont il témoigne

Il en serait ainsi dans le cas de l'analyse quasi-implicative pour des observations d'implication inverse, mais aussi, de manière moins visible donc plus insidieuse, d'observations de co-variations non assimilables au jeu de l'implication directe, voire

[13] Comme, par exemple, la croissance, l'amélioration, comme la chute, la stagnation, ou des irrégularités soudaines d'expression de performances cognitives.

[14] « Erreurs » de mesure, variations explicables dans un système par ailleurs bien connu.

[15] Terme générique employé ici et renvoyant à toutes formes de structure, système, configuration, etc.

des observations d'incompatibilités caractérisant des jeux contradictoires, de configurations, ou sous-configurations, potentiellement concurrentielles. Ces évènements particuliers seraient, en quelque sorte, *« cachés »* sous la tendance générale du nuage de points[16] représentant la variabilité des occurrences observées autour de la tendance implicative générale perceptible par l'observation directe ou par le traitement par calcul (CHIC) de l'ensemble des co-mesures effectuées.

Dans l'esprit de la problématique esquissée dans les développements qui précèdent[17], une opérationnalisation formelle, partant de l'approche piagétienne initiale, mais prenant en compte les avancées des recherches en termes de développement de l'enfant et de l'adolescent et tenant compte du schéma-cadre méthodologique ci-dessus, peut être envisagée. Il se trouve, en effet, que, le raisonnement opératoire piagétien renvoyant à *un sujet épistémique*-en développement selon un mode implicatif strict peut être appliqué, au moins partiellement, dans certaines conditions, référant notamment à la situation scolaire (cf. notamment, à ce sujet, Pellois (2013)), et sous une certaine forme limitée, de son expression cognitive convergente, à *un sujet psychologique*. Cela ne peut se faire cependant qu'à l'intérieur d'un cadre formalisé de manière schématique, et dans une vision encore restreinte (voir ci-dessous) de l'expression de la personne, mais une vision moins restreinte que celle relevant uniquement du modèle logique d'organisation formelle de la pensée opératoire piagétienne pris au sens strict, et, a fortiori, de la seule opération logique de l'implication.

Il s'agit de la « vision » d'un *sujet psychologique*, certes, toujours conçu dans l'approche objectivée d'une pensée convergente aux prises avec ses réalités internes/externes, mais dont les réalités de l'expression de la pensée sont abordées sous un angle fonctionnel plus large que les conceptions piagétiennes, tout au moins initiales. Ce cadre inclut également, et se limite aussi, à des conceptions objectives récentes telles que l'envisagent notamment les sciences humaines, dont la psychologie cognitive, mais aussi les neurosciences (voir, par exemple, au regard de ces deux cadres, Huteau et Lautrey (1999), Houde *et al.* (2002)). Ce qui élargit mais également réduit toujours le cadre d'un sujet pris dans son expression complète agissant dans ses multiples contextes. En effet, ces approches objectivées, même les plus récentes, n'incluent pas des conceptions psychologiques anciennes ou actuelles, plus cliniques ou à l'articulation de l'objectif et du subjectif. Elles n'incluent pas des conceptions, par exemple inscrites davantage dans un cadre relevant de la phénoménologie et/ou du subjectif pour lui-même et de l'intersubjectif, c'est-à-dire renvoyant à la définition d'un champ conceptuel, peut-être au sens de A. Vergnaud (cf. Gras et Acioly-Régnier (2015)), par certains aspects, largement différent. Dans une vision ouverte de l'approche psychologique de la personne humaine, ces approches pourraient bien avoir leurs pertinences à côté de celles renvoyant aux conceptions objectives classiques ou moins classiques (par exemple : Delacour (1998), Varela (1993)). Si tel est le cas, ce ne peut être qu'à condition de reconnaître, tout à la fois, pour la personne inscrite dans sa subjectivité propre et singulière, la radicalité des différences des points de vue de chacun et leur pertinence (momentanée ?), au moins au regard de la réalité subjective de celui qui l'exprime. En effet, cette psychologie, comme la réflexion épistémologique l'accompagnant, traite, pour avancer dans leur conceptualisation, des rapports de

[16] Correspondant aux occurrences d'observations/mesures.

[17] Et sous l'angle de la réflexion épistémologique maintenant bien connue à laquelle elle fait, pour partie, référence, telle que exposée, notamment, par Bachelard (1987).

complémentarité, mais aussi de disjonction ou même de contradiction, voire de formes *« a-rationnelles »* opérantes, y compris par rapport à des aspects davantage rationnels et objectifs (par exemple : Damasio (1995), Bandura (2003), Olry-Louis *et al.* (2013)). Ces postures, élargies à ce qui dissocie ou oppose, est susceptible d'amener jusqu'aux confrontations de positions des sujets comme des différents courants de pensée considérés. Elles peuvent amener, à la juxtaposition des points de vue, montrant par là que chaque sujet ne soumet pas sa réflexion, son raisonnement, à celui de l'autre (par exemple, C. Pellois (2014), notamment, tome 1, p. 48 et suivantes, p. 271, p. 286, tome 2, p. 7, p. 205, dans ouvrage collectif) et donc pas totalement à la pensée de celui qui a conçu un courant de pensée donné. Ceci renvoie davantage à des logiques de confrontation, de conflit et de compétition entre *modèles* « locaux », mais aussi de tolérance sans inféodation abusive. Elles défendent, ces postures, la reconnaissance de différences non assimilables les unes aux autres, et le fait que chacune peut présenter une part de légitimité au regard de contextes d'application différents. Cependant, et peut-être plus radicalement, ces postures voudraient renvoyer, certes sur un mode subjectif, à l'émergence d'un sujet dynamique *original* qui, tout à la fois, pense, ressent et agit dans des contextes de plus en plus complexes(cf. Morin (1994), Le Moigne (1995), Pellois, (2002, 2003, 2004))[18], par exemple, paradoxaux[19], disjoints et complémentaires à la fois, mais aussi de plus en plus largement ouverts à l'ensemble large, donc pas seulement cognitives, des situations humaines vécues. Ce qui vaut pour l'individu ne vaudrait-il pas, alors, également, pour ceux qui pensent les conceptions théoriques *dans l'interprétation* de faits *« objectivement »* observés ?

Mais, ce débat, qui méritait d'être posé, au regard de la réflexion à caractère épistémique concernant la méthodologie d'utilisation de l'analyse implicative et du logiciel CHIC, dépasse très largement le propos qui peut être développé ici. Aussi il convient de revenir au sujet *« psychologique »* élargi mais traité dans un cadre qui reste, limité à l'approche objective.

2.5 Retour à une complexité moindre mais permettant le traitement statistique de données : une vision élargie mais qui resterait « piagétienne » ?

A cette étape de la réflexion, et au regard de ce qui précède, gardons simplement en mémoire que, si le raisonnement opératoire, à caractère implicatif peut être appliqué au *sujet psychologique*, même en élargissant le champ des variables en jeu, ce ne peut être que de manière limitée, voire dans une conception plus ou moins ou moins réduite, et, en ce sens, plus ou moins *artificielle*. Cela ne peut, donc, se faire qu'en faisant abstraction d'un contexte encore plus large, d'une approche plus complexe que le simple jeu d'élaboration puis d'expression, parmi bien d'autres choses, des opérations de la logique binaire, s'appliquant au raisonnement hypothético-déductif. Ce mode de raisonnement relève alors, effectivement, encore ici, d'une forme schématique, au contexte défini de manière emboîtée. Il emprunte cependant davantage à la psychologie du sujet qu'à la structuration, puis la structure acquise, abstraite, d'un *sujet épistémique*. Même si ce *sujet psychologique* est encore conçu, ici, selon une organisation générale, qui peut être perçue comme simplifiée, des états acquis (les stades), il renvoie, déjà, à

[18] Mais aussi le courant, dit de l'intersubjectivité.
[19] Voir, peut-être, sur ce point, l'approche des thérapies familiales.

des phases plus complexes, ici multifactorielles, de son élaboration, ou de l'expression complexe de sa fonctionnalité.

Pour autant, le modèle piagétien initial du développement du *sujet épistémique*, bien que très strictement, donc étroitement formalisé, est, lui-même, complexe, car, c'est l'évidence, celui-ci ne se limite pas au simple développement du raisonnement implicatif[20] pas plus qu'il ne se limite à l'expression du raisonnement hypothético-déductif dans le domaine des sciences[21]. Mais la complexité est déjà singulièrement plus grande lorsque ce modèle est interrogé, du point de vue d'autres conceptions traitant de contenus et de modalités de conduites adaptatives. Ces modalités sont décrites dans des conceptions portant sur les modes d'expression du *sujet psychologique*, considérés comme basiques mais à champ d'application large. Il s'agit davantage de *« fonctions »* (cf., par exemple, à ce sujet, Lautrey et Chartier (1987), 1995, Richard (1990), Baddeley (1993)) exercées par lui, comme l'attention, les différentes formes de mémoire, de fluidité et/ou flexibilité mentale, les différentes modalités de contenus et de traitement de l'information[22], les différentes sensibilités aux états, contenus, formes, expressions émotionnelles (Damasio (1995)), etc. Ces différentes dimensions de l'expression du *sujet psychologique* sont nécessairement mobilisées, du fait de leur polyvalence de fonctionnalité, à la fois dans l'élaboration du développement et dans son expression fonctionnelle à partir d'*« organisations cognitives »* acquises, conçues dans un sens culturel élargi. L'expression *« organisation cognitive »*, ici mise entre guillemets, recouvre-t-elle pour autant toute l'expression culturelle du sujet, soit l'ensemble de ses expressions plus ou moins organisées ? Sans doute pas, mais cela recouvre, pour le *sujet psychologique*, un champ varié et vaste d'expressions élaborées plus ou moins complexes qui en sont, en quelque sorte, les *« produits »* plus ou moins adaptatifs —tel le raisonnement hypothético-déductif, mais pas uniquement— aux différents contextes humains auxquels ces *« organisations cognitives »* correspondent, soit une grande part du raisonnement scientifique actuel, mais pas seulement. Il convient, enfin, d'insister sur le fait que ces *« organisations cognitives »* sont mobilisées, à la fois, au moins dans l'expression fonctionnelle, du sujet, et dans son développement y compris, donc, opératoire au sens piagétien du terme.

Mais les contextes, auxquels ces *« produits*[23] *»* de l'élaboration *« cognitive »* acquise et/ou de son développement, peuvent s'appliquer, présentent des particularités qui conditionnent justement le caractère plus ou moins adaptatif de ces *« produits »*, pris séparément ou associés dans des *produits* d'organisation d'ordre supérieur. Ils peuvent conditionner la forme de leurs mobilisations plus ou moins conjointes, dans ces contextes, correspondant aux conduites supérieures[24] du sujet dans lesquels ils sont associés : plus ou moins en covariance, plus ou moins emboîtés dans un sens ou dans un autre[25], ou indépendants. Voire, certains de ces *« produits »* de l'*« organisation cognitive »* peuvent être incompatibles entre eux. Il s'agit, donc, de cerner ici les problématiques liées à des *produits* variés de l'expression du sujet, mais aussi de

[20] Notamment, les épreuves intitulées Test des Opérations Formelles (T. O. F.) de F. Longeot.
[21] Travaux de Piaget concernant le jugement moral chez l'enfant, la perception, etc. (Piaget, 1932, 1961).
[22] Spatial/verbal-numérique, analogique/propositionnel, successif/simultané, etc.
[23] Culturels ? Pas seulement ?
[24] Vie sociale, expression d'une pensée complexe, stratégies, raisonnements complexes, créativité, relations, sentiments, etc.
[25] Suivant les personnes, les contextes.

« *produits de produits* », relevant d'un niveau d'organisation, considéré comme supérieur, liés à l'expression de *« fonctions »* psychiques, et appliqués à des contenus et selon des modalités multiples. Ces « *produits* » seraient, ici, pris dans un sens et un contexte très large, par exemple, du raisonnement rationnel, au sens de la logique binaire, qui inclut le raisonnement implicatif, jusqu'aux capacités subtiles à sentir, saisir le sens, du poète…

2.6 Le sujet en situation : ses parts de liberté et ses parts de contraintes variables et variées

Il s'agit bien de reconnaître des contextes, et des cadres d'expression du sujet y afférant (renvoyant ici, de nouveau, à des champs conceptuels au sens de A. Vergnaud (Gras et Acioly-Régnier, 2015), dans lesquels, notamment, jouent de manière variée les conditions de contraintes plus ou moins prégnantes, nécessitant la compréhension des règles permettant de s'y exprimer, mais aussi les parts de liberté[26] permettant la modification des *« règles »*, ou même l'élaboration de nouvelles *« règles »*[27] dont disposerait alors le sujet pour, à la fois, s'exprimer et se développer.

Les jeux et les niveaux de ***contraintes*** des situations, les ***parts de liberté*** qui peuvent apparaître, dans le domaine scolaire, mais aussi la vie courante (Pellois (2008, 2010, 2013), sont bien variables en fonction des contextes. Ils peuvent relever de cadres complètement contraignants. Il en est ainsi, par exemple, du cadre du raisonnement implicatif formel pris au sens strict. Ils peuvent relever de situations de conception, d'expression plus ouvertes, mobilisant d'autres habiletés que le raisonnement même élargit à l'ensemble des opérations de la logique propositionnelle, voire pas de raisonnement du tout. Ce peut être le cas de situations de conception vécues dans l'architecture, l'inventivité, mais aussi d'expression de l'éthique, de la créativité, de l'esthétique, dans l'art, ou simplement de l'imaginaire. Ces situations d'émergence de la pensée ne réfèrent pas, ou pas *nécessairement*, à celles des sciences, mais *la mobilisation de ce qui s'y exprime peut aussi être nécessaire aux sciences*. Sur ce plan, vraisemblablement, à l'intérieur même des sciences, les contraintes et donc les parts de liberté à prendre en compte en sciences humaines ne sont pas les mêmes qu'en biologie, celles-ci n'étant pas nécessairement les mêmes que celles de la physique, qui à leur tour ne sont pas nécessairement les mêmes que celles correspondant à la logique, aux mathématiques peut-être[28], etc.

---

[26] Comme en biologie ou les parts de liberté que peut faire jouer un organisme, dans un contexte donné (Par exemple : les fluctuations des réplications chromosomiques aux mutations génétiques) permet l'expression de la variabilité nécessaire pour favoriser ensuite l'adaptation plus ou moins optimale, par sélection naturelle des formes (sauf, peut-être (?), dans des conditions particulières, ou une seule solution (toujours la même) est possible au regard de contraintes particulièrement strictes et rigoureuses observées (par exemple, les cas d'invariants ou de constantes en physique, en mathématique ?) et de contextes relativement simples et étroitement définis, il peut, dans des contextes plus complexes, laissant davantage d'espaces de libertés, y en avoir communément plusieurs, vicariance oblige…), des caractères les plus adaptés aux conditions de contrainte du contexte et à ses changements.

[27] Exemple, au plan épistémologique, la pertinence, dans le domaine de la physique, du principe de non contradiction confronté à l'apparente contradiction, de l'incertitude, de l'*« intrication »*, etc. considérés comme opérant à un certain niveau de développement conceptuel de la théorie quantique, et l'interêt des phénomènes paradoxaux de la communication dans le cadre des thérapies familiales.

[28] Et qu'en est-il à l'intérieur des mathématiques elles-mêmes ?

3 Vers le traitement des données, tenant compte des complexités différentielles des sujets psychologiques : modéliser pour une utilisation adaptée de l'analyse implicative

3.1 Quelques mots pour rappeler le cadre ici retenu en vue de l'utilisation du logiciel

Tout cela peut amener, alors, au moins dans un premier temps, à une utilisation de CHIC, non pas, il convient de le répéter ici, comme *forme de modélisation* d'un quelconque *sujet épistémique* plus ou moins étroitement ou largement défini, mais comme *outil* utilisé en vue d'aborder les différentes formes de développement et d'expression d'un *sujet psychologique* aux conduites multiples et variées dans des contextes les plus divers. Cela peut être envisagé, certes, au départ, de manière restreinte, et seulement dans le cadre contextuel du développement et de l'expression opératoire, au sens piagétien du terme, et donc pas seulement sur le plan du raisonnement strictement implicatif. Mais cela peut être considéré plus largement, ensuite, dans le cadre du développement et de l'expression de compétences scientifiques prises au sens large et appliquées dans les contextes les plus variés, ou dans le cadre d'une expression culturelle plus ouverte encore. L'application du logiciel est, cependant, toujours restreint aux limites du contexte de son application, relevant de l'objectivité et du rationnel, mais pouvant tout de même opérer dans des *« univers »* élargis dans lesquels semble bien encore *« fonctionner »*, au moins *« localement »*, la logique implicative. Ceci n'est pas nécessairement vrai pour l'ensemble de toutes les conduites, y compris adaptatives, du *sujet psychologique*, c'est-à-dire de l'ensemble des *sujets psychologiques* en prise avec tous leurs modes d'expression possibles, actuellement connus ou non connus.

3.2 Homogénéité, hétérogénéité et vicariance : un exemple formel de modélisation avant traitement des données

Ramenée au cadre du raisonnement implicatif, mais pas seulement, une illustration formelle de cette utilisation, certes limitée, peut être concrétisée par le jeu très simplifié qui suit, jeu fait de questionnements et de propositions, d'hypothèses et d'inférences.

Soit des capacités fonctionnelles {a, b, c, d, e} : comme des capacités d'attention, de mémoire, de contrôle, d'inhibition/facilitation de représentations, de schèmes, de flexibilité d'évocation mentale, de planification, etc. dont disposerait un sujet. Soit, par ailleurs, des états de structurations opératoires (au sens piagétien) *« partielles »* {*S, T, U, V, W*} : comme des éléments de logique de classe, de relation, des capacités de permutation, de combinatoire, voire des sous-groupes des 16 opérations de la logique binaire, que maîtriserait un sujet. Ces états *« partiels »* seraient susceptibles de lui permettre d'aboutir ensuite à l'acquisition et la maîtrise cognitive d'une organisation conceptuelle globale, ici la pensée opératoire formelle piagétienne {**X**}, donc de pouvoir mettre en œuvre des aspects fondamentaux du raisonnement hypothético-déductif scientifique de l'adulte. L'observation, la mesure, montre que l'acquisition de {**X**}, au sens des analyses statistiques traditionnellement utilisées, est corrélée aux acquisitions de {*S, T, U, V, W*}, pris séparément, en sous-ensembles, ou plus complètement. Ces acquisitions sont elles-mêmes corrélées à l'expression des performances du sujet en {a, b, c, d, e}. La question se pose du sens de cette corrélation conjointe entre toutes ces

variables considérées, conduisant à un jeu d'hypothèses habituelles[29] lorsque l'on traite de co-variation de performances, hypothèses entre lesquelles il conviendrait de choisir en fonction de résultats d'observation obtenus et de l'analyse logique qui peut en être faite.

Une question classique qui peut être posée, et à laquelle un traitement par CHIC peut donner des résultats permettant d'induire une possible réponse, peut être la suivante : derrière les corrélations observées, quels sont les rapports, au plan implicatif, entre ces différentes classes ou catégories de variables {a, b…S, T… **X**} prises dans leur ensemble ? Mais la question peut être aussi la suivante. Renvoient-ils, ces résultats, dans un contexte donné, à une implication/causalité (?) complexe entre *des variables* fonctionnelles et *d'autres*, en tant que *« produits »* d'organisation (ici cognitive), tenant à un exercice de *ces variables* fonctionnelles (***covariation primaire***) permettant l'expression évolutive conjointe des capacités qui leur correspondent ? Renvoient-ils à l'interaction (par exemple, par causalité circulaire ?) permettant la mutualisation des effets évolutifs des dimensions considérées à leur bénéfice et à celui des *autres variables* ? Ou bien existe-t-il une variable latente (type variable *« produit »* ?), commune à ces dimensions, et susceptible d'intervenir conjointement sur chacune d'elles ? Ou bien, encore, de manière plus restrictive, et dans un schéma piagétien référencé au schéma ***implicatif*** strict ou même néo-piagétien, une autre formulation du jeu de questions, dans une élaboration élargie donc plus complexe que celle du schéma opératoire piagétien initial, et relevant de l'exemple formel présenté ci-dessus, peut être la suivante. L'expression de certaines capacités ({a, b, c, d, e}[30] nécessaires) rend-elle possible le développement d'autres capacités culturellement organisées et d'un niveau d'organisation supérieur ({*S, T, U, V, W*} (soit des sous structures du développement opératoires) ? Ces capacités acquises permettraient-elles, elles-mêmes, par leur organisation, le développement complet, puis l'expression efficace et étendue, d'une structure d'ensemble (logico-mathématique formelle) supérieure dans son organisation et ses qualités cognitives {**X**} (***covariation secondaire***) ? Il semblerait que l'analyse implicative utilisée dans une conception méthodologique rigoureuse, mais ouverte, permet assez bien de répondre au type de problématique évoqué là.

Mais cela peut renvoyer, de nouveau, à une série de questions que l'analyse implicative, mais éventuellement associée à d'autres modes d'analyse de relations entre variables, peut traiter. Par exemple, une fois la *superstructure* atteinte, celle-ci par son utilisation, permet-elle en retour, du fait de leur exercice (***covariation tertiaire fonctionnelle***), des capacités conjointes améliorées des dimensions, initiales pour celui qui en maîtrise, en quelque sorte, le *« produit »* structuré ? Ou bien cette *superstructure* ne peut-elle s'exprimer comme conduite culturelle que parce qu'une autre organisation plus psychologique voire biopsychologique sous-jacente le permet ? Ou, à l'opposé, cette *superstructure* atteinte[31] ne transcende-t- elle pas, en quelque sorte, les autres capacités en décuplant l'expression de leurs qualités et donc de leur efficacité (***covariation tertiaire structurale***) ? Ce jeu de questions n'est, bien entendu, pas

[29] Notamment simple covariation dépendant d'un facteur latent, relation causale directe supposée et sens de la causalité déduit de conceptions théoriques inhérentes au modèle considéré.

[30] Habiletés cognitives, capacités diverses de mémoire, d'attention, etc.

[31] Pour Piaget la structure de groupe INRC au sommet de la combinatoire des 16 opérations propositionnelles entre elles.

limitatif et peut intégrer d'autres dimensions que celles strictement cognitives[32]. Mais cette conception qui reste cependant limitée de l'expression du *sujet psychologique*, plus largement envisagée, ne peut elle pas aussi contrarier d'autres modes d'expression de ce même *sujet psychologique*, et ces mêmes capacités fonctionnelles initiales, voire d'autres *« capacités »* ? Ce qui renvoie alors à l'opération de ***non-compatibilité*** au moins partielle de la structure avec ces autres dimensions. Plus largement, des capacités seraient pertinentes dans certaines situations adaptatives, mais pas nécessairement dans d'autres, voire feraient obstacle à l'expression d'autres capacités alors utiles en d'autres circonstances.

3.3 Du modèle formel à un exemple : esquisse d'une méthode d'approche du traitement de donnée utilisant, notamment, l'analyse implicative

De manière plus restreinte, concernant le thème de la réflexion développée ici, une perspective davantage cernée et cadrée, plus *« locale »*, traitant de sous-ensembles d'une population donnée, est susceptible de mieux s'inscrire dans une approche méthodologique différentielle. Une approche qui correspond davantage à la préoccupation principale du propos développé ici en rapport avec l'utilisation de CHIC. Cette approche peut alors amener à d'autres observations qui seront traitées, mais de manière plus ou moins complète, tenant à la limite des choix des sous-groupes retenus, par une analyse implicative, voir à ce sujet Chamsine (2012), cette fois-ci différentielle.

Cette perspective peut être, de nouveau, celle posée par le fonctionnement vicariant qui peut être décrit formellement, sur les mêmes séries de variables que précédemment, de la façon suivante. Il peut être observé, d'une part, que la maîtrise des sous-structures {*S* <u>*et*</u> *T*} sont davantage corrélées à {a <u>*et*</u> b} et les sous-structures {*V* <u>*et*</u> *W*} davantage corrélées à {d <u>*et*</u> e}. D'autre part, lorsque l'on analyse finement sur un plan implicatif la structuration progressive au stade formel, il est constaté que, pour la population considérée et à partir de contenus présentés sous une forme donnée, si une implication ou plutôt une *quasi-implication*, existe entre {a, b, c, d, e} et {**X**} d'un côté et {*S, T, U, V, W*} *et* {**X**}**,** d'un autre côté, lorsque l'on relève des indices d'implication concernant des sous-ensembles de structures partielles {*S, T*} ou {*V, W*}, chacune de ces structures partielles est plutôt associée à un groupe d'expression de performances fonctionnelles différent : {a, b} ou {d, e}. Ensuite une analyse implicative globale associant ces différents groupes de variables à **X** montre des relations de *quasi-implication* modérée de chacune de ces catégories de variables, prises en sous-groupes séparés, avec **X.**

Mais que signifient ces différents constats de *quasi-implication* ? Traduisent-ils, au regard de l'expression d'une certaine homogénéité des données, une implication stricte entre les variables retenues, exception faite des erreurs fortuites dans la prise de mesure conduisant à l'expression de cas aberrants au regard de la conception de l'implication stricte ? Ou bien ces cas supposés, alors à tort comme aberrants, ne sont-ils pas révélateurs d'une vicariance, certes, pouvant porter sur une dissymétrie de voies de cheminement dans le développement, mais vicariance cependant, attestée par

[32] Une autre hypothèse possible, en effet, peut être la suivante : suivant les moments du développement, le fonctionnement lié à l'action des deux dimensions considérées, univoque pour chacune des deux questions qui précèdent, peut évoluer vers une fonctionalité alternative, intégrant d'autres dimensions que des dimensions strictement cognitives, voire un fonctionnement mutuel conjoint (circularité), traduisant des mécanismes de **covariation quaternaire complexes**, de l'ordre de dimensions culturelles élargies, et ceci, bien au-delà des jeux d'inter-causalité circulaire de la **covariation primaire**.

l'expression d'une implication inverse, entre les dimensions prises en compte, pour une petite partie de la population ? Ou bien la dispersion des résultats ne s'expliquerait-elle pas à la fois par les deux raisons invoquées ?

Une manière de trancher peut-être de saisir des caractéristiques de sous-populations du groupe de référence. Ces caractéristiques sont susceptibles, par exemple, de valider l'hypothèse d'une hétérogénéité du développement à l'œuvre. Celle-ci se traduirait par une voie qui peut être plus dominante du développement, reflétant une certaine homogénéité partielle, dans le contexte considéré[33], et une voie qui peut être plus minoritaire, et que le constat des relations implicatives directes et inverses, pour des sous-groupes différents, permet de mettre à jour. Pour confirmer cela, il conviendrait que des analyses implicatives différentielles réalisées à partir de sous-populations adéquatement sélectionnées dans la population parente valident un chemin reliant les variables {a, b} *et* {*S, T*} puis {*S, T*} *et* {*V, W*}, enfin {*V, W*} *et* {**X**} d'une part. Il conviendrait, d'autre part, qu'elles valident également un chemin reliant les variables {d, e} *et* {*V, W*}, puis {*V, W*}, *et* {*S, T*} enfin {*S, T*} *et* {**X**}. Seul l'ensemble des résultats obtenus peut permettre de trancher, tout au moins en termes de **développement**, entre l'hypothèse d'homogénéité générale et celle d'hétérogénéité différentielle, des évolutions. Mais, une fois l'organisation {**X**} acquise, un autre aspect des différences peut apparaître alors, comme cela est évoqué par ailleurs. L'élaboration, le développement de la structure **X**, via le cheminement structurant et le style dominant s'appuyant sur des habiletés plus prononcées en {a, b} passe par l'expression structurée partielle de type {*S, T*} puis de {*S, T*} *à* {*V, W*}, et enfin *de* {*V, W*} *à* l'expression globale de {**X**}.

Cette voie particulière peut permettre également une **expression** aisée et préférentielle dans des contenus et sur des modalités propres à {a *et* b} et beaucoup moins sur des modalités propres à {d *et* e}. Par contre, l'élaboration, le développement de la structure **X**, via le cheminement structurant et le style dominant s'appuyant sur des habiletés plus prononcées en {d, e} passerait, elle, par l'expression structurée partielle de type {*V, W*} puis de {*V, W*} *à* {*S, T*}, et enfin *de* {*S, T*} *à* l'expression globale de {**X**}. Cette voie permettrait l'expression préférentielle différente de la précédente, c'est-à-dire une voie d'expression plus aisée, et dynamisant le développement, en {d *et* e}, et moins aisée en {a *et* b}**.** Mais, en élargissant le champ, une autre perspective peut compléter la précédente, l'exercice des performances et des compétences qui permettent le développement opératoire selon un certain cheminement menant à **X**[34], que ce soit le premier ci-dessus ou le second, n'est-il pas **incompatible**[35]**,** au moins partiellement[36] et pour une part des sujets considérés, avec ce qui permet le développement ailleurs, par exemple de **Y**[37] ? La réponse affirmative à cette question, après analyse statistique adéquate, relèverait de la mise en évidence différentielle, non plus de la corrélation, ni, non plus, de l'implication ou de la vicariance, mais bien de l'incompatibilité, déjà évoquée précédemment, au moins partielle, qui existerait entre l'exercice efficace et

[33] C'est-à-dire des contraintes situationnelles, de forme, d'expression, etc.

[34] La pensée opératoire dans le raisonnement.

[35] ***Non-compatibilité*** évoquée précédemment.

[36] En effet l'incompatibilité peut être plus ou moins radicale, ou relative, auquel cas une voie ***minoritaire*** du développement vers l'une des organisations peut être la voie ***majoritaire*** dans l'autre organisation et vice versa.

[37] Par exemple : les qualités d'empathie dans la relation.

adapté de ces deux compétences larges. Là aussi, la recherche et l'analyse statistique des données, mais plus uniquement sur le mode implicatif, concernant l'homogénéité/l'hétérogénéité[38], ici, tendancielle, qu'elles soient dominantes ou minoritaires, renvoie à une approche féconde à la fois : de la diversité, donc de la variabilité, de la pluralité des différences et des problèmes que cela pose aux différentes formes de développement du *sujet psychologique*.

4 De l'analyse des données au retour à la modélisation : Le modèle piagétien inscrit dans une conception élargie ? Une nouvelle conception ?

In fine, cette dernière piste de recherche peut aboutir à une perspective encore plus radicalement différentielle[39] sur les voies variées du développement pris au sens large, voies menant à des configurations différentes (par exemple : raisonnement et empathie ?), ces configurations permettant des conduites adaptatives différentes à des situations d'application de ces conduites elles-mêmes différentes. La question qui resterait alors en suspens est la suivante. Existe-t-il des sujets capables de maîtriser conjointement les deux dimensions considérées malgré un jeu d'incompatibilité momentané ou seulement local ? Plus clairement, y aurait-il, pour la plupart des cas considérés, des jeux d'habiletés différentes, pour des sujets différents, mais également, pour quelque cas, des capacités de maîtrise conjointe ? Cela dépendrait-il alors d'une structure d'ensemble {***Z***}, incluant {**X**} <u>*et*</u> {**Y**}, que quelques (?) personnes seraient susceptibles de maîtriser, ou de l'expression, simplement, de deux habiletés, plus ou moins compatibles, se développant conjointement ? Auquel cas, les conceptions piagétiennes relèveraient d'une modélisation, d'ordre épistémique, *locale,* principalement active lors de l'expression des comportements hypothético-déductifs et de leur genèse. Se pose alors la question de ce qu'est véritablement la réalité *psychologique* effective d'un sujet se plaçant en deçà (maîtrise progressive du développement opératoire) mais aussi au-delà (maîtrise conjointe de situations très variées, tel que ci-dessus évoqué) d'un tel modèle. Cela s'inscrit bien dans un contexte méthodologique qui renvoie à l'expression d'un raisonnement ayant à voir avec la pensée hypothético-déductive du *sujet épistémique* telle qu'elle est décrite formellement par ailleurs, mais une pensée qui s'exprime dans un ensemble plus vaste des comportements humains, l'expression *psychologique du sujet* dans son ensemble…

De même, si le modèle « standard » général et épistémique piagétien par sa logique emboîtée nécessairement implicative, tel que le représente, par exemple, Laroche (1966, p. 67), se prête bien à un traitement univoque selon la logique implicative de CHIC, le développement formel qui précède confirme bien, comme cela a déjà été évoqué, la pertinence d'utiliser CHIC, pas nécessairement, répétons-le, comme *« modèle »* univoque, mais plutôt comme *outil* de traitement. L'application de CHIC se fait, simplement, dans un ensemble différentiel plus complexe du fait psychologique

[38] Cette approche méthodologique de l'utilisation de CHIC donne lieu par ailleurs à une simulation formelle avec calculs par le logiciel CHIC sur des données fictives (cf. Pellois (2015) mais représentatives de distributions effectivement possibles dans la réalité (cf. Chamsine (2012)) des observations réalisées dans la recherche en psychologie.

[39] De Thurstone (par exemple : Huteau et Lautey, 1999, p.142 et suivantes), à Gardner (1996), etc.

constitué de multiples configurations possibles, différentes, et non nécessairement hiérarchisables, voire plus ou moins partiellement compatibles. Cela pose, alors, les bases d'une certaine manière d'utiliser CHIC dans une conception méthodologique plus large que le traitement sous le simple aspect de l'analyse implicative, en particulier au sens strict, des conduites du sujet. L'utilisation de CHIC renverrait alors non plus à des perspectives d'une modélisation qui ramènerait aux conceptions piagétiennes, fussent-elles élargies, développées à un moment donné de la réflexion dans le domaine de la psychologie, mais bien davantage à des perspectives incertaines et hypothétiques davantage de l'ordre de l'ouverture, vers de nouvelles perspectives modélisatrices. En exprimant ce point de vue pour conclure, il apparaît bien que perdure, par delà les débats et les remises en causes, une conception bien piagétienne, certes de l'ordre d'une réflexion davantage épistémique, celle articulant dans le développement du sujet psychologique, assimilation et accommodation.

Références

[1] Bachelard, G. (1987), *Le nouvel esprit scientifique*, PUF, Paris.

[2] Baddeley, A. (1993), *La mémoire humaine théorie et pratique*, PUG, Grenoble.

[3] Bandura, A. (2003), *Auto-efficacité. Le sentiment d'efficacité personnelle*, De Boeck Université, Paris.

[4] Bideaud, J., O. Houdé, et J.-L. Pedinielli (1993), *L'homme en développement*, PUF, Paris.

[5] Chamsine, D. (2012), *Situation d'échec scolaire et sentiment d'échec scolaire de l'élève. Etude des interactions de divers facteurs qui entre en jeu*, Thèse de doctorat, Université de Caen.

[6] Damasio, A. R. (1995), *L'erreur de Descartes, la raison des émotions*, Odile Jacob, Paris.

[7] Delacour, J. (1998), *Une introduction aux neurosciences cognitives*, de Boeck Université, Paris.

[8] Gardner, H. (1996), *Les intelligences multiples. Pour changer l'école : la prise en compte des différentes formes d'intelligence*, Retz, Paris.

[9] Gras, R. et Acioly-Régnier N. (2015), Hiérarchie de règles en A.S.I. et conceptualisation, *Actes de A.S.I. 8*, Radès (Tunisie).

[10] Houdé, O. (1992), *Catégorisation et développement cognitif*, PUF, Paris.

[11] Houdé, O., B. Mazoyer, et N. Tzourio-Mazoyer (2002), *Cerveau et psychologie. Introduction à l'imagerie cérébrale anatomique et fonctionnelle*, PUF, Paris.

[12] Huteau, M. et J. Lautrey (1999), *Evaluer l'intelligence. Psychométrie cognitive*, PUF, Paris.

[13] Juhel, J. (2005), Les systèmes dynamiques non linéaires dans l'étude de la variabilité cognitive, *Cahiers de la MRSH*, n° spécial, Caen, 193-216.

[14] Laroche, J.-L. (1966), Diagnostic des structures opératoires d'adultes de niveau CEP, *C.P.T.R.*, Metz bibliothèque, 664, 1-83.

[15] Lautrey, J. et D. Chartier (1987), Images mentales de transformations et opérations cognitives: une revue critique des études développementales, *L'Année Psychologique*, 87, 581-602.

[16] Lautrey, J. (sous la direction de) (1995), *Universel et différentiel en psychologie*, PUF, Paris.

[17] Leenhardt, J. (2005), Modalités de prise en charge du cognitif dans l'art. Chemin que cela esquisse pour une recherche des sciences de la cognition sur l'expérience esthétique. Dans : *Approche cognitive de la création artistique*, sous la direction de Mario Barillo, Pierre Mardaga éditeur, Sprimont (Belgique), 57-64.

[18] Longeot, F. (1978), *Les stades opératoires de Piaget et les facteurs de l'intelligence*, PUG, Paris.

[19] Le Moigne, J.-L. (1995), *Le constructivisme*, tome 2 : *des épistémologies*, ESF, Paris.

[20] Morin, E. (1994), *Introduction à la pensée complexe*, ESF, Paris.

[21] Olry-Louis, I., V. Guillon, et E. Loarer (2013), *Psychologie du conseil en orientation*, de Boeck Supérieur, Bruxelles.

[22] Perrot, A., Fontayne, P., Bertsch, J., Bot, G. (2006), Évolution de la configuration des aptitudes cognitives et psychomotrices au cours de l'acquisition d'une habileté motrice. L'exemple de la « cascade à trois balles » en jonglage, *Movement & Sport Sciences*, **2** (58), 81-92.

[23] Pellois, C. (2002), *Apprentissage et développement de la personne dans l'enseignement et la formation*, tome 1 : *Du rationnel au complexe*, l'Harmattan, Paris.

[24] Pellois, C. (2003), *Enseignement et formation : le développement de la personne*, tome 2 : *Le complexe*, l'Harmattan, Paris.

[25] Pellois, C. (2004), *Enseignement et formation : le développement de la personne*, tome 3 : *Le sensible et le rationnel*, l'Harmattan, Paris.

[26] Pellois, C. (2008), Contrainte et liberté du sujet : entre incertitude et prévisibilité ? Dans : *Cognition, incertitude et prévisibilité*, B. Cadet, G. Chasseigne, G. Foliot, Publiboock, Sciences Humaines et Sociales, Paris, 77-100.

[27] Pellois, C. (2010), Sens et incertitude, une forme de complexité en psychologie : des contraintes aux parts de liberté, le développement et ses contextes. Dans : *Traitement de la complexité dans les sciences humaines*, Cadet, C., et G. Chasseigne, Editions Publibook Université, 177-207.

[28] Pellois, C. (2013), La psychologie et l'usage du traitement mathématique des données statistiques. Nouvelles perspectives conceptuelles, *Revista Brasiliera de Ensino de Cienca e Technologia*, **6** (1), 230-259.

[29] Pellois, C. (2014), Comment mieux traiter de la question de l'autre, et, donc, des différences en entretien d'aide et de conseil à la personne ? Un point de vue épistémologique ? Dans ouvrage collectif : *Comment mieux prendre en compte les*

différences en entretien d'aide et de conseil ? Tome 1, l'Harmattan, Paris, 191-250.

[30] Pellois, C. (2015), Le traitement statistique de données et la complexité de l'expression et du développement différentiels du sujet psychologique, *Actes de A.S.I. 8*, Radès (Tunisie).

[31] Piaget, J. (1932), *Le jugement moral chez l'enfant.* Delachaux et Niestlé, Neuchâtel.

[32] Piaget, J. (1961), *Les mécanismes perceptifs : modèles probabilistes, analyse génétique, relations avec l'intelligence*, PUF, Paris.

[33] Richard, J.-F. (1990), *Les activités mentales, comprendre, raisonner, trouver des solutions*, Armand Colin, Paris.

[34] Sternberg, R. J. (1994), La théorie triarchique de l'intelligence, *l'Orientation Scolaire et Professionnelle*, **23** (1), 119-136.

[35] Varela, F., Thomson, E., Roch, E. (1993), *L'inscription corporelle de l'esprit : sciences cognitives et expérience*, Éditions du Seuil, Paris.

BEM-ESTAR SUBJETIVO DE ESTUDANTES BRASILEIROS DA ÁREA DE CIÊNCIAS EXATAS: ABORDAGEM DOS DADOS NO QUADRO DA ANÁLISE ESTATÍSTICA IMPLICATIVA

Fabiana ALEXANDRE FERREIRA[1]; Céline FAURE[2]; Sofiane BOUZID[3]

BIEN-ETRE SUBJECTIF DES ETUDIANTS BRESILIENS DANS LE DOMAINE DES SCIENCES EXACTES: L'APPROCHE DES DONNEES DANS L'ANALYSE STATISTIQUE IMPLICATIVE

SUBJECTIVE WELL-BEING OF BRAZILIAN STUDENTS IN THE FIELD OF EXACT SCIENCES: THE APPROACH OF THE DATA IN THE IMPLICATIVE STATISTICAL ANALYSIS

RESUMO

O presente artigo apresenta dados parciais de um projeto de pós-doutoramento, em andamento, que visa validar um índice de bem-estar para o Brasil e para a França, comparando os resultados entre os dois países após a validação. O termo bem-estar ou bem-estar subjetivo costuma ser utilizado na literatura como sinônimo de felicidade, pressupondo que o próprio indivíduo é quem pode melhor avaliar sua satisfação para com a vida. O termo está ligado à psicologia positiva e é um conceito complexo que integra as dimensões cognitiva e afetiva, sendo incorporado num abrangente campo de estudo que abarca outros grandes conceitos e domínios de estudo, tais como o afeto positivo, o afeto negativo e a qualidade de vida. A escala que será validada é o Índice de Felicidade Pemberton. Para este artigo foram selecionadas apenas as respostas dos estudantes brasileiros da área de ciências exatas aos itens referentes ao bem-estar lembrado do índice. Os dados foram analisados com o apoio do software CHIC (Classification Hierárquique, Implicative et Cohérsive), enfocando as similaridades e as implicações entre as itens.

***Palavras-chaves**: Bem-estar subjetivo. Felicidade. Psicologia Positiva. Análise Implicativa. Índice de Felicidade Pemberton.*

RÉSUMÉ

Cet article présente des données partielles d'un projet post-doctoral, en cours, visant à valider un indice du bien-être pour le Brésil et la France, en comparant les résultats entre les deux pays. Le bien-être ou le bien-être subjectif est souvent utilisé dans la littérature comme synonyme du bonheur, en supposant que l'individu lui-même est celui qui peut le mieux évaluer son bien-être. Ce terme du bien-être subjectif est liée à la psychologie positive, il s'agit d'un concept complexe qui intègre les dimensions cognitives et affectives qui inclut d'autres concepts et domaines d'études tels que l'affect positif, l'affect négatif et la qualité de vie. L'échelle à valider est l'indice du Bonheur de Pemberton. Pour cet article, nous avons sélectionné un échantillon des étudiants brésiliens issues de la filière des sciences exactes et leurs réponses concernant la partie « souvenir de bien-être » de l'indice. Les données ont été analysées avec le soutien du logiciel CHIC (Classification

[1] Instituto Federal de Educação, Ciência e Tecnologia de São Paulo. Rua Américo Ambrósio, 269, Jardim Canaã, CEP 14169-206, Sertãozinho, São Paulo, Brasil. fabiana.ifsp@gmail.com

[2] Institut des Sciences de L'Homme. 14 avenue Berthelot, 69363, Lyon, CEDEX 07, France. celine.faure@ish-lyon.cnrs.fr

[3] Institut des Sciences de L'Homme. 14 avenue Berthelot, 69363, Lyon, CEDEX 07, France. Sofiane.bouzid@ish-lyon.cnrs.fr

Hiérarchique, Implicative et Cohésive), mettant l'accent sur les similitudes et les implications entre les éléments.

Mots-clés : *Bien-être Subjective. Bonheur. Psychologie positive. Analyse implicative. Indice de Bonheur Pemberton.*

ABSTRACT

This research paper presents partial data from a post-doctoral project, in progress, aimed at validating a well-being index for Brazil and France, comparing the results between the two countries after validation. The well-being term or subjective well-being is often used in the literature as a synonym for happiness, assuming that the individual himself is the one who can best assess their satisfaction with life. The term is linked to positive psychology and is a complex concept that integrates the cognitive and affective dimensions, being incorporated into a comprehensive field of study that includes other major concepts and fields of study such as positive affect, negative affect and the quality of life. The scale to be validated is the Pemberton Happiness Index. For this article we selected only the responses of Brazilian students in the area of exact sciences to the items for the well-being remembered of the index. Data were analyzed with the support of the CHIC software (Classification Hierárquique, implicative et Cohérsive), focusing on the similarities and implications between the items.

Keywords : *Subjective Well-Being. Happiness. Positive Psychology. Implicative analysis. Pemberton Happiness Index.*

1 Introdução

O presente artigo apresenta dados parciais de uma pesquisa intitulada “Validação do Índice de Felicidade Pemberton para o Brasil”. Os dados empíricos deste estudo fazem parte de um conjunto maior de dados, coletados para a realização de um projeto de pós-doutoramento, que está em andamento[4]. O termo bem-estar ou bem-estar subjetivo costuma ser utilizado na literatura como sinônimo de felicidade, pressupondo que o próprio indivíduo é quem pode melhor avaliar sua satisfação para com a vida. O termo está ligado à psicologia positiva e é um conceito complexo que integra as dimensões cognitiva e afetiva, sendo incorporado num abrangente campo de estudo que abarca outros grandes conceitos e domínios de estudo, tais como o afeto positivo, o negativo e a qualidade de vida. Existem diferentes meios para se avaliar o bem-estar subjetivo havendo uma tendência pela aproximação entre essas formas de medidas (Diener & Seligman, 2004). Escalas envolvendo itens múltiplos como o Questionário de Felicidade Oxford proposto por Hills e Argyle (Hills & Argyle, 2002), a Escala de Felicidade Subjetiva construída por Lyubomirsky e Lepper (Lyubomirsky & Lepper, 1999) e Índice de Felicidade Pemberton desenvolvido por Gonzalo Hervás e Carmelo Vázquez (Hervás & Vázquez, 2013) costumam ser indicadas para a avaliação quando mais de um componente do bem-estar subjetivo é de interesse. Porém, instrumentos como o Item de Felicidade Global, desenvolvido por Bradburn (Bradburn, 1969) e que envolve uma única questão, são considerados vantajosos por serem de rápida aplicação, mas existem muitas críticas quanto a seu uso.

[4] Pós-Doutorado em andamento pelo Departamento de Psicologia da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo sob a supervisão do Dr. José Aparecido da Silva. Esta pesquisa inclui ainda um estágio no Institut des Sciences de L’Homme, Lyon, França sob a supervisão do Dr. Djamel Abdelkader Zighed e apoio da FAPESP.

Quanto ao tamanho das escalas e à quantidade de itens, o Índice de Felicidade Pemberton envolve uma escala de 21 itens onde as primeiras 11 questões são respondidas em uma Escala Likert de dez pontos e as demais 10 questões são respondidas de forma dicotômica, o Questionário de Felicidade Oxford apresenta um total de 29 itens para serem respondidos em uma Escala Likert de seis pontos; a Escala de Felicidade Subjetiva é um modelo de quatro itens derivado de uma combinação inicial de 13 itens onde cada item deve ser respondido dentro de uma Escala Likert de sete pontos. O Item de Felicidade Global consiste unicamente na questão "Considerando todas as coisas, quão feliz você está nos dias atuais?" que deve ser respondida em uma Escala Likert de sete pontos, variando de "Não muito feliz" até "Muito feliz".

O campo do bem-estar subjetivo, enquanto disciplina científica, tem crescido rapidamente e há várias justificativas para tal. A primeira está ligada ao fato de estarmos entrando num mundo pós-materialista, em que as pessoas se interessam, muito mais que outrora, por questões relativas à qualidade de vida. A segunda justificativa para este crescimento está na constatação de que estimativas do bem-estar subjetivo são meios, particularmente democráticos, por respeitar o que as pessoas pensam sobre suas próprias vidas. Por fim, o estudo do bem-estar subjetivo floresceu devido ao crescimento do individualismo ao redor do mundo, pois quando estes indivíduos se preocupam com seus próprios sentimentos e crenças, colocam os estudos do bem-estar subjetivo em perfeita sintonia com a cultura ocidental (Diener, Lucas & Oishi, 2002).

2 Fundamentação teórica

A Psicologia Positiva é o estudo de sentimentos, emoções, instituições e comportamentos positivos que tem como objetivo final a promoção da felicidade humana (Seligman, 2004). É o estudo científico das forças e virtudes próprias do indivíduo, numa perspectiva mais apreciativa em relação ao potencial, motivação e capacidades humanas (Sheldon & King, 2001). A utilização do termo Psicologia Positiva pressupõe a existência de uma Psicologia Negativa. Numa breve análise da palavra psicologia, considerando sua origem etimológica, encontramos em sua gênese as palavras psiqué e logos. Apesar da tradução exata de psiché ser alma, Graziano (2005) explica que o conceito de alma para o grego está muito mais próximo do que hoje chamamos de mente, uma vez que a palavra alma acabou impregnada de um caráter quase que exclusivamente religioso. Ao traduzirmos logos como estudo, alcançamos o popular significado da palavra psicologia como sendo o estudo da mente. Mas, se a Psicologia é a ciência que estuda a mente, que deve contemplar toda a complexidade da mente humana e que deve compreender todos os nossos processos internos, então esta distinção entre positivo e negativo não deveria fazer sentido. Então, Seligman (2004) justifica a utilização do termo Psicologia Positiva, quando realiza uma análise histórica da atuação do psicólogo nos Estados Unidos, apontando para o momento em que estes profissionais passaram a atuar como psicoterapeutas para suprir a demanda de apoio aos soldados que retornavam da Segunda Guerra Mundial. Foi neste momento que o foco da psicologia passou a ser o tratamento da doença mental. Ele aponta a criação do Instituto Nacional de Saúde Mental (National Institute of Mental Health), pelo Congresso Americano, como sendo uma ação utilizada para garantir a manutenção dessa tendência na Psicologia. Este instituto, dirigido por

psiquiatras, passou a financiar apenas pesquisas que demonstrassem relevância para a causa e a cura das doenças mentais. Assim, apesar deste modo de conduzir a Psicologia ter promovido enormes avanços tanto para a compreensão quanto para o tratamento das doenças mentais, a atenção que se empregou no alívio destes transtornos fez diminuir a preocupação e maior atenção com situações que fazem a vida valer a pena (Seligman, 2004, p.11). Desta forma, surgiu nos Estados Unidos, em 1998, o movimento científico que recebeu o nome de Psicologia Positiva, iniciado por Martin Seligman que, ao lado de outros cientistas como Mihaly Csikszentmihalyi, Ray Fowler, Chris Peterson, George Vaillant e Ed Diener, começou a desenvolver pesquisas utilizando o métodos quantitativos, a fim de promover uma mudança de foco na Psicologia atual centrada no estudo de algumas das piores coisas da vida para o estudo focado no que faz a vida ser compensadora e gratificante (Seligman, 2004).

Um esclarecimento necessário, diz respeito ao uso que a Psicologia Positiva faz dos termos positivo e negativo para qualificar as emoções. De acordo com Fredrickson (2001), as chamadas emoções positivas são aquelas que favorecem a aproximação e, consequentemente, a convivência com o outro, enquanto que as emoções negativas fazem o contrário. As emoções positivas fortalecem nossos repertórios físicos, sociais e intelectuais, criando recursos que podemos acessar quando houver uma oportunidade, ou uma ameaça ambiental. Inclusive, o estado de espírito positivo desperta a afetividade do outro, criando um ambiente favorável que, além de contribuir para o desenvolvimento da resiliência psicológica, desencadeia uma espiral ascendente em direção à melhoria do bem-estar emocional. Do ponto de vista do desenvolvimento humano, isso nos faz concluir que um funcionamento ótimo jamais poderia prescindir das emoções positivas. Seligman e Csikszentmihalyi (2000) apontam ainda para o fato da Psicologia Positiva ter também o mérito de resgatar o caráter preventivo da Psicologia. Para eles, através do estudo das características humanas positivas, a ciência aprenderá a prevenir doenças físicas e mentais, e os psicólogos, então, aprenderão a desenvolver qualidades que ajudem indivíduos e comunidades a não somente resistir e sobreviver, mas também, florescer. Seligman (2004) diz que a Psicologia Positiva está pautada sobre três bases. São elas: 1) O estudo da emoção positiva; 2) O estudo dos traços ou qualidades positivas, principalmente, forças e virtudes, incluindo também habilidades como inteligência e capacidade atlética; 3) O estudo das chamadas instituições positivas, como a democracia, a família e a liberdade que dão suporte às virtudes que, por sua vez, apoiam as emoções positivas. Sendo que, embora se reconheça a importância destas três bases, este estudo foca a felicidade, apoiando-se na primeira base apontada.

A preocupação humana com a felicidade é bastante antiga, como atesta a Filosofia. Para Aristóteles, que viveu de 384 a 322 a.C., a felicidade era o objetivo da ética que, para ele, era a vida boa, a vida digna. Aristóteles acreditava que a felicidade só seria conquistada através do exercício das virtudes, em sintonia com a vida em sociedade (Aristóteles: Ética a Nicômaco, 1987). Para Epicuro, que viveu de 341 a 270 a.C., uma existência feliz seria possível através da recusa dos excessos, medos e compromissos que podiam levar a sofrimentos inúteis (Epicuro, Carta sobre a Felicidade, 1994). Já Sêneca, que viveu do ano 2 a.C. a 65 d.C., e negava a sociedade em que vivia como padrão de referência por observá-la infeliz, dizia que, para ser feliz, a primeira coisa que o indivíduo deveria fazer seria recusar-se seguir a multidão (Sêneca: Da vida feliz, 1997). Já no início do século XX, o filósofo Bertrand Russel acreditava que a felicidade

é um bem a ser conquistado e não uma dádiva divina, apontando para a importância do homem integrar-se com a sociedade, buscando a felicidade de outrem na mesma medida em que busca a sua própria felicidade (Russel, 2003). Graziano (2005) afirma que não seria exagero imaginarmos que a busca pela felicidade tenha sido a grande mola propulsora a conduzir a espécie humana em sua trajetória pelo mundo. Que o desejo por uma vida melhor foi a norteadora das grandes invenções. Freud, no entanto, numa perspectiva mais contemporânea, acreditando na essência humana como algo eminentemente ruim, condenou o homem a uma angústia eterna, sofrendo numa luta igualmente eterna para manter sua essência sob controle. Na obra Além do Princípio do Prazer (Freud, 1998 [1920] apud Graziano, 2005), ele discorre sobre o que chamou de instinto de morte, contestando que o objetivo da vida pudesse ser alcançar um estado que jamais fora alcançado antes. Considerando sua grande influência no pensamento ocidental, as ideias de Freud contribuíram para que uma onda de pessimismo se disseminasse neste período. Quanto à natureza da felicidade, Graziano (2005) observou que a concepção humana esteve baseada em premissas de natureza extrínseca e natureza intrínseca. A premissa que crê na natureza extrínseca da felicidade leva o indivíduo a buscá-la para além de si mesmo como, por exemplo, o hedonismo, por acreditar que o segredo da felicidade consiste no aumento dos bons e diminuição dos maus momentos da vida. Já a concepção de felicidade que parte de uma premissa intrínseca, enxerga o próprio indivíduo como sua fonte, conferindo a ele a tarefa de trabalhar a si mesmo de forma a conquistar uma vida feliz.

A Psicologia Positiva trabalha com o conceito de bem-estar subjetivo que consiste nas avaliações que as pessoas fazem da própria vida, tanto no aspecto afetivo quanto no aspecto cognitivo sendo frequentemente usado como sinônimo de felicidade, satisfação e afeto positivo. Constata-se o bem-estar subjetivo quando se sente prazer e poucas emoções desagradáveis, quando se está satisfeito com a própria vida (Diener & Lucas, 2000). De acordo com Diener (2000), embora as pessoas façam julgamentos sobre suas vidas como um todo, o bem-estar subjetivo possui diferentes componentes: a) Satisfação com a vida que corresponde aos julgamentos que se faz em relação à própria vida; b) Satisfação em áreas importantes, em diferentes domínios, como o trabalho, por exemplo; c) Emoções positivas envolvendo muitas experiências de emoções e humores agradáveis e; d) Baixos níveis de emoções negativas envolvendo experiência de poucas emoções e humores desprazeirosos.

As pesquisas no campo da felicidade tanto podem investigar esses componentes separadamente, quanto podem usar medidas globais para avaliar o bem-estar subjetivo. Neste sentido, ainda é válido apresentar a constatação de que, apesar de as condições de vida exercer uma grande influência no bem-estar subjetivo (Diener, Lucas & Oishi, 2002), tal influência não é suficiente para que encontremos uma relação direta entre variáveis tais como prosperidade e felicidade (Csikszentmihalyi, 1999). O que se observa é que ocorre uma estabilidade nos níveis de bem-estar das pessoas, que não pode ser explicado através da estabilidade das condições sob as quais elas vivem, nos levando a entender que o bem-estar subjetivo é fortemente influenciado por disposições de personalidade estáveis que refletem a tendência do indivíduo ao reagir cognitiva e emocionalmente em relação às circunstâncias da vida (Diener, Lucas & Oishi, 2002).

Segundo Seligman (2004) em menos de três meses, eventos importantes, sejam eles positivos ou negativos, perdem o impacto sobre os níveis de bem-estar subjetivo, fazendo com que as pessoas voltem a apresentar seus níveis médios de felicidade

prévios. Esta estabilidade está ligada ao fato de o bem-estar subjetivo envolver um componente genético substancial que faz com que, em algum grau as pessoas nasçam prontas para serem mais ou menos felizes (Diener, Lucas & Oishi, 2002). Quanto ao Brasil, a edição 2013 do Relatório sobre a Felicidade Mundial (World Happiness Report) aponta o país na 24ª posição (Jonh, Layard & Sachs, 2013). Trata-se de uma pesquisa mundial baseada nos estudos de cultura política coordenada pela Universidade de Michigan nos Estados Unidos e que conta com a adesão de universidades de dezenas de países que contribuem para a elaboração, tradução para a língua do seu país e aplicação de um questionário a cada quatro ou seis anos. Dos 156 países analisados, o documento mostrou que Dinamarca, Noruega, Suíça, Holanda e Suécia ocupam as primeiras cinco posições, respectivamente. Ao se comparar amplas amostras populacionais de diferentes países verifica-se grandes e estáveis diferenças nos níveis de felicidade. De acordo com Ingleman e Klingemann (2000), é comum nações mais ricas reportarem índices de felicidade mais elevados quando comparadas a nações mais pobres. Indivíduos que vivem nas regiões mais ricas do globo como América do Norte, Austrália, Europa Ocidental e Japão manifestam escores de felicidade mais altos do que aqueles indivíduos que vivem em regiões mais pobres como África e Ásia. Contudo, como aponta Veenhoven (1991), esta informação não explica como muitos países da América Latina apresentam índices médios de felicidade comparáveis aos da Austrália e do Japão. De acordo com Ferraz, Tavares e Zilberman (2007) não existem pesquisas científicas sobre felicidade considerando globalmente a população brasileira, exceto os dados de pesquisas internacionais como o Relatório sobre a Felicidade Mundial. O relatório recomenda inclusive, que os líderes políticos devem se atentar aos índices que afetam a felicidade já que pessoas felizes vivem mais, são mais produtivas, ganham mais e também são melhores cidadãs.

2.1 O Índice de Felicidade Pemberton

O Índice de Felicidade Pemberton foi desenvolvido pelos pesquisadores Gonzalo Hervás e Carmelo Vázquez da Universidade Complutense de Madrid com o apoio do Instituto Coca-Cola da Felicidade. O índice envolve 21 declarações (Quadro 1), onde onze delas estão relacionadas a diferentes domínios do bem-estar lembrado, tais como geral, hedonista, eudaimonico e bem-estar social, e dez relacionadas ao bem-estar vivenciado, ou seja, os eventos emocionais positivos e negativos que possivelmente aconteceram no dia anterior. A soma destes itens produz um índice de bem-estar combinado (Hervás & Vázquez, 2013). A característica distintiva deste estudo é que, para construir a escala, um conjunto inicial de itens, que abrange os domínios de bem-estar lembrado e vivenciado, foi submetido a um processo de validação e seleção completa. Esses itens foram baseados em escalas validadas como a Escala de Afeto Positivo e Negativo (PANAS), a Escala de Satisfação com a Vida, a Escala de Bem-estar Subjetivo e a Escala de Bem-Estar Psicológico (Hervás & Vázquez, 2013). De acordo com Hervás e Vázquez (2013), os itens iniciais e as escalas de referência foram traduzidos em sete línguas e completados via Internet por uma amostra (N = 4052), com idade entre 16 e 60 anos de nove países, incluindo Alemanha, Índia, Japão, México, Rússia, Espanha, Suécia, Turquia e EUA. Os participantes foram capturados em painéis de pesquisa, ou seja, grupos de pessoas que aceitam participar, regularmente, de pesquisas sociais, do Millward Brown, uma companhia de pesquisa que atua no mundo

inteiro. A composição da amostra foi heterogênea no que se refere ao gênero, níveis socioeconômico e educacionais.

Os resultados iniciais deste estudo demonstraram bons indicadores psicométricos do IFP, tais como consistência interna, estrutura unifatorial, validade convergente e incremental. Com relação à permissão de uso, IFP é de domínio público, sem direitos autorais e, portanto, pesquisadores e profissionais são livres para usá-lo, desde que créditos sejam atribuídos aos autores da escala. Dado as boas propriedades psicométricas do IFP, os autores inclusive, sugerem a utilização desta escala considerada simples e integrativa, como instrumento para monitorar as alterações no bem-estar (Hervás & Vázquez, 2013).

DOMÍNIOS E SUBDOMÍNIOS	ITENS
BEM-ESTAR LEMBRADO	
Bem-Estar Geral	
	Estou muito satisfeito/a com a minha vida.
	Tenho energia para realizar minhas tarefas diárias.
Bem-Estar Eudaimônico	
Sentido da Vida	Acho que minha vida é útil e vale a pena.
Auto Aceitação	Estou satisfeito/a comigo mesmo/a.
Crescimento Pessoal	Minha vida é cheia de experiências de aprendizado e desafios que me fazem crescer.
Relacionamentos	Eu me sinto muito conectado/a com as pessoas ao meu redor.
Percepção de Controle	Eu me sinto capaz de resolver a maioria dos meus problemas diários.
Autonomia	Eu acho que posso ser eu mesmo/a com relação a coisas importantes.
Bem-Estar Hedônico	
Afeto Positivo	Eu aprecio muitas coisas todos os dias.
Afeto Negativo	Tenho muitos maus momentos na minha vida diária.
Bem-Estar Social	Eu acho que vivo numa sociedade que permite realizar plenamente meu potencial.
BEM-ESTAR VIVENCIADO	
Experiências Positivas	
	Algo que fiz me deixou orgulhoso/a.
	Às vezes, me senti sobrecarregado/a.
	Eu fiz algo divertido com alguém.
	Fiquei entediado/a por muito tempo.
	Fiz algo que eu realmente apreciei fazer.
Experiências Negativas	
	Eu estava preocupado/a com assuntos de caráter pessoal.
	Aprendi algo interessante.

	Eu me presenteei com uma guloseima.
	Algumas coisas aconteceram que me deixaram muito bravo/a.
	Senti-me desrespeitado/a por alguém.

Quadro 1- Domínios, Subdomínios e Itens do Índice de Felicidade Pemberton.(**Fonte:** Tradução de Hervás e Vázquez, 2013.)

3 Metodologia

3.1 Origem do estudo

O presente artigo apresenta dados parciais de uma pesquisa intitulada "Validação do Índice de Felicidade Pemberton para o Brasil". Os dados empíricos deste estudo fazem parte de um conjunto maior de dados, coletados para a realização de um projeto de pós-doutoramento, que está em andamento. Trata-se de um projeto que teve inicialmente como objetivo geral disponibilizar o Índice de Felicidade Pemberton (IFP), para uso no Brasil. O IFP é uma medida integrada de bem-estar subjetivo que incorpora o bem-estar geral, hedônico, eudaimônico e social, envolvendo perspectivas de bem-estar vivenciado e lembrado. Como objetivos específicos, o projeto propôs realizar a tradução e adaptação cultural do IFP; avaliar a fidedignidade do instrumento através dos coeficientes de correlação para consistência interna, divisão pela metade e teste-reteste, entre outros; analisar a validade de conteúdo e a validade aparente dos itens que compõem o instrumento; analisar a validade total do instrumento a partir das análises fatorial exploratória e confirmatória; analisar a validade concorrente aferida a partir da comparação com o Questionário de Felicidade Oxford. Posteriormente, foi acrescentado aos objetivos do projeto, realizar a validação do IFP também na França e após a validação, comparar com os índices do Brasil. Escolheu-se a França para realizar esta parte do projeto, visto que o Índice de Felicidade Pemberton já foi validado para nove idiomas e para sete países que não incluem os idiomas português e francês, nem os países Brasil e França. Além disso, a versão 2013 do Mapa Mundial da Felicidade da Universidade de Leicester mostrou que ambos os países apresentam índices de felicidade próximos – o Brasil ocupa 24ª posição e a França ocupa a 25ª posição do ranking – sendo que uma análise comparativa mais aprofundada pode trazer informações importantes para caracterizar melhor as diferenças e as similaridades entre os dois países. Para a elaboração deste artigo e apresentação deste resultado parcial, selecionamos apenas os dados de estudantes brasileiros da área de ciências exatas que participaram da fase do teste e que correspondem à 199 participações.

3.2 Amostra, local e período da coleta

A amostragem foi regida por critérios de conveniência e/ou de disponibilidade dos inquiridos diante da condição de inclusão de que o participante fosse brasileiro, estudante universitário na região de Ribeirão Preto, estado de São Paulo, Brasil, com idade entre 16 e 60 anos. Os questionários foram aplicados entre os meses de agosto e setembro de 2014 e para sua aplicação, utilizou-se o aplicativo Google forms. Participaram do teste 544 estudantes universitários das áreas de humanas, exatas e biológicas. O re-teste foi aplicado à 72 participantes nos mês de outubro de 2014. As

respostas dos estudantes da área de ciências exatas analisadas neste artigo foram extraídas desta amostra, sendo que o Quadro 2 aponta a quantidade de participações por área de conhecimento.

Quantidade	Área
340	Ciências Humanas
199	Ciências Exatas
05	Ciências Biológicas

Quadro 2- Características da amostra geral por área do conhecimento. (**Fonte:** Produzido pelos autores.)

3.3 Instrumentos de coleta de dados

Os instrumentos para a coleta dos dados foram:

Questionário Sociodemográfico com questões sobre dados pessoais, econômicos, situacionais, de contexto social, saúde, religiosidade e postura sociopolítica.

Índice de Felicidade Pemberton, após tradução e adaptação cultural para língua portuguesa e brasileira.

Questionário de Felicidade Oxford.

Termo de Consentimento Livre e Esclarecido.

4 Resultados e Discussão

Os itens do IFP, constantes no Quadro 3, foram usados para analisar o bem-estar lembrado dos estudantes brasileiros. Os itens foram respondidos em escala Likert de 10 pontos, conforme o modelo proposto pelos autores do índice, onde o avaliado foi convidado para manifestar o seu grau de concordância desde discordo totalmente (nível 1), até concordo totalmente (nível 10).

BEM-ESTAR LEMBRADO	
Pp1	Estou muito satisfeito/a com a minha vida.
Pp2	Tenho energia para realizar minhas tarefas diárias.
Pp3	Acho que minha vida é útil e vale a pena.
Pp4	Estou satisfeito/a comigo mesmo/a.
Pp5	Minha vida é cheia de experiências de aprendizado e desafios que me fazem crescer.
Pp6	Eu me sinto muito conectado/a com as pessoas ao meu redor.
Pp7	Eu me sinto capaz de resolver a maioria dos meus problemas diários.
Pp8	Eu acho que posso ser eu mesmo/a com relação a coisas importantes.
Pp9	Eu aprecio muitas coisas todos os dias.
Pp10	Tenho muitos maus momentos na minha vida diária.
Pp11	Eu acho que vivo numa sociedade que permite realizar plenamente meu potencial.

Quadro 3- Itens do Bem-Estar Lembrado.(**Fonte:** Produzido pelos autores.)

Também apresentamos no Quadro 4 as ocorrências, a média e o desvio padrão da amostra.

n col 11, n fil 200

	Ocorrência	Média	Desvio padrão
Pp1	152.81	0.76	0.20
Pp2	100.98	0.50	0.32
Pp3	50.49	0.25	0.35
Pp4	102.30	0.51	0.34
Pp5	82.28	0.41	0.37
Pp6	96.47	0.48	0.34
Pp7	108.46	0.54	0.35
Pp8	86.68	0.43	0.39
Pp9	93.83	0.47	0.35
Pp10	74.47	0.37	0.27
Pp11	99.99	0.50	0.25

Quadro 4- Ocorrência, média e desvio padrão da amostra. (**Fonte:** Gerado pelo software CHIC.)

4.1 Análise das similaridades

A análise de similaridade efetua uma ponderação das proximidades segundo I. C. LERMAN visando formar partições mais e mais refinadas sobre um conjunto de n variáveis, de maneira ascendente e apresentadas sob a forma de uma "árvore", a partir de um critério de similaridade entre variáveis (COUTURIER et al., 2004). O valor máximo dos níveis de similaridade é igual a 1. A árvore gerada pelo software CHIC (Figura 1) apresenta as relações de similaridade encontradas entre todos os itens que compõem o bem-estar vivenciado. As relações similares identificadas como as mais significativas estatisticamente, os nós significativos, foram destacadas na análise. Esses nós se formam a partir dos índices de similaridade entre os itens, sendo que, quanto mais próximo de 1 for este índice, maior será a similaridade entre os itens ou os grupos de itens.

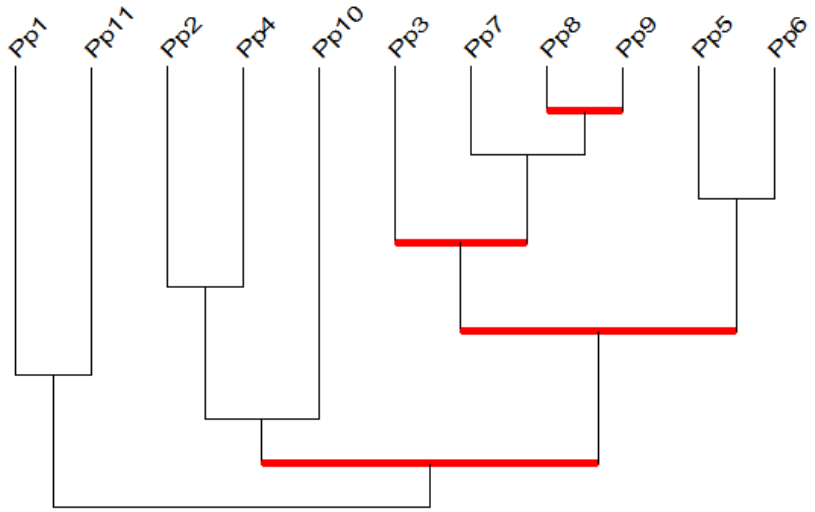


Figura 1- Árvore de Similaridade. (**Fonte:** Produzida pelo software CHIC.)

O nó mais relevante é o que une os itens Pp8 e Pp9 (Pp8 Pp9), cujo nó tem grau de similaridade 0.966. Observamos, então, que os indivíduos fizeram apontamentos muito próximos sobre as afirmativas "Eu acho que posso ser eu mesmo/a com relação a coisas

importantes” e “Eu aprecio muitas coisas todos os dias”. O segundo nó mais relevante une o item PP7 “Eu me sinto capaz de resolver a maioria dos meus problemas diários” com os itens do primeiro nó (Pp7 (Pp8 Pp9), com grau de similaridade 0.917. Observamos, a partir dessas análises que há uma aproximação entre a percepção da mostra sobre percepção de controle, autonomia e afeto positivo.

O terceiro nível, na árvore de similaridade aposta os itens Pp5 e PP6 (Pp5 Pp6) com similaridade de 0.8652. Temos então padrões similares de percepção de crescimento pessoal com relacionamentos interpessoais representados pelos itens “Minha vida é cheia de experiências de aprendizado e desafios que me fazem crescer” e “Eu me sinto muito conectado com as pessoas ao meu redor”.

Os demais níveis de similaridade foram menores que 0.85: nível 4 (Pp3 (Pp7 (Pp8 Pp9))) similaridade 0.8489; nível 5 (Pp2 Pp4) similaridade 0.8167; nível 6 ((Pp3 (Pp7 (Pp8 Pp9))) (Pp5 Pp6)) similaridade 0.6612; nível:7 (Pp1 Pp11) similaridade 0.5724; nível 8 ((Pp2 Pp4) Pp10) similaridade 0.4504; nível 9 (((Pp2 Pp4) Pp10) ((Pp3 (Pp7 (Pp8 Pp9))) (Pp5 Pp6))) similaridade 0.2246; nível 10 ((Pp1 Pp11) (((Pp2 Pp4) Pp10) ((Pp3 (Pp7 (Pp8 Pp9))) (Pp5 Pp6)))) similaridade 2.5221. O coeficiente de correlação dos itens pode ser observado a seguir no Quadro 5.

	Pp1	Pp2	Pp3	Pp4	Pp5	Pp6	Pp7	Pp8	Pp9	Pp10	Pp11
Pp1		-0.11	-0.40	-0.16	-0.20	-0.25	-0.14	-0.15	-0.17	-0.21	0.16
Pp2			0.10	0.29	0.27	0.24	0.20	0.25	0.43	0.16	0.06
Pp3				0.20	0.16	0.17	0.18	0.28	0.22	0.18	-0.08
Pp4					0.20	0.27	0.31	0.28	0.27	0.08	0.03
Pp5						0.27	0.26	0.33	0.32	0.13	0.01
Pp6							0.37	0.26	0.29	0.14	-0.01
Pp7								0.43	0.34	0.07	-0.06
Pp8									0.42	0.09	-0.01
Pp9										0.11	-0.05
Pp10											0.03
Pp11											

Quadro 5- Coeficiente de correlação dos itens.(**Fonte:** Produzido pelo software CHIC.)

4.2 Análise das quasi-implicações

A árvore implicativa exprime graficamente o encaixamento sucessivo das classes constituídas com sentido implicativo. O índice implicativo é realizado pelo cálculo de coesão, e pode ser calculado através da lei de Poisson (COUTURIER et al., 2004). O grafo implicativo apresentado na Figura 2 indica que todas ligações com flechas vermelhas possuem um intervalo de confiança de 0,95, e as ligações com flecha em azul apresentam um intervalo de 0,90, e portanto, apresentam uma forte relação. A interpretação dos resultados do grafo inicia pela parte superior, e vai implicando até a base do grafo. A percepção de Pp5 “Minha vida é cheia de experiências de aprendizado e desafios que me fazem crescer” implica na percepção de Pp8 “Eu acho que posso ser eu mesmo/a com relação a coisas importantes” que por sua vez implica na percepção de Pp9 “Eu aprecio muitas coisas todos os dias” que sugere a percepção de Pp2 Tenho energia para realizar minhas tarefas diárias”. Da mesma forma, o item Pp3 “Acho que minha vida é útil e vale a pena” implica os itens Pp8, Pp9 e Pp2.

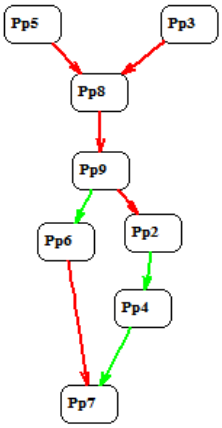


Figura 2- Grafo Implicativo. (**Fonte:** Produzido pelo software CHIC.)

Os índices das implicações entre os itens do IFP obtidos segundo a teoria clássica e calculados usando a lei de Poisson podem ser observados no Quadro 6.

	Pp1	Pp2	Pp3	Pp4	Pp5	Pp6	Pp7	Pp8	Pp9	Pp10	Pp11
Pp1	0	41	27	37	35	32	38	38	37	38	60
Pp2	29	0	63	90	86	85	82	86	96	68	58
Pp3	0	74	0	91	83	85	89	96	92	78	35
Pp4	21	90	75	0	80	88	93	90	88	60	53
Pp5	12	92	74	86	0	92	94	96	96	68	52
Pp6	10	86	72	90	88	0	97	88	91	68	48
Pp7	24	79	72	91	86	94	0	97	92	58	43
Pp8	17	90	87	94	96	91	99	0	99	63	48
Pp9	18	98	79	91	93	92	96	98	0	65	43
Pp10	13	76	73	65	71	74	63	67	70	0	55
Pp11	74	58	42	54	52	48	42	48	44	53	0

Quadro 6- Índices de implicações. (**Fonte:** Produzido pelo software CHIC.)

4.3 Análise da coesão

Os índices de coesão segundo a teoria clássica e calculados usando a lei de Poisson podem ser observados no Quadro 7.

	Pp1	Pp2	Pp3	Pp4	Pp5	Pp6	Pp7	Pp8	Pp9	Pp10	Pp11
Pp1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24
Pp2	0.00	0.00	0.31	0.88	0.82	0.79	0.73	0.81	0.97	0.43	0.20
Pp3	0.00	0.57	0.00	0.90	0.76	0.80	0.87	0.98	0.91	0.66	0.00
Pp4	0.00	0.88	0.59	0.00	0.68	0.85	0.93	0.88	0.85	0.23	0.08
Pp5	0.00	0.91	0.56	0.81	0.00	0.92	0.94	0.98	0.96	0.43	0.06
Pp6	0.00	0.82	0.51	0.88	0.85	0.00	0.98	0.86	0.90	0.42	0.00
Pp7	0.00	0.68	0.52	0.90	0.81	0.94	0.00	0.98	0.92	0.19	0.00
Pp8	0.00	0.88	0.82	0.94	0.96	0.90	1.00	0.00	0.99	0.32	0.00
Pp9	0.00	0.99	0.66	0.89	0.93	0.91	0.97	0.99	0.00	0.35	0.00
Pp10	0.00	0.61	0.54	0.35	0.50	0.57	0.31	0.40	0.46	0.00	0.12
Pp11	0.57	0.20	0.00	0.09	0.05	0.00	0.00	0.00	0.00	0.08	0.00

Quadro 7 - Índices de coesão. (**Fonte:** Produzido pelo software CHIC.)

Na árvore coesitiva apresentada na Figura 3, podemos visualizar as inter-relações dos itens estudados em uma árvore hierárquica. A árvore coesitiva, também chamada

hierarquia ascendente, demonstra graficamente o ajeitamento sucessivo das classes constituídas de acordo com o critério de coesão, em sentido decrescente, conforme os níveis da hierarquia, sendo que isso ocorre no sentido contrário da formação das classes de variáveis. As setas demonstram a direção da implicação e a cor vermelha identifica as mais significativas. O valor máximo do índice de coesão é igual a 1.

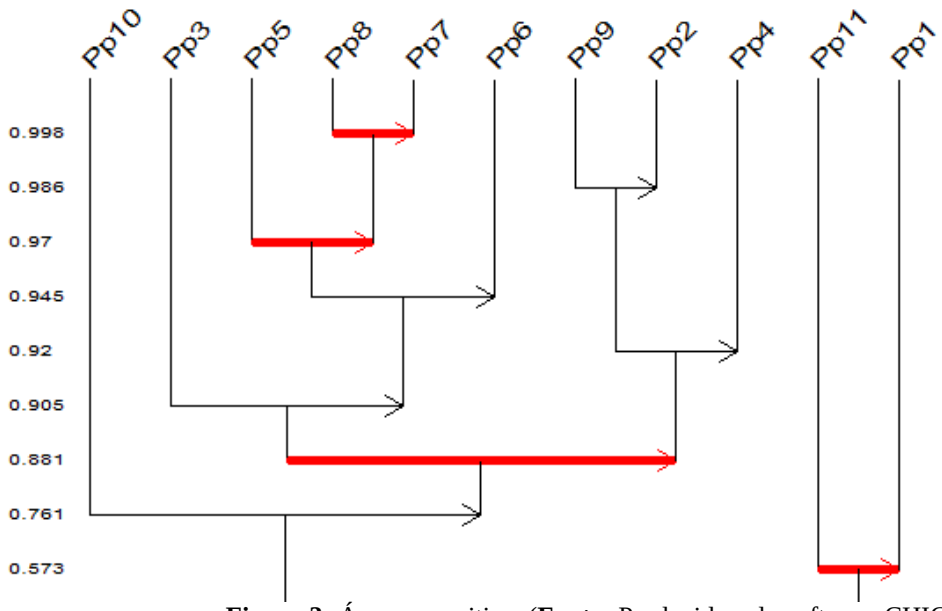


Figura 3- Árvore coesitiva. (**Fonte:** Produzida pelo software CHIC.)

Aqui, as avariáveis Pp10, Pp3, Pp5, Pp8, Pp7, Pp6, Pp9, Pp2 e pp4 formam uma classe, com subclasses apresentando implicações mais fortes. As variáveis Pp11 e Pp1 formam outra classe. A leitura da árvore deve ser pautada por meio das setas. As maiores associações entre as variáveis são identificadas de duas formas: uma forma referente à seta destacada na cor vermelha, que é denominada pelo programa, como nós significativos utilizados para chamar a atenção para a classe, pois ali se encontra as variáveis com os indícios de maiores implicações. A segunda forma, referente à distância do eixo superior à base da hierarquia, quanto mais próxima do eixo, maior é a implicação entre as variáveis. Nesse sentido, a análise das classes deve ser realizada respeitando a sua ordem de qualidade de coesão decrescente.

O nó mais significativo (Pp8 Pp7) com o índice de coesão 0.998 é dado pela relação entre as percepções de “Eu acho que posso ser eu mesmo/a com relação a coisas importantes” e “Eu me sinto capaz de resolver a maioria dos meus problemas diários”. Assim, pode-se dizer que a percepção de autonomia implica a percepção de controle.

O segundo nó, faz a ligação entre os itens Pp9 e Pp2 (Pp9 Pp2) com índice de coesão 0.986 onde a percepção de Afeto Positivo implica na percepção do Bem-estar Geral. Essa classificação de nível 2 faz a relação da afirmação “Eu aprecio muitas coisas todos os dias” com a afirmação “Tenho energia para realizar minhas tarefas diárias.

A classificação ao nível 3, com coesão 0.97, é representado pelo nó (Pp5 (Pp8 Pp7)) onde a afirmativa “Minha vida é cheia de experiências de aprendizado e desafios que me fazem crescer” implica nas afirmativas do primeiro nó “Eu acho que posso ser eu mesmo/a com relação a coisas importantes” e “Eu me sinto capaz de resolver a maioria dos meus problemas diários”.

Os demais níveis de coesão são representados pelas setas no nível 4 ((Pp5 (Pp8 Pp7)) Pp6) com coesão 0.945, ao nível 5 ((Pp9 Pp2) Pp4) com coesão 0.92, ao nível 6 (Pp3 ((Pp5 (Pp8 Pp7)) Pp6)) com coesão 0.905, ao nível 7 ((Pp3 ((Pp5 (Pp8 Pp7)) Pp6)) ((Pp9 Pp2) Pp4)) com coesão 0.881, ao nível 8 (Pp10 ((Pp3 ((Pp5 (Pp8 Pp7)) Pp6)) ((Pp9 Pp2) Pp4))) com coesão 0.761 e ao nível 9 (Pp11 Pp1) com coesão 0.573.

Références

[1] Aristóteles (1987). *Ética a Nicômaco. Coleção os Pensadores*. São Paulo: Nova Cultural

[2] Bradburn, N. M. (1969). The *structure of psychological well-being*. Chicago: Aldine.

[3] Couturier, R.; et al. *A classificação hierárquica implicativa e coesiva.* Manual Curso CHIC versão 2.3; 2004. Tradução S. Ag Almouloud.

[4] Csikszentmihalyi, M. (1999). If we are so rich, why aren t we happy?. *American Psychologist*, **54,** 821 835.

[5] Diener, E. (2000). Subjetive well-being: The science of happiness and a proposal for a national index. . *American Psychologist*, **55,** 34 - 43.

[6] Diener, E.; Lucas R. E. & Oishi (2002). Subjetive well-being: The science of happiness and life satisfaction. In: Snyder, C. R. & Lopez, S. orgs. *Handbook of Positive Psychology.* New York, Oxford,p. 63- 73.

[7] Diener, E. & Lucas, R. E. (2000). Subjective emotional well-being. In: Lewis M; Haviland. J.M. (Eds.). *Handbook of emotions.* 2 ed. New York: Guilford, p.325-337.

[8] Diener, E.; & Seligman, M. E. P. (2004). Beyond money: toward an economy of well-being. *Psychological Science in the Public Interest*, **5 (1)**, 1-31.

[9] Epicuro (1994). *Carta sobre a felicidade.* Tradução de João Forte. Lisboa: Relógio D Água Editores.

[10] Ferraz, R. B.; Tavares, H. & Zilberman, M. L. (2007). Felicidade: Uma Revisão. *Revista de Psiquiatria Clínica*, São Paulo, v. 34, **n. 5**. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-60832007000500005&lng=en&nrm=iso>. Acesso em: 2013-03-14.

[11] Fredrickson, B. L.(2001). The role of positive emotions in positive psychology: The broaden-and-built theory of positive emotions. *American Psychologist*, **55,** 218 - 226.

[12] Graziano, L. D. (2005). *A felicidade revisitada: um estudo sobre bem-estar subjetivo na visão da psicologia positiva.* Tese (Doutorado em Psicologia Escolar e do Desenvolvimento Humano) - Instituto de Psicologia, Universidade de São Paulo, 2005. Disponível em: <http://www.teses.usp.br/teses/disponiveis/47/47131/tde-23052006-164724/>. Acesso em: 2012-12-11.

[13] Hervás, G. & Vázquez, C. (2013). Construction and validation of a measure of integrative well-being in seven languages: The Pemberton Happiness Index. *Health & Quality of Life Outcomes,* **11,** 66.

[14] Hills, P. & Argyle, M. (2002). *Personality and Individual Differences,* **33 (7)**, 1073-1082.

[15] Ingleman, R. & Klingemann, H.D. (2000).Genes, culture, democracy and happiness. In: Diener, E.; Suh, E.M. (eds.) *Culture and subjective well-being.* Cambridge: MIT Press, , pp. 165-167.

[16] John, H.; Layard R. & Sachs J. (2013). *World Happiness Report.* Disponível em: <http://www.earthinstitute.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf>. Acesso em: 2013-09-29.

[17] Lyubomirsky, S. & Lepper, H. S. A. (1999). *Measure of subjective happiness: preliminary reliability and construct validation.* Social Indicators Research, **46 (2)**, 137-155.

[18] Municípios da Região Administrativa de Ribeirão Preto**.** Disponível em: <http://www.ribeiraopreto.sp.gov.br/crp/dados/local/i01regiao.htm>. Acesso em: 2013-03-14.

[19] Passareli-Carrazzoni, P. ; Silva, J. A. Bem-estar subjetivo: autoavaliação em estudantes universitários. *Estudos de Psicologia* (PUCCAMP. Impresso), **v. 29**, p. 415-425, 2012.

[20] Rodrigues, A. (2007). *O bem-estar subjetivo de comerciantes e comerciários de Ribeirão Preto e região.* Dissertação (Mestrado em Psicologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007. Disponível em: <http://www.teses.usp.br/teses/disponiveis/59/59137/tde-15102007-113421/>. Acesso em: 2013-02-06.

[21] Russel, B. (2003). *A conquista da felicidade.* Tradução: Luis Guerra. Rio de Janeiro: Ediouro.

[22] Seligman, M. (2004). *Felicidade Autêntica:* Usando a nova Psicologia Positiva para a realização permanente. Rio de Janeiro: Objetiva.

[23] Seligman, M. & Csikszentmihalyi, M. (2000). Positive Psychology: An introduction. *American Psychologist,* **55,** 5-14.

[24] Sêneca (1994). *Da vida feliz.* Tradução João Forte. Lisboa: Relógio D Água Editores.

[25] Sheldon, K. M. & King, L. (2001) Why positive psychology is necessary. *American Psychologist,* **56,** 216-217.

[26] Veenhoven, R. (1991). Questions on happiness: classical topics, modern answers, blind spots. In: Strack, F.; Argyle, M.; Schwartz, N. (eds.). *Subjective well-being: an interdisciplinary perspective.* Pergamon, New York, pp. 7-26.

LES POTENTIELS PEDAGOGIQUES DE L'USAGE DU TABLEAU NUMERIQUE INTERACTIF A L'ECOLE PRIMAIRE EN FRANCE A LA LUMIERE DE L'ANALYSE STATISTIQUE IMPLICATIVE

Hassan ALCHEGHRI[1]

THE PEDAGOGICAL POTENTIAL OF USE OF THE INTERACTIVE WHITEBOARD IN PRIMARY SCHOOL IN FRANCE IN THE LIGHT OF THE STATISTICAL IMPLICATIVE ANALYSIS

RÉSUMÉ

Le but de cette recherche est d'étudier les potentiels pédagogiques de l'usage du TNI à l'école primaire du point de vue des enseignants utilisant ce dispositif techno-sémio-pédagogique. Nous nous intéressons, dans un premier temps, à étudier les représentations des enseignants envers l'usage du TNI. Nous nous focalisons dans un deuxième temps sur les relations existantes entre certains potentiels pédagogiques de l'usage du TNI. Pour traiter et analyser nos données obtenues à partir de l'échelle de préférence, nous avons eu recours à deux approches statistiques : le test W de Kendall et l'analyse statistique implicative. L'approche du W de Kendall nous fournit un rangement global des potentiels pédagogiques de l'usage du TNI. Elle nous montre que les enseignants apprécient plus particulièrement le TNI parce qu'il encourage la participation des élèves, attire leurs attentions et favorise les interactions entre les différents acteurs dans la salle de classe. Grâce à l'utilisation de l'analyse statistique implicative instrumentée par le logiciel CHIC, nous avons pu obtenir des informations plus analytiques sur le rangement fourni par les enseignants. Avec cette approche nous avons pu voir les similarités et les différentes relations d'implication entre les potentiels du TNI. Grâce à cette méthode nous avons aussi analysé le degré de cohésion entre les différents aspects pédagogiques du TNI.

***Mots-clés** : tableau numérique interactif, potentiels pédagogiques, test W de Kendall, analyse statistique implicative, C.H.I.C.*

ABSTRACT

The objective of this research is to study the pedagogical potential of the use of TNI in primary school from the perspective of teachers using this techno-semio- pedagogic device. We are interested, at first, to study the representations of teachers towards the use of TNI. We focus in a second time on the relationships between some pedagogical potential of the use of TNI. To treat and analyze our data obtained from the preference scale, we used two statistical approaches: Kendall's W and implicative statistical analysis. The approach of Kendall W provides us with an overall ranking of the pedagogical potential of TNI use. It shows us that teachers appreciate especially the TNI because it encourages student participation, attracts their attention and promotes interactions between different actors in the classroom. Through the use of statistical analysis implicative instrumented by the CHIC software, we were able to get more analytical information on the classification execute by teachers. With this approach we could see the similarities and the different implication relationships between the potential of the TNI. Through this method we have also analyzed the degree of cohesion between the different pedagogical aspects of TNI.

***Keywords:** interactive whiteboard, pedagogical potential, W Kendall test, implicative statistical analysis, CHIC.*

[1] Université Lumière Lyon 2, Ecole doctorale ED 485 EPIC, 86 Rue Pasteur, 69007 Lyon. Laboratoire UMR 5191 ICAR, 15, Parvis René Descartes BP 7000 69342 LYON cedex 07. hassanalcheghri@yahoo.com

1 Introduction et problématique

L'intégration d'outils numériques dans le domaine de l'éducation est en constante évolution. De nouveaux outils apparaissent continuellement et sont mis à la disposition des enseignants et de leurs élèves. Au début des années 90 le tableau numérique interactif (abrégé TNI) parfois appelé le tableau blanc interactif (abrégé TBI) a vu le jour aux États-Unis. Cependant son usage en éducation à travers le monde n'a commencé qu'à la fin des années 1990 et au début du 21ème siècle. Le TNI est devenu, dans certains pays, un outil technologique très répandu en milieu scolaire. Cet outil a introduit de réels changements dans les pratiques pédagogiques et a permis de faire évoluer la pédagogie elle-même. Le TNI, en tant que dispositif techno-pédagogique, a permis de faire évoluer la pédagogie transmissive vers des modèles constructivistes. Cette construction du savoir par le TNI est à la fois individuelle et collective. Le TNI met à la disposition de l'enseignant et des élèves une série de potentiels pédagogiques (interactivité, supports multimodaux, modalité du travail, accès à des ressources très variées, logiciels installés, etc.) qui peuvent faciliter l'enseignement et l'apprentissage. Dans notre recherche doctorale conduite sur la thématique des usages pédagogiques du TNI pour l'enseignement et l'apprentissage des mathématiques au cycle III de l'école primaire en France, les perceptions des enseignants par rapport aux potentiels du TNI dans les leçons de mathématiques ont été analysées. Nous avons demandé aux enseignants de classer par ordre d'importance croissante une série de fonctions pédagogiques du TNI. Dans cette communication nous étudierons donc les représentations des enseignants de l'école primaire envers l'usage du TNI. Nous présenterons les rangements des enseignants en faisant différents croisements entre certains potentiels pédagogiques du TNI. Concernant cette deuxième question, nous chercherons à savoir, d'une part, si le fait que le TNI encourage à la participation des élèves dans les leçons occupe une place plus importante que l'attraction de l'attention des élèves. D'autre part, si le fait que les interactions du TNI en classe est rangé à un meilleur rang que l'aspect visuel et kinesthésique de l'apprentissage. Pour analyser nos données, nous nous sommes focalisés sur deux méthodes de traitement statistique : le test W de Kendall et l'analyse statistique implicative (ASI). L'approche du W de Kendall fournit un rangement global des potentiels pédagogiques de l'usage du TNI. L'analyse statistique implicative nous a permis de faire une lecture très fine de nos données et de fournir une information plus analytique sur le phénomène en question. Avec cette méthode, nous avons pu étudier *l'arbre de similarité, le graphe implicatif et l'arbre de cohésitif* par le logiciel C.H.I.C. Ce dernier nous a apporté un moyen pour mieux comprendre nos résultats obtenus et de mieux les interpréter.

2 Le TNI : définition et principes du fonctionnement

The British Educational Communications and Technology Agency (Becta) dans un rapport publié en 2003 définit le TNI comme suit : « *un tableau blanc interactif est un grand tableau tactile, relié à un projecteur numérique et à un ordinateur. Le projecteur affiche l'image de l'écran de l'ordinateur sur le tableau. L'ordinateur peut être contrôlé en touchant le tableau, soit directement, soit avec un stylo spécial* » (Becta, 2003, p. 1, traduction libre). Cette définition nous montre que le TNI ressemble beaucoup à un système ordinateur-vidéoprojecteur utilisé avec un grand écran blanc tactile fixe ou

mobile (voir figure 1). Les éléments constituants le TNI travaillent simultanément. Le tableau tactile ou le récepteur (selon le modèle) capte toutes les manipulations effectuées sur la surface et il les transmet simultanément à l'ordinateur auquel il est connecté. L'ordinateur à son tour s'occupe des informations reçues puis il génère une autre image qu'il envoie au vidéoprojecteur. Ce dernier recrée l'image transmise par l'ordinateur en la projetant sur la surface du TNI ou sur le mur (Burton-Monney & Jauquier, 2010).

Le TNI permet à son utilisateur de faire des manipulations par deux moyens :

- En utilisant le clavier de l'ordinateur. Dans ce cas toutes les manipulations faites sur l'ordinateur sont immédiatement visibles sur le tableau numérique.
- Ou encore, en écrivant directement sur le tableau numérique à la main ou avec un stylet spécial. Toutes ces écritures sont reproduites sur l'écran de l'ordinateur.

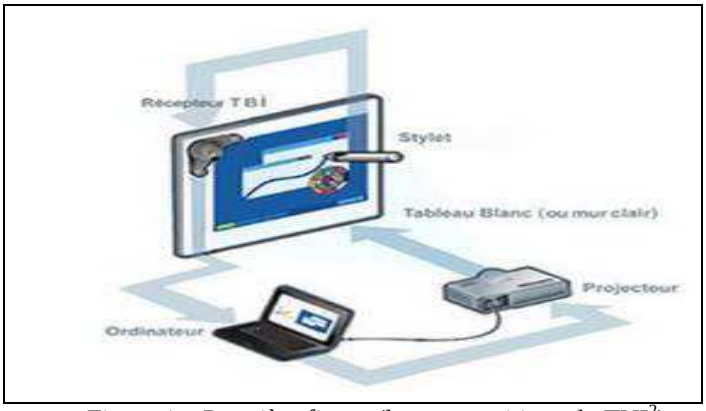


Figure 1 – Première figure (les compositions du TNI[2])

Le TNI permet effectivement de montrer des objets, des images, des textes, des enregistrements audio ou vidéo au sens le plus simple du terme, avec l'utilisation de la main (Beauchamp, 2004 ; Becta, 2003). Cet accès direct aux reproductions affichées au tableau est très intéressant pour les élèves et les enseignants parce que : « *d'une part, cela apporte une grande précision, par exemple concernant ce que l'enseignant veut effectivement montrer aux élèves : il est beaucoup plus facile pour eux de voir ce qui « au bout du doigt » de celui qui est au tableau, que de suivre le mouvement d'un curseur sur un écran (moins de fatigue visuelle pour « chercher » ce dont il parle…), d'une autre part, le TBI permet à celui qui est en action de « rester au tableau », d'oublier la présence de l'ordinateur, d'être concentré sur l'activité en cours* » (Heutte & Tempez, 2008, p. 104).

Le TNI peut être effectivement relié à des périphériques complémentaires qui peuvent aider l'utilisateur à effectuer certaines opérations nécessaires. Comme par exemple, des boîtiers de réponses des élèves, une tablette à stylet sans fil (tablette interactive) ou ardoise sans fil. Ces périphériques contribuent à augmenter l'interaction

[2] Source de l'illustration http://www.cndp.fr/crdp-besancon/index.php?id=tableau-numerique-interactif consulté le 27/01/2014.

dans la classe et à mesurer les connaissances acquises. Elles aident aussi à résoudre certains problèmes dans la classe (Dostál, 2011).

Le TNI est équipé aussi par certains logiciels assurant soit la gestion, soit l'usage pédagogique de cet outil. Ces logiciels permettent de modéliser des idées et concepts abstraits. Ils favorisent l'interaction et la discussion en classe (Becta, 2004) et créent des situations pédagogiques dynamiques, actives et entreprenantes. De même, ces logiciels « *jouent un rôle important pour transformer le TBI en un outil pédagogique performant* » (Heutte & Tempez, 2008, p. 102). Les logiciels pédagogiques sont l'un des grands avantages du TNI. Ils favorisent et « *décuplent le potentiel pédagogique que recèle cette technologie* » (Toussaint, 2010, p. 2). Les élèves dans l'étude de Wall et ses collaborateurs (2005) ont souligné que l'utilisation de différents logiciels éducatifs aide à faciliter leurs apprentissages. Les logiciels éducatifs permettent donc de changer la façon d'apprendre (Becta, 2004).

3 Un TNI pour quoi faire ? Quels sont ses attributs ?

Dans notre recherche, le TNI n'est pas un instrument isolé des autres éléments existants en classe mais il est un dispositif techno-sémio-pragmatique (Peraya, 1999). C'est dans cette acception que le TNI réunit trois grandes entités : la sémiotique, le social et la technique. De ce point de vue, le TNI est donc vu comme l'ensemble des interactions entre trois univers : technologique, système de relations (cadre technosocial) et système de représentations (de l'ordre du sémiocognitif). Cet instrument numérique occupe de plus en plus une place importante dans la sphère scolaire à travers le monde. Quelles sont donc les raisons qui donnent au TNI un tel engouement ?

1. **Interactivité** ; l'un des principaux avantages demandés au TNI comme outil d'enseignement-apprentissage, c'est l'interactivité (Becta, 2003). Selon (Bakadam & Asiri, 2012 ; Hall & Higgins, 2005) l'usage du TNI augmente le niveau d'interactivité dans la classe. L'interactivité soutenue par le TNI regroupe l'ensemble de la communauté dans la classe : enseignant-élèves et élèves-élèves. Smith et al (2005) estiment que l'interactivité enseignant-élèves est le principal avantage du TNI. Ils trouvent que l'interactivité fournie par le TNI a deux dimensions : interactivité technique et interactivité pédagogique. Le TNI favorise l'intersection entre ces deux types d'interactivité. Pour Beauchamp et Kennewell (2008), l'interactivité technique assure l'interactivité pédagogique, quand par exemple, des images, des films projetés obligent les élèves à discuter et à poser des questions.
2. **Enregistrement du travail** ; la possibilité d'enregistrer et de mémoriser les activités effectuées dans la classe semble un atout très important de l'usage du TNI. L'enseignant peut partager cette mémoire avec les élèves en l'imprimant, en l'envoyant par courriel, ou en la mettant en ligne. Cela aide les élèves à reprendre les leçons antérieures quand ils souhaitent réviser ou trouver une information. Les élèves absents, de leur côté, peuvent récupérer et apprendre les leçons manquées. En outre, l'enseignant peut feuilleter sur l'écran, devant toute la classe, ces traces pour faire une révision collective des concepts traités durant l'année. Le TNI devient donc « *la mémoire de la classe d'une heure à l'autre*

car on pourra très facilement revenir sur une trace, même lointaine […] *Il sera alors plus facile de faire des ponts entre les séances, entre les chapitres, de mettre les élèves en conditions de créer des liens entre les notions* » (Leroux, 2009, p. 51).

3. **Supports multimodaux** ; l'une des particularités propre au TNI est la possibilité de construire et d'utiliser des supports multimodaux (visuels, sonores et manipulables). Ces divers supports ont un effet considérable sur l'apprentissage des élèves.
4. **Diversité des ressources** ; le TNI permet d'accéder à des ressources très variées comme le texte, l'image fixe et animée, la vidéo, le son, etc. (Tataroglu & Erduran, 2010 ; Levy, 2002). Cette diversité permet d'enrichir les leçons et de rendre l'apprentissage plus dynamique. En outre, le TNI permet d'accéder rapidement aux ressources pédagogiques sans avoir besoin d'acheter d'autres supports. L'accès à l'Internet permet d'accéder instantanément à des ressources infinies.
5. **Visibilité** ; le grand écran du TNI est fondamental car il permet à tous les élèves (mêmes pour ceux qui sont assis au fond de la salle ou les malvoyants) de bien voir tous les écrits au tableau.
6. **Modalité du travail** ; le TNI permet de diversifier les modalités de réalisation des activités en classe. Ces modalités prennent trois formes : en grand groupe (classe entière), en petit groupe et enfin individuellement.

4 Quelques fonctions pédagogiques du TNI

Dans cette sous-section nous essayerons de mettre en lumière certains potentiels pédagogiques du TNI qui sont liés à notre questionnement. Ces bienfaits du TNI ont un effet notable sur le processus d'enseignement et d'apprentissage.

4.1 Représentation visuelle et manipulable

Le TNI permet de créer des environnements d'apprentissage où diverses représentations dynamiques peuvent être utilisées (visuelles, kinesthésique, sonores). Autrement dit, le TNI permet de proposer des représentations diverses selon l'objet d'étude. Un élève dans l'étude de Wall et ses collaborateurs (2005) a indiqué cette capacité du TNI en montrant qu' « *il aide parce que vous pouvez voir les choses, entendre et faire bouger les choses autour du tableau* » (Ibid., page 862, traduction libre). Cet instrument permet également de faire des combinaisons entre ces supports multimodaux. L'enseignant peut, lors d'une séance de découverte du monde, joindre la représentation externe audio à la représentation externe visuelle, de telle façon que, lorsqu'un élève touche la carte de monde, il entend le phonème prononcé (continent, pays, ville…). Dans ce cas, ces deux représentations (visuelle et sonore) forment un seul élément virtuel que la personne peut exploiter via deux canaux sensoriels (Magnat, 2013). En outre, le TNI permet également de manipuler physiquement les éléments de l'objet étudié. Ces représentations kinesthésiques peuvent être effectuées par

l'enseignant ou par ses élèves. Dans ce cas, l'utilisation du TNI offre l'occasion de créer un environnement d'apprentissage intéressant pour les élèves. Dans cet environnement médiatisé par le TNI, l'élève devient un acteur actif et il s'engage dans toutes les procédures de l'apprentissage. « *Les TBI permettent aux enseignants de créer des séances dynamiques où chaque leçon devient une expérience d'apprentissage sur mesure. S'il est utilisé de la bonne façon, l'apprentissage peut devenir plus interactif, amusant et attirant, ce qui permet d'augmenter l'engagement et la mémorisation du contenu* » (McIntyre-Brown, 2011, p. 3, traduction libre). Pour Miller et son collaborateur (2005a, 2005b) les représentations multimodales du TNI ont un effet sur la motivation et l'attention des élèves. La combinaison entre ses supports constitue aussi un élément qui peut encourager les élèves à apprendre.

4.2 Interaction

Les potentiels pédagogiques du TNI ont permis d'introduire des changements sur les interactions en classe. Autrement dit, ce dispositif numérique augmente à la fois l'interaction enseignant-élèves, et, élèves-élèves (Lewin, Somekh, & Steadman, 2008). Becta (2003) souligne que le TNI favorise l'interaction et la discussion en classe. L'interaction élèves-élèves est plus élevée chez les élèves qui apprennent en utilisant le TNI que chez ceux qui apprennent en suivant les méthodes usuelles (Zittle, 2004). Le TNI favorise l'interaction enseignant-élèves en raison de différentes façons de présenter les informations par l'enseignant (Lefebvre & Samson, 2013 ; Glover, Miller, & Averis, 2003). En outre, le niveau d'interaction entre l'enseignant et les élèves est également plus élevé quand le TNI est utilisé dans la classe (Smith, Hardman, & Higgins, 2006). Cet instrument conduit donc à favoriser l'aspect social dans le processus d'enseignement-apprentissage. Il permet de produire ensemble, d'échanger et de partager (Toussaint, 2010) les connaissances avec les autres. Il permet aussi d'apprendre en faisant des erreurs « ensemble ». Dans l'étude de Levy (2002), les élèves ont souligné que la discussion et le partage du travail avec les autres leur ont permis d'examiner des idées et de donner des bonnes explications. La disponibilité du TNI en soi-même dans la salle de classe n'est pas suffisante pour assurer une amélioration considérable dans les interactions. La manière d'utiliser le TNI par l'enseignant est le facteur clé pour favoriser cet aspect. L'enseignant qui utilise le TNI comme un outil complémentaire dans la classe plutôt que comme un outil intégré dans les pratiques pédagogiques réduit l'interaction prévue par l'utilisation du TNI (Greiffenhagen, 2000). Pour cela, l'enseignant doit apprendre à arbitrer et à gérer les nombreux axes d'interactions en classe (Lewin, Somekh, & Steadman, 2008). En l'absence d'interactivité le TNI devient seulement un outil pour présenter les informations (Beauchamp & Kennewell, 2010). Il devient un simple outil visuel (Kennewell & Beauchamp, 2007).

4.3 Estime de soi

L'usage du TNI aide au développement positif du concept de soi (Jeunier et al, 2005). Les chercheurs ont conclu que les élèves, dans les classes équipées du TNI, demandent très fréquemment de passer au tableau même s'ils refusent habituellement de le faire. « *Six des huit garçons qui ne lèvent pas le doigt d'habitude ont levé le doigt au moins une fois, et deux d'entre eux l'ont fait de façon spontanée, sans que l'enseignant*

propose une activité ou sollicite quoi que ce soit. Le premier de ces deux élèves a souhaité mettre en marche le tableau » (Ibid., p.26). Le TNI parait le point focal lorsqu'il est utilisé, donc l'élève « *se sent moins visé d'une part, et trouve, avec cet outil, une forme de contenance d'autre part* » (Ibid.).

4.4 Motivation et attention

L'usage du TNI augmente la motivation, l'attention, le plaisir et l'intérêt à l'apprentissage chez les élèves (Beauchamp & Kennewell, 2008 ; Hall & Higgins, 2005). Il accroît aussi leur engagement, leur participation et leur concentration (Kennewell & Beauchamp, 2007 ; Glover & Miller, 2001 ; Wall, Higgins, & Smith, 2005). Tataroglu & Erduran (2010) ont conclu que l'intérêt des élèves (surtout ceux qui avaient un faible niveau en mathématiques) à l'apprentissage des mathématiques a été amélioré après avoir utilisé le TNI pendant 5 semaines. De leur côté, Lewin et ses collaborateurs (2008), ont constaté que les élèves ayant des besoins éducatifs spéciaux sont très motivés par la possibilité de démontrer leurs compétences et leurs connaissances avec le TNI. Torff et Tirotta (2010) ont conclu que les élèves du groupe expérimental qui ont appris les mathématiques avec le TNI ont signalé des niveaux plus élevés de motivation pour apprendre les mathématiques par rapport à ceux du groupe témoin qui ont appris les mathématiques avec les méthodes classiques. Ils ont trouvé qu'il y a un rapport –même s'il est faible - entre le degré d'appropriation des enseignants à l'utilisation du TNI et le niveau de la motivation chez leurs élèves. Sur ce point, ils ont indiqué que les élèves des enseignants qui étaient favorables au TNI ont eu des niveaux de motivation plus élevés que les élèves des enseignants qui étaient moins favorables au TNI. Pour leur part, Miller, Glover, & Averis (2005b) ont démontré dans leur observation lors des leçons animées par le TNI qu'il y avait un intérêt et un plaisir chez les élèves envers les mathématiques. Pourtant les chercheurs signalent que le TNI en soi même n'est pas suffisant pour s'assurer que les élèves sont motivés. C'est les pratiques pédagogiques et la qualité de l'enseignement qui peuvent renforcer la motivation des élèves. Pour Bakadam & Asiri (2012) le TNI est un outil pédagogique efficace qui a le pouvoir d'engager et de motiver les élèves dans le processus d'apprentissage.

5 Perceptions des enseignants envers l'utilisation et la contribution du TNI

Selon nos lectures, nous avons trouvé que le TNI est généralement perçu de manière (très) positive par les enseignants. Ces derniers ont un avis favorable à l'usage du TNI dans la salle de classe. Ils pensent que le TNI s'intègre aisément dans les courantes pratiques pédagogiques, et que son utilisation contribue à améliorer l'enseignement et l'apprentissage. Dans cette section nous montrerons quelques résultats des recherches touchant la question suivante : comment les enseignants de l'école primaire perçoivent l'utilité de l'usage du TNI dans la salle de classe. Pour ce faire, nous montrerons dans un premier temps les représentations des enseignants envers l'utilisation du TNI en général et dans un deuxième temps nous mettrons en lumière les impressions des

enseignants sur la contribution de cet outil notamment dans le domaine des mathématiques.

Les enseignants sont enthousiastes à utiliser le TNI (Glover & Miller, 2001 ; Slay, Sieborger, & Hodgkinson-Williams, 2008) et ils pensent que le TNI favorise et soutient leur enseignement. Dans leurs recherches (Türel & Johnson, 2012 ; Lewin, Somekh, & Steadman, 2008) les enseignants estiment qu'ils gagnent du temps en utilisant le TNI. Par contre, au début, il leur faut plus de temps pour préparer les cours animés par le TNI. Le temps investi par les enseignants à la préparation des cours devrait être réduit en raison de la possibilité de sauvegarder, partager et réutiliser les travaux réalisés en classe (Bakadam & Asiri, 2012 ; Somekh, et al., 2007 ; Lefebvre & Samson, 2013). Les enseignants de l'étude de Türel & Johnson (2012) ont également confirmé que le TNI améliore beaucoup la présentation des cours par la visualisation de leurs contenus sur l'écran. Ce qui permet d'attirer l'attention des élèves vers les notions « de jeux ». Ainsi il facilite la compréhension des élèves des concepts abordés, principalement les concepts complexes et abstraits. Dans ce contexte Leroux (2009) souligne que « *le tableau numérique interactif (TNI) peut nous séduire par la facilité avec laquelle il permet de rendre dynamique une présentation magistrale* » (Ibid., p. 45). En outre, Türel & Johnson (2012), dans leur recherche, ont conclu que les enseignants estiment que le TNI peut faciliter l'apprentissage des élèves sous certaines conditions. Ils ont souligné trois conditions principales : « *la collaboration avec les collègues, la formation sur les stratégies pédagogiques efficaces s'agissant de l'utilisation du TNI, et, l'utilisation régulière du TNI par les enseignants pour améliorer leurs compétences* » (Ibid., page 381, traduction libre).

Dans l'étude de Bakadam & Asiri (2012), les enseignants trouvent que le TNI constitue un moyen efficace et pratique pour transmettre le contenu d'apprentissage et qu'il augmente le niveau d'interaction en classe. Ce niveau élevé de l'interaction augmente l'expérience d'apprentissage. Le TNI est un vecteur visuel et dynamique du point de vue des enseignants et il est le meilleur support à la réflexion par rapport aux autres outils traditionnels (Beauchamp & Kennewell, 2008). Dans l'étude de Slay et ses collaborateurs (2008) les enseignants ont mis en évidence l'efficacité, la flexibilité, la polyvalence du TNI et la possibilité d'accéder à des ressources multimédias (vidéo, audio…) ainsi que la possibilité de gérer la classe plus facilement.

Lewin, Somekh, & Steadman (2008) ont réalisé une recherche mixte (quantitative et qualitative) dans 100 classes des écoles primaires réparties au niveau national en Angleterre. Les chercheurs ont conclu que, lorsque les enseignants ont enseigné en utilisant le TNI pendant deux ans, ils ont montré que les élèves avaient fait des progrès exceptionnels dans leurs résultats en mathématiques, en anglais et en sciences. Dans son étude Zittle (Zittle, 2004) a travaillé avec deux groupes d'élèves (avec et sans TNI) pour l'apprentissage des mathématiques. Pour ces deux groupes, il a appliqué un pré-test et un post-test pour mesurer les connaissances apprises par les élèves. Il a conclu que l'amélioration du niveau des élèves en mathématiques dans le post-test est plus grande chez le groupe du TNI que chez le groupe sans TNI. A cet égard, López (2010), a confirmé que les résultats des élèves, en mathématiques et en lecture, dans les classes animées par le TNI ont été augmentés significativement par rapport aux résultats des élèves dans les classes traditionnelles. L'auteur indique que les élèves dans les classes qui utilisent le TNI ont dépassé en mathématiques leurs camarades des classes

traditionnelles. Pour leur part, Swan et ses collaborateurs (2008) ont conclu que la comparaison entre les résultats du groupe ayant appris les mathématiques en utilisant le TNI et ceux n'ayant pas utilisé le TNI, montrent qu'il y a une différence légèrement favorable pour ceux qui ont appris avec le TNI. Pour d'autres chercheurs le TNI a un effet indirect sur les résultats et l'apprentissage des élèves : le TNI améliore les pratiques pédagogiques, les modalités de l'enseignement et de l'apprentissage, l'interaction dans la classe, etc. Cela conduit évidement à améliorer les résultats et les performances des élèves dans les domaines enseignés. « *Les tableaux blancs interactifs ont un grand potentiel comme outil pour améliorer les pratiques pédagogiques en classe et, finalement, améliorent les résultats des élèves* » (Marzano, 2009, p. 82, traduction libre).

6 Méthodologie

Dans cette section nous aborderons les questions méthodologiques de notre recherche qui a comme objectif de s'interroger sur les préférences des enseignants dans l'usage du TNI. Cet objectif nous amène évidemment à nous interroger sur certains croisements possibles entre les items de notre échelle pour bien comprendre et interpréter les relations mutuelles entre les fonctions pédagogiques de ce dispositif mis en œuvre. Nous présenterons nos hypothèses de recherche, nos outils méthodologiques utilisés pour le recueil des données et nous montrerons enfin les variables indépendantes et l'échantillon sur lequel nous avons travaillé.

6.1 Hypothèses

Pour répondre à notre problématique et aux questions de notre recherche nous avons formulé les hypothèses de recherche suivantes :

- Hypothèse 1 ; les enseignants de l'école primaire ont des points de vue communs en ce qui concerne les potentiels pédagogiques de l'usage du tableau numérique interactif dans la salle de classe.
- Hypothèse 2 ; le fait que le tableau numérique interactif encourage la participation des élèves dans les leçons est rangé à un meilleur rang plutôt qu'il n'attire l'attention des élèves.
- Hypothèse 3 ; le fait que le tableau numérique interactif encourage les interactions en classe est rangé à un meilleur rang plutôt qu'il ne favorise l'aspect visuel et kinesthésique de l'apprentissage.

Une lecture de la littérature du TNI nous a permis d'aboutir à deux grandes conclusions. La première est que beaucoup de recherches mettent l'accent sur le fait que ce dispositif attire l'attention, la concentration des élèves et qu'il favorise l'aspect visuel chez les élèves. La deuxième conclusion est la rareté des recherches qui mettent en lumière la contribution du TNI dans l'encouragement à la participation des élèves dans les leçons ainsi qu'aux interactions dans la classe. Autrement dit, les recherches mettent plutôt l'accent sur l'aspect où l'élève est un spectateur et ignorent les aspects où l'élève est acteur. Les recherches mettent aussi en évidence les contributions pédagogiques du TNI

à court terme plutôt que ses contributions permanentes, sachant que du point de vue pédagogique on cherche à mettre l'élève en action pour favoriser ses apprentissages, surtout si on sait que l'enthousiasme et la passion des élèves vers cet outil semblent s'amoindrir avec le temps (Dostál, 2011). Dans les hypothèses 2 et 3 nous essayons de voir les perceptions des enseignants envers les potentiels pédagogiques les plus importants de leurs points de vue. Donnent-ils une priorité aux potentiels pédagogiques permettant aux élèves de participer dans les activités en classes ou plutôt mettent-ils l'accent sur l'aspect affectif de l'usage du TNI en classe ?

6.2 Instruments de recueil de données

Dans le cadre de notre recherche, il nous a fallu construire des données qui nous ont permis d'étayer les réponses hypothétiques que nous avons formulées à l'égard des questions que nous nous sommes posées. Nous avons eu besoin de connaître les appréciations des enseignants envers l'utilisation du TNI, les résultats, les supports multimodaux utilisés et ses effets sur l'apprentissage des élèves, etc. Nous avons tout d'abord réalisé une observation directe dans deux classes de l'école primaire (CM1 et CM2) pendant l'enseignement-apprentissage des leçons sur les fractions durant l'année scolaire 2014-2015. Nous avons dans un deuxième temps construit une échelle pour évaluer les potentiels pédagogiques du TNI du point de vue des enseignants. Cette enquête a été complétée par une enquête à l'aide d'entretiens semi-directifs auprès d'élèves et d'enseignants. Le terrain d'observation et d'entretien est une école primaire publique de la région lyonnaise. Dans l'étude de l'échelle de préférence c'est l'ensemble des écoles primaires publiques de la région lyonnaise qui ont été interrogées. Nous décrivons très brièvement les outils de construction de données utilisés dans cette recherche à la fois quantitative et qualitative. Il est important de noter que les résultats mentionnés ci-dessous sont uniquement des résultats issus de notre échelle de préférence construite et adressée aux enseignants du primaire.

6.2.1 Construction des données par observations directes

Nous avons réalisé une observation directe des élèves en situation d'apprentissage. Cette observation s'est faite en deux groupes. Un macro-groupe où nous avons observé la classe entière puis nous avons concentré notre observation sur un groupe de 4 élèves (2 filles et 2 garçons) qui travaillaient et apprenaient ensemble. Leur niveau scolaire en mathématiques était aussi hétérogène : un élève de bon niveau, deux élèves de niveau moyen et un élève de faible niveau.

6.2.2 Construction des données par échelle de préférence

Nos lectures des recherches conduites sur l'exploitation du TNI en mathématiques, nous ont permis d'identifier certains potentiels pédagogiques de l'usage du TNI dans les leçons des mathématiques. Selon nos observations sur le terrain nous avons également pu remarquer d'autres fonctions pédagogiques de cet outil. Nous avons regroupé l'ensemble des potentiels pédagogiques du TNI dans une échelle de préférence (voir tableau 1) en excluant les potentiels identiques.

Enoncés des objectifs	Rang
V40A. Encourager la participation des élèves	
V40B. Promouvoir l'interaction élèves-élèves	
V40C. Aider à comprendre des notions complexes	
V40D. Concrétiser les concepts abstraits	
V40E. Transmettre plus facilement le contenu des cours	
V40F. Varier les ressources pédagogiques utilisées en classe	
V40G. Favoriser l'aspect visuel et kinesthésique de l'apprentissage	
V40H. Attirer l'attention des élèves	
V40I. Susciter la motivation des élèves pour apprendre	
V40J. Aider à représenter facilement les relations fractionnaires	
V40K. Encourager l'interaction enseignant-élèves	
V40L. Diversifier les méthodes d'enseignement-apprentissage	

Tableau 1 - Premier tableau (tableau des q=12 items à ranger par ordre d'importance croissante)

Nous avons proposé ensuite à l'ensemble de notre échantillon une échelle comprenant 12 items qu'il fallait classer de 1 à 12 afin de mesurer leurs appréciations envers l'usage du TNI en classe. Ces douze items se regroupaient par potentiels pédagogiques du TNI sur l'aspect individuel et collectif (voir tableau 1). Ils avaient pour consigne de classer les 12 éléments par ordre d'importance croissante. Après avoir recueilli les données, nous les avons analysées dans un premier temps selon le test W de Kendall pour savoir s'il existait une tendance à un choix préférentiel dans l'échantillon de notre étude. Nous nous sommes tournés dans un deuxième temps vers une analyse des données focalisée sur l'approche d'analyse statistique implicative. Cette approche est développée par Régis Gras et ses collaborateurs (Gras, Régnier, & Guillet, 2009 ; Gras, Régnier, Marinicia, & Guillet, 2013) et instrumentée par le logiciel C.H.I.C.

6.2.3 Construction des données par enquête à l'aide d'entretiens semi-directifs

Pour tenter de mieux comprendre le sens des réponses issues d'échelle et d'observations directes, nous avons réalisé des entretiens semi-directifs individuels : enseignants des classes observées et élèves de ces deux classes. Avec les élèves[3] nous avons effectué dix entretiens. L'entretien avec les enseignants comportait plusieurs groupes de questions : profil personnel et professionnel de l'enseignant, différents aspects de l'usage pédagogique du TNI, etc. Avec les élèves, nous avons procédé en deux parties. Une première partie concernait leurs perceptions face à l'usage du TNI en interrogeant les élèves sur les différentes fonctions pédagogiques du TNI. Dans la deuxième partie, lors d'un entretien-feuilletage nous avons demandé aux élèves de feuilleter leurs cahiers de mathématiques pour tenter d'établir une corrélation entre l'usage du TNI et le cahier des élèves.

6.3 Variables et échantillon

[3] Nous avons sollicité l'accord parental pour interviewer les enfants.

L'échantillon de cette étude a été construit à l'aide de 5 variables totalisant 13 variables binaires.

1. Sexe (homme, femme)
2. Âge (âge 1, âge 2, âge 3)
3. Ancienneté à l'utilisation du TNI (ancienneté 1, ancienneté 2, ancienneté 3)
4. Niveau de classe enseignée (CM1, CM2, CM1-CM2)
5. Formation sur le TNI (formation_TNI_Oui, formation_TNI_Non)

S'agissant des variables d'âge et d'ancienneté, nous avons posé ces deux questions sous forme de questions ouvertes. Nous avons ensuite catégorisé les réponses en variables intervalles. Concernant l'ancienneté, nous l'avons regroupé en 3 variables : ancienneté 1 = de 0,5 à 1,41 ans, ancienneté 2 = de 2 à 3 ans, ancienneté 3 = de 4 à 6 ans. Nous avons aussi regroupé l'âge en 3 variables : âge 1 = de 30 à 37 ans, âge 2 = de 39 à 46 ans, âge 3 = de 47 à 59 ans.

Nous pouvons préciser que durant notre recherche, nous avons adressé un questionnaire de 45 questions aux enseignants de l'école primaire utilisant le TNI pour animer leurs cours de mathématiques. Nous avons crée deux versions de ce questionnaire : électronique et papier. Les enseignants ont été donc soumis à un questionnaire dont une question consistait à ranger les 12 apports du TNI selon un ordre croissant d'importance. Nous avons reçu 67 réponses. De cet ensemble, nous avons obtenu un sous-échantillon de 48 sujets qui ont répondu à la version électronique de notre questionnaire. Nous avons également 19 personnes qui ont répondu à la version papier mais nous n'avons pas pris en compte faute de temps cette catégorie dans cette communication.

En fonction de nos variables prises en compte, les 48 sujets se répartissent comme suit :

	Âge			Ancienneté			Classe			Formation	
	Âge 1	Âge 2	Âge 3	Ancienneté 1	Ancienneté 2	Ancienneté 3	CM1	CM2	CM1-CM2	OUI	NON
Homme	5	10	5	5	8	6	8	8	4	17	3
Femme	16	5	7	4	21	4	13	13	2	24	4
	21	15	12	9	29	10	21	21	6	41	7

Tableau 2 - Deuxième tableau (répartition de l'échantillon selon nos variables)

7 Résultats et discussion

Dans cette section nous présenterons plus exclusivement nos résultats obtenus à partir de l'échelle de préférence pour nous aider à déterminer le degré d'appréciation des potentialités pédagogiques du TNI chez les enseignants. Nous présenterons nos résultats en se focalisant sur deux approches de traitement des données : le test W de Kendall et l'analyse statistique implicative. Ces deux méthodes statistiques nous ont permis d'analyser nos données obtenues par l'échelle d'évaluation pour l'ensemble de notre échantillon. Le fait de recourir à deux approches statistiques nous a permis d'approfondir nos analyses et interprétations et donc d'avoir une lecture plus fine de nos

données. L'analyse statistique implicative nous permet de « *fournir une information plus analytique sur la structure de quai-ordre affectée d'un niveau de confiance 1- α, qui organise l'ensemble des q(q-1) paires d'objets, que ne le permet l'approche par la procédure du W de Kendall qui fournit un rangement global synthétique* » (Régnier & Gras, 2013, p. 193). L'idée de notre recherche était de voir s'il y avait un accord entre les enseignants de l'école primaire s'agissant des potentiels pédagogiques et de comprendre certaines relations entre ces potentialités. Nous constatons dans le tableau 1, qu'il y a douze items de l'échelle de l'appréciation des fonctions pédagogiques allant de l'item 1 à l'item 12 (de Q40A à V40L). Le tableau 1 nous montre également que les sujets de notre échantillon étaient invités à attribuer un rang de préférence entre 1 et 12 pour chaque item. Il est important de souligner que les variables de rangs dans notre recherche se situent dans le cas du rangement complet et avec *ex aequo*.

7.1 Le recours au test W de Kendall

Le calcul des sommes des objets rangés (de 1 à 12) selon le test W de Kendall nous donne les résultats suivants :

Objectifs	O1 V40A	O2 V40B	O3 V40C	O4 V40D	O5 V40E	O6 V40F	O7 V40G	O8 V40H	O9 V40I	O10 V40J	O11 V40K	O12 V40L
Sommes	183	216	337,5	340,5	426,5	404,5	277	187,5	192	407	258	514,5
Rangs	1	4	7	8	11	9	6	2	3	10	5	12

Tableau 3 - Troisième tableau (tableau des 12 items rangés par ordre d'importance croissante selon l'approche W de Kendall

La mise en œuvre de la procédure du test W de Kendall fournit les informations suivantes. La réalisation empirique de la statistique W vaut w = 0.4038 alors que la valeur critique au seuil de $\alpha = 0.01$ vaut $w_c = 0.0468$ et au seuil de $\alpha = 0.05$ vaut $w_c = 0.0372$. Selon le principe du test W de Kendall on rejette dans les deux niveaux de risque l'hypothèse H0 d'indépendance ou d'hétérogénéité des rangements fournis par les enseignants de l'école primaire. C'est-à-dire, qu'il y a certains accords entre les sujets sur le rangement.

Rang	1	2	3	4	5	6	7	8	9	10	11	12
Global	V40A	V40H	V40I	V40B	V40K	V40G	V40C	V40D	V40F	V40J	V40E	V40L
Homme	V40A	V40I	V40B	V40H	V40K	V40G	V40C	V40D	V40E	V40F	V40J	V40L
Femme	V40H	V40A	V40I	V40B	V40G	V40K	V40D	V40C	V40J	V40F	V40E	V40L
CM1	V40A	V40B	V40I	V40H	V40K	V40G	V40D	V40C	V40J	V40F	V40E	V40L
CM2	V40H	V40I	V40A	V40B	V40G	V40K	V40D	V40C	V40F	V40J	V40E	V40L
CM1_CM2	V40I	V40H	V40A	V40G	V40C	V40B	V40K	V40F	V40E	V40D	V40J	V40L
Formation_oui	V40I	V40A	V40H	V40B	V40K	V40G	V40C	V40D	V40F	V40J	V40E	V40L
Formation_non	V40H	V40A	V40B	V40K	V40G	V40I	V40J	V40D	V40C	V40F	V40E	V40L
Âge 1	V40B	V40A	V40I	V40H	V40K	V40G	V40D	V40C	V40F	V40J	V40E	V40L
Âge 2	V40H	V40I	V40A	V40G	V40B	V40K	V40C	V40D	V40E	V40F	V40J	V40L
Âge 3	V40A	V40H	V40B	V40I	V40K	V40G	V40J	V40C	V40D	V40F	V40E	V40L
Ancienneté1	V40I	V40H	V40A	V40B	V40G	V40C	V40D	V40K	V40E	V40F	V40J	V40L
Ancienneté 2	V40A	V40B	V40H	V40I	V40K	V40G	V40D	V40C	V40F	V40J	V40E	V40L
Ancienneté 3	V40H	V40I	V40A	V40G	V40D	V40C	V40K	V40B	V40J	V40F	V40E	V40L

Tableau 4 - Quatrième tableau (les rangements des items au sein de différents groupes de l'échantillon)

Pour ce qui concerne le tableau 4, nous trouvons en jaune le rang des items du point de vue de l'ensemble de notre échantillon. Les autres couleurs nous montrent le classement des items du point de vue de chaque groupe de notre échantillon (en rouge selon le sexe et ainsi de suite). Le tableau 5, quant à lui, nous fournit les informations suivantes : W correspond au W de Kendall, P au nombre d'individus (échantillon ou sous-échantillon), w_c à la valeur critique et enfin 0.05 et 0.01 correspondent au niveau de risque.

La lecture du tableau 4 selon l'approche du W de Kendall nous permet de fournir la conclusion suivante : plus nous nous approchons de 12, plus la concordance est forte. Nous traduisons cette conclusion par des exemples concrets du tableau 5. Dans ce cas, nous voyons (en vert pâle) que la concordance est plus forte chez les hommes (0.5741) et moins forte (en gris) chez les maîtres enseignants dans une classe mixte, soit CM1-CM2 (0.2934). Nous précisons que dans les rangs numérotés 1, 2, 3, 7 et 8, il y a 4 items qui se répètent chez les groupes de l'échantillon. Alors que pour les rangs numérotés 4, 5, 6 et 8, il y a 5 items qui se répètent. Concernant les rangs numérotés 10, 11, 12, il y a respectivement 3, 2, 1 items qui se répètent. Autrement dit, à partir du rang numéro 8, les nombres d'items décroissent de 5 à 1 alors que dans les sept premiers rangs les nombres d'items fluctuent entre 4 et 5.

	p=	W=	wc : α = 0,05	wc : α = 0,01
Global	48	0,4038	0,0372	0,0468
Homme	20	0,5741	0,0894	0,1123
Femme	28	0,3171	0,0638	0,0802
CM1	21	0,5498	0,0851	0,1070
CM2	21	0,3495	0,0851	0,1070

CM1_CM2	6	0,2934	0,2981	0,3746
Formation_oui	41	0,4214	0,0436	0,0548
Formation_non	7	0,346	0,2555	0,3211
Âge 1	21	0,4442	0,0851	0,1070
Âge 2	15	0,4032	0,1192	0,1498
Âge 3	12	0,4115	0,1490	0,1873
Ancienneté1	9	0,4995	0,1987	0,2497
Ancienneté2	29	0,4430	0,0616	0,0775
Ancienneté3	10	0,3787	0,1788	0,2247

Tableau 5 - Cinquième tableau (valeurs empiriques et critiques au seil de α = 0.01 et de α = 0.05

En termes d'importance comme le montre le tableau 4, la majorité absolue de l'échantillon s'accorde pour placer V40L "*diversifier les méthodes d'enseignement-apprentissage*" en dernière position, V40E "*transmettre plus facilement le contenu des cours*" en avant dernière position et en neuvième position V40F "*varier les ressources pédagogiques utilisées en classe*". V40A ("*encourager la participation des élèves*") et V40I ("*susciter la motivation des élèves pour apprendre*") occupent toujours les premiers rangs. V40A balance entre la première et troisième position tandis que V40I se trouve entre la première et quatrième position. V40G ("*favoriser l'aspect visuel et kinesthésique de l'apprentissage*") occupe toujours les trois rangs suivants quatrième, cinquième et sixième position. Alors que V40K "*encourager l'interaction enseignant-élèves*" se trouve au milieu des rangs entre quatrième et huitième position (plus généralement en cinquième position). V40B ("*promouvoir l'interaction élèves-élèves*") et V40H ("*attirer l'attention des élèves*") se positionnent dans les premiers rangs (en général entre première et quatrième position). V40C "*aider à comprendre des notions complexes*" et V40D "*concrétiser les concepts abstraits*" occupent les rangs du milieu (septième ou huitième position). Alors que V40J se situe dans les rangs suivants (entre neuvième et onzième position).

7.2 Le recours à l'analyse statistique implicative (ASI)

L'analyse statistique implicative nous a permis d'avoir des informations plus riches avec plus de détails sur les rangs et les liens implicatifs entre les items rangés. Pour traiter nos variables de rangs selon l'approche d'ASI, nous avons transformé les variables binaires en variables modales à valeur sur [0 ; 1] comme le suggère Régnier et Gras (2005, 2013). Avec cette méthode, nous avons pu étudier *l'arbre de similarité, le graphe implicatif et l'arbre cohésitif* comme nous les présenterons ci-après. Il est important de souligner que nous avons analysé nos données selon l'approche d'ASI en se focalisant sur l'approche de l'analyse de la propension statistique de J-B. Lagrange (Lagrange, 1998). Celui-ci étudie la relation de propension aux variables de rangs avec des valeurs modales de [0 ; 1]. Nous avons pensé au début de nous tourner vers l'approche "permutationnelle" de la propension de Régnier mais compte-tenu du manque de temps nous nous sommes focalisé sur celle de Lagrange. L'approche de Régnier nécessite des procédures plus fines pour calculer l'intensité de propension. Régnier et Gras (2013) mettent en lumière la différence entre la relation de propension de Lagrange et l'approche permutationnelle en disant que « *la seule différence réside dans le fait que le calcul de l'espérance et de la variance statistique Z utilisée par J.B.*

Lagrange tient compte de la particularité de l'information fondée sur statistique de rangs » (Idid., p. 201).

Dans la suite nous présentons les résultats issus grâce à l'approche d'ASI. Nous montrons dans un premier temps les résultats de l'arbre de similarité, puis ceux du graphe implicatif et enfin nous mettrons en lumière les résultats de l'arbre cohésitif.

7.2.1 L'arbre des similitudes : étude des similarités entre les variables

L'arbre de similarité nous a permis, d'une part, de voir s'il y a des similarités entre nos variables. D'autre part, d'identifier le niveau de similarité entre les couples de variables. L'arbre des similarités calcule pour chaque couple de variables la similarité entre celles-ci. Elle regroupe aussi les variables similaires par des groupes de variables qui se ressemblent le plus. De même elle classe la similarité entre les variables du niveau le plus élevé vers le moins élevé.

Dans notre exemple, nous avons trois groupes de similarité qui se détachent (voir figure 2). Horizontalement nos variables se répartissent en 11 niveaux de similarité (voir figure 2 et tableau 6). Le premier composé de V40A, V40G, V40B et V40K. Le deuxième de V40C, V40D, V40E et V40F. Le troisième de V40H, V40I, V40J et V40L. Pour bien comprendre la similarité entre ces variables nous précisons que les lignes rouges qui relient les variables signifient que la similarité est très forte et signifiante entre les variables. En outre, plus les lignes du graphique descendent vers le bas, moins il y a de similarité entre les variables. A ce stade, nous pouvons dire que V40J et V40L sont plus similaires que V40H et V40I. Nous observons également que le deuxième et le troisième groupe ont des similarités très fortes, mais cela se fait à un niveau très bas. Ces deux groupes se ressemblent au niveau 11 avec le premier groupe. Les trois groupes constituent à la fin un seul bloc, ce qui signifie qu'il y a une similarité entre ces groupes.

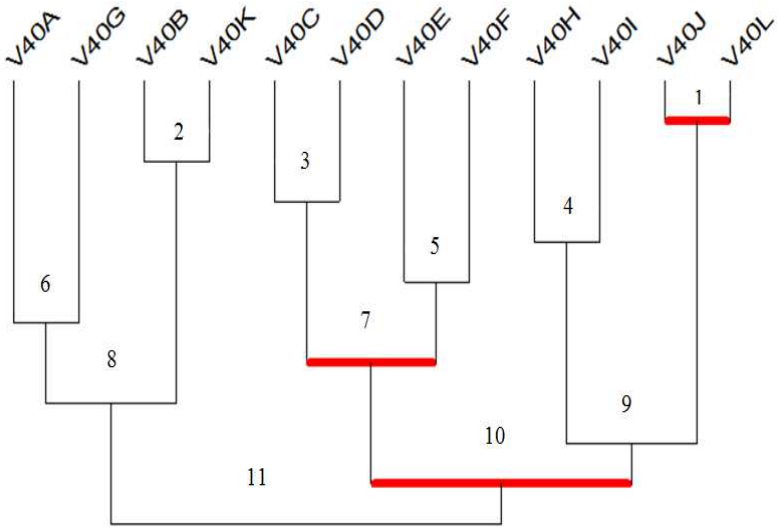


Figure 2 - Deuxième figure (arbre de similarité organisant les 12 potentiels pédagogique du TNI en classes)

Notre arbre de similarité montre que le nœud le plus significatif est au niveau 1. C'est-à-dire, entre V40J "*aider à représenter facilement les relations fractionnaires*" et V40L "*diversifier les méthodes d'enseignement-apprentissage*". Il y a d'autres nœuds significatifs : au niveau 7 et au niveau 10. Nous remarquons que V40B "*promouvoir*

l'interaction élèves-élèves" et V40K "*encourager l'interaction enseignant-élèves*" sont très similaires. V40C "*aider à comprendre des notions complexes*" et V40D "*concrétiser les concepts abstraits*" ont également un niveau fort de similarité. Une similarité assez forte existe aussi entre V40H "*attirer l'attention des élèves*" et V40I "*susciter la motivation des élèves pour apprendre*". Comme nous l'avons indiqué ci-dessus, plus nous descendons vers le bas du graphique, moins la similarité est identifiée entre les potentiels pédagogiques du TNI.

Niveau de classification	Variables	Similarité
Classification au niveau 1	(V40J V40L)	0.725953
Classification au niveau 2	(V40B V40K)	0.677835
Classification au niveau 3	(V40C V40D)	0.653313
Classification au niveau 4	(V40H V40I)	0.613723
Classification au niveau 5	(V40E V40F)	0.564988
Classification au niveau 6	(V40A V40G)	0.427426
Classification au niveau 7	((V40C V40D) (V40E V40F))	0.171002
Classification au niveau 8	((V40A V40G) (V40B V40K))	0.142545
Classification au niveau 9	((V40H V40I) (V40J V40L))	0.0447803
Classification au niveau 10	(((V40C V40D) (V40E V40F)) ((V40H V40I) (V40J V40L)))	0.000427605
Classification au niveau 11	(((V40A V40G) (V40B V40K)) (((V40C V40D) (V40E V40F)) ((V40H V40I) (V40J V40L))))	2.9069e-08

Tableau 6 - Sixième tableau (niveaux de similarité entre nos variables)

7.2.2 Le graphe implicatif : réseaux des variables

Quant au graphe implicatif, il nous montre les tendances à une intensité d'implication supérieure à un certain seuil entre les différentes variables (voir figure 3). À l'origine, nous avons fixé l'indice d'implication à un niveau de confiance 0.90. Mais alors aucune variable n'apparaissait. Nous avons alors décidé de l'abaisser à 0.80, mais seulement deux variables apparaissaient, ce qui n'apportait que peu d'éléments pour développer notre interprétation dans le champ de l'usage du TNI. Nous avons décidé à nouveau d'abaisser le niveau de confiance jusqu'au niveau (0.62) où on voit alors toutes nos variables principales. Les résultats ont été plus explicites : toutes les variables apparaissent en deux groupes distincts.

Le premier groupe réunit 5 variables : ("*attirer l'attention des élèves*", "*encourager la participation des élèves*", "*susciter la motivation des élèves pour apprendre*", "*promouvoir l'interaction élèves-élèves*" et "*encourager l'interaction enseignant-élèves*"). De cette manière, nous observons qu'il y une tendance entre les sujets à classer

V40H ("*attirer l'attention des élèves*") au premier rang. Puis vient V40A ("*encourager la participation des élèves*"). Ce qui signifie que V40H est préféré à V40A dans notre échantillon. De son côté, V40A est préféré respectivement à V40B ("*promouvoir l'interaction élèves-élèves*") et à V40I ("*susciter la motivation des élèves pour apprendre*"). Nous observons alors qu'avec un degré de confiance à 0.85, il y a une tendance chez les enseignants à ranger V40K ("*encourager l'interaction enseignant-élèves*") juste après V40B. Une fermeture transitive à 0.64 lie directement V40K à V40A comme le montre les flèches pointillées (figure 3). Une autre fermeture transitive à 0.79 s'effectue entre V40I et V40H.

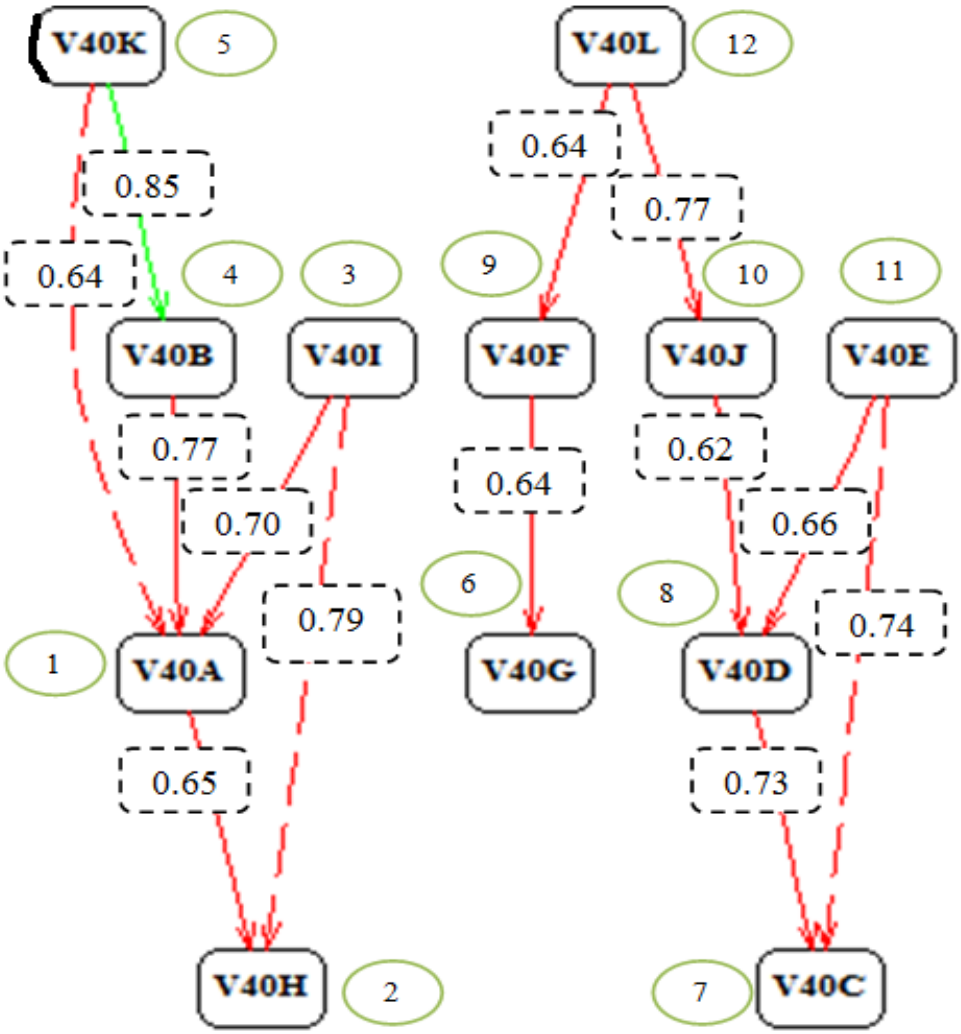


Figure 3 – Troisième figure (graphe implicatif organisant les 12 potentiels pédagogiques du TNI au niveau de confiance de ≥0.62)

Sept variables composent le second groupe. Dans ce groupe implicatif on voit que les sujets ont tendance à placer V40L "*diversifier les méthodes d'enseignement-apprentissage*", en dernière position précédée de V40F "*varier les ressources pédagogiques utilisées en classe*" et de V40J "*aider à représenter facilement les relations fractionnaires*". Dans ce groupe, nous voyons également une fermeture transitive à 0.74 entre V40E ("*transmettre plus facilement le contenu des cours*") et V40C ("*aider à comprendre des notions complexes*").

Concernant les chemins de propension figurés dans le graphe implicatif nous pouvons dire qu'une variable supplémentaire peut (homme, âge, etc.) contribuer à la création d'un chemin comme le montre le tableau 7(ci-après)[4]. Pour le chemin V40K-V40A-V40H, les hommes sont ceux qui contribuent le plus à ce chemin avec un niveau de risque de 0.079 alors que la variable « ancienneté1 » contribue le moins à ce chemin avec un risque de 0.926. « Ancienneté2 » participe le plus au chemin V40K-V40B-V40A-V40H avec un risque de 0.149.

Chemin	Contribution des variables supplémentaires
V40K-V40A-V40H	Ancienneté1 [0.926], Ancienneté2 [0.182], Ancienneté3 [0.211], Âge1 [0.617], Âge2 [0.28], Âge3 [0.282], Femme [0.784], Homme [0.079], CM1 [0.272], CM2 [0.617], CM1-CM2 [0.274], Formation_oui [0.394], Formation_non [0.475].
V40K-V40B-V40A-V40H	Ancienneté1 [0.821], Ancienneté2 [0.149], Ancienneté3 [0.483], Âge1 [0.481], Âge2 [0.426], Âge3 [0.284], Femme [0.601], Homme [0.229], CM1 [0.313], CM2 [0.313], CM1-CM2 [0.731], Formation_oui [0.402], Formation_non [0.591].
V40K-V40B-V40A	Ancienneté1 [0.282], Ancienneté2 [0.449], Ancienneté3 [0.447], Âge1 [0.441], Âge2 [0.483], Âge3 [0.282], Femme [0.354], Homme [0.508], CM1 [0.0596], CM2 [0.617], CM1-CM2 [0.854], Formation_oui [0.394], Formation_non [0.475].
V40I-V40A-V40H	Ancienneté1 [0.798], Ancienneté2 [0.0998], Ancienneté3 [0.659], Âge1 [0.218], Âge2 [0.647], Âge3 [0.38], Femme [0.139], Homme [0.796], CM1 [0.112], CM2 [0.706], CM1-CM2 [0.49], Formation_oui [0.444], Formation_non [0.323].
V40L-V40F-V40G	Ancienneté1 [0.229], Ancienneté2 [0.657], Ancienneté3 [0.0999], Âge1 [0.782], Âge2 [0.134], Âge3 [0.173], Femme [0.139], Homme [0.753], CM1 [0.925], CM2 [0.0821], CM1-CM2 [0.0841], Formation_oui [0.359], Formation_non [0.389].
V40E-V40D-V40C	Ancienneté1 [0], Ancienneté2 [0.33], Ancienneté3 [0.98], Âge1 [0.419], Âge2 [0.209], Âge3 [0.632], Femme [0.377], Homme [0.479], CM1 [0.419], CM2 [0.121], CM1-CM2 [0.9], Formation_oui [0.485], Formation_non [0.263].

Tableau 7 - Septième tableau (contribution des variables supplémentaires aux chemins de propension)

Il est important de souligner qu'il y a certaines inversions dans le rang des variables selon le graphe implicatif par rapport au rang donné par le test de Kendall (dans la figure 3 on voit les rangs donnés par Kendall dans des cercles verts). Par exemple, le test de Kendall classe V40A ("*encourager la participation des élèves*") en première

[4] Dans notre graphe il y a d'autres chemins mais dans le tableau 7 nous n'avons présenté que les chemins les plus longs et les contributions des variables à chaque chemin. On voit en rouge la variable qui contribue le plus au chemin envisagé.

position et V40H ("*attirer l'attention des élèves*") en deuxième. Cependant dans l'ASI nous trouvons l'inverse.

7.2.3 L'arbre cohésitif : classification hiérarchique orientée par la relation de quasi-implication

Par rapport aux résultats de l'arbre cohésitif, celui-ci nous montre le degré de cohésion entre les variables principales (voir figure 4). Dans la figure 4 nous voyons que la majorité absolue des variables sont liées. Nous avons une seule variable isolée : V40G. C'est-à-dire, que selon l'arbre cohésitif cette fonction pédagogique du TNI n'implique aucun autre potentiel pédagogique de l'usage du TNI.

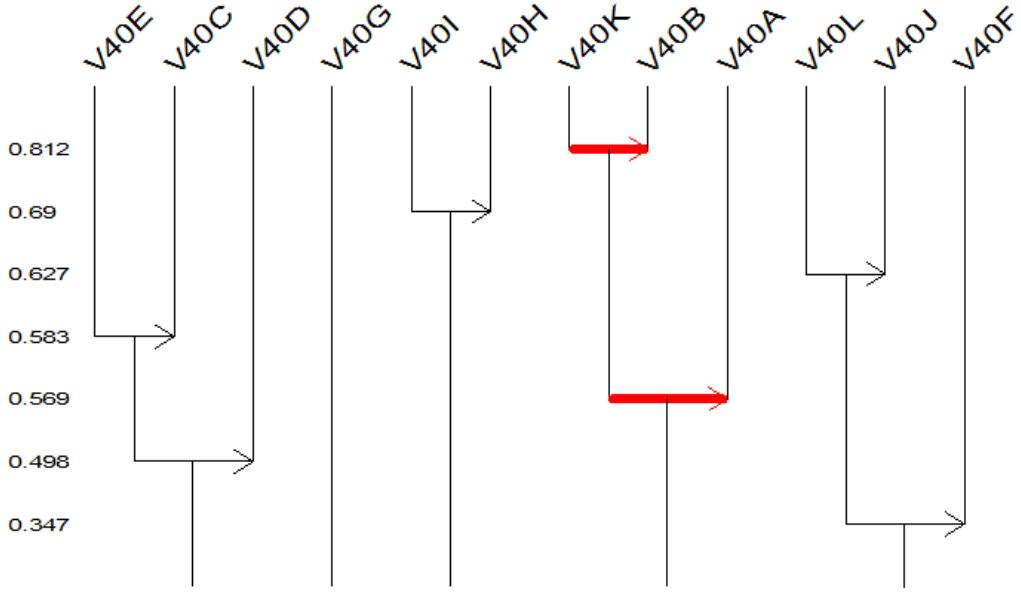


Figure 4 - Quatrième figure (arbre cohésitif organisant les 12 potentiels pédagogiques du TNI selon une classification hiérachique orientée par la relation de quasi-implication)

Comme le montre la figure 4, les niveaux de cohésion dans notre exemple varient de 0.812 à 0.347. Le degré de cohésion le plus fort correspond à 1 : plus on descend, moins il y a de cohésions identifiées entre les variables. De cette manière, au premier niveau de la hiérarchie, nous remarquons que la classe (V40K, V40B) est créée. Elle représente le fait que la variable V40K ("*encourager l'interaction enseignant-élèves*") implique la variable V40B ("*promouvoir l'interaction élèves-élèves*") avec une intensité plus forte que tous les autres couples de variables. Ce premier niveau de la hiérarchie est d'ailleurs significatif comme l'indique la flèche rouge. Au deuxième niveau, nous observons que le fait d'affirmer que le TNI suscite la motivation des élèves (V40I) implique le fait que le TNI attire l'attention des élèves (V40H). De même, V40L "*diversifier les méthodes d'enseignement-apprentissage*" et V40J "*aider à représenter facilement les relations fractionnaires*" ont une cohésion assez forte. Au quatrième niveau V40K, V40B et V40A forment une classe. Cette classe à trois composantes admet la plus forte cohésion parmi celles de toutes les classes possibles à trois composantes. Ce niveau de la hiérarchie est d'ailleurs significatif comme l'indique la flèche rouge. V40L, V40J et V40F forment enfin une dernière classe.

8 Conclusion

Dans le cadre de cette recherche portant sur les potentiels pédagogiques du TNI à l'école primaire du point de vue des enseignants nous avons eu recours à deux approches statistiques pour traiter et analyser les données obtenues à partir de l'échelle de préférence : le test W de Kendall et l'analyse statistique implicative. Comme nous l'avons vu, les enseignants ont des points de vue communs s'agissant des potentialités pédagogiques de ce dispositif techno-sémio-pédagogique. C'est-à-dire, qu'il y a certains accords entre eux s'agissant de la pertinence de telles ou telles fonctions pédagogiques. Pour eux, le TNI n'est pas seulement un instrument technique mais plutôt un instrument pédagogique qui peut introduire de réels changements dans les pratiques pédagogiques et dans la pédagogie elle-même. Les enseignants estiment que le plus pertinent dans l'usage du TNI est d'encourager la participation des élèves et de conserver leurs attentions. Les enseignants apprécient également que cet outil nomade joue un rôle en favorisant les relations dans la classe entre l'enseignant et les élèves, et entre les élèves eux-mêmes.

Grâce à l'approche de l'analyse statistique implicative, nous avons pu étudier les différents aspects pédagogiques du TNI et bien comprendre les relations entre les nombreuses fonctions pédagogiques de cet instrument numérique. L'utilisation des trois graphes proposés par l'analyse statistique implicative met en avant leurs différentes applications selon les questions abordées. Cela nous conduit bien entendu à une meilleure compréhension de la nature des relations existantes entre les potentialités pédagogiques de cet instrument. L'arbre de similarité nous a permis de comprendre et d'interpréter l'organisation des douze potentiels pédagogiques du TNI. Cette organisation nous a amené à identifier les potentiels se ressemblant le plus. Il nous a aidés aussi à mettre en lumière les potentiels dissemblables. Le graphe implicatif quant à lui nous a permis d'obtenir un schéma sur les potentiels pédagogiques du TNI ayant une intensité d'implication supérieure à un certain seuil. Il nous a montré les différents chemins possibles entre les variables étudiées en précisant la contribution de chaque variable supplémentaire à la construction de chemins. L'arbre cohésitif nous a permis aussi de voir le niveau de cohésion entre les potentiels pédagogiques de l'instrument mis en œuvre. Cela a conduit à nous interroger sur les classes de potentiels pédagogiques du TNI constituées à partir des implications de ceux-ci. Ces trois graphes nous ont dirigés vers des interprétations plus fines et plus détaillées de nos résultats.

Les résultats issus de ces deux approches statistiques constituent enfin une preuve de l'importance et de la pertinence de l'usage du TNI en classe. Ils nous montrent que les effets du TNI dépassent l'aspect affectif vers des perspectives pédagogiques qui touchent directement les processus d'enseignement et d'apprentissage. L'élève, dans ces perspectives, est un producteur et un acteur au lieu d'être un récepteur ou un spectateur de ce qui se passe en classe. De plus, ces résultats nous donnent un bon exemple de la durabilité des effets de l'usage du TNI dans la salle de classe. En outre, ces deux traitements traitement statistiques conduisent également à la mise en évidence de similarités et de liaisons implicatives entre les potentiels du TNI.

Références

[1] Bakadam, E., & Asiri, M. J. (2012). Teachers' Perceptions Regarding the Benefits of Using the Interactive Whiteboard (IWB) : The Case of a Saudi Intermediate School. *Procedia - Social and Behavioral Sciences , 64*, 179-185.

[2] Beauchamp, G. (2004). Teacher use of the interactive whiteboard in primary schools : towards an effective transition framework. *Technology, Pedagogy and Education , 13* (3), 327-348.

[3] Beauchamp, G., & Kennewell, S. (2010). Interactivity in the classroom and its impact on learning. *Computers & Education , 54*, 759–766.

[4] Beauchamp, G., & Kennewell, S. (2008). The influence of ICT on the interactivity of teaching. *Education and Information Technologies , 13* (4), 305-315.

[5] Becta. (2004). *Getting the most from your interactive whiteboard. A guide for primary schools.* British Educational Communications and Technology Agency, Coventry : Grande-Bretagne.

[6] Becta. (2003). *What the research says about interactive whiteboards.* British Educational Communications and Technology Agency Coventry, Grande-Bretagne.

[7] Burton-Monney, S., & Jauquier, L. (2010). *TBI : Enseigner avec les tableaux blancs interactifs.* Berne: Educa.Guides - Institut suisse des médias pour la formation et la culture en Suisse.

[8] Dostál, J. (2011). Reflections on the Use of Interactive Whiteboards in Instruction in International Context. *The New Educational Review , 25* (3), 205-220.

[9] Glover, D., & Miller, D. (2001). Running with Technology : the pedagogic impact of the large-scale introduction of interactive whiteboards in one secondary school. *Journal of Information Technology for Teacher Education , 10* (3), 257-278.

[10] Glover, D., Miller, D., & Averis, D. (2003). The impact of interactive whiteboards on classroom practice: examples drawn from the teaching of mathematics in secondary schools in England. Brno, Czech Republic: The Mathematics Education into the 21st Century Project Proceedings of the International Conference.

[11] Gras, R., Régnier, J.-C., & Guillet, F. (2009). *Analyse statistique implicative: méthode exploratoire et confirmatoire à la recherche de causalités.*Revue des Nouvelles Technologies de l'Information RNTI-E-16. Toulouse, France: Cépaduès.

[12] Gras, R., Régnier, J.-C., Marinicia, C., & Guillet, F. (2013). *L'analyse statistique implicative: méthode exploratoire et confirmatoire à la recherche de causalités*. Toulouse, France: Cépaduès.

[13] Greiffenhagen, C. (2000). Interactive whiteboards in mathematics education : possibilities and dangers, Paper presented at WGA 11 (ICME-9), The Use of Technology in Mathematics Education.

[14] Hall, I., & Higgins, S. (2005). Primary school students' perceptions of interactive whiteboards. *Journal of Computer Assisted learning , 21 (2005),* PP102-107.

[15] Heutte, J., & Tempez, F. (2008). Quand une technologie rassurante renforce le sentiment d'efficacité personnelle et le plaisir d'enseigner. Dossiers de l'ingénierie éducative, hors série. Dans P. Claus, *Les TICE au service des èlèves du primaire* (pp. 101-106). Paris: Futuroscope : Scérén-CNDP.

[16] Jeunier, B., Morcillo-Bareille, A., Camps, J.-F., Galy-Marié, E., & Tricot, A. (2005). *Expertise relative aux usages du tableau blanc interactif en école primaire.* Paris. Rapport remis à la Direction de la Technologie / SDTICE Ministère de l'Education Nationale dans le cadre du projet PrimTice.

[17] Kennewell, S., & Beauchamp, G. (2007). The features of interactive whiteboards and their influence on learning. *Learning, Media and Technology , 32* (3), 227–241.

[18] Lagrange, J.-B. (1998). Analyse implicative d'un ensemble de variables numériques ; aaplication au traitement d'un questionnaire aux réponses modales ordonnées. *Revue de statistique appliquée , 46* (1), 71-93.

[19] Lefebvre, S., & Samson, G. (2013). État des connaissances sur l'implantation du tableau numérique interactif (TNI) à l'école. *Sticef , 20*.

[20] Leroux, L. (2009). Le tableau numérique interactif : quelles spécifites vis-à-vis d'autres dispositifs? Quand et pourquoi l'utiliser? *REPERES - IREM Grenoble , 74*, pp. 45-52.

[21] Levy, P. (2002). *Interactive whiteboards in learning and teaching in two Sheffield schools: a developmental study*. Sheffield : Department of Information Studies, University of Sheffield .

[22] Lewin, C., Somekh, B., & Steadman, S. (2008). Embedding interactive whiteboards in teaching and learning : The process of change in pedagogic practice. *Education and Information Technologies , 13* (4), 291-303.

[23] López, O. S. (2010). The Digital Learning Classroom: Improving English Language Learners' academic success in mathematics and reading using interactive whiteboard technology. *Computers & Education , 54*, 901-915.

[24] Magnat, E. (2013). *Le TBI comme instrument du développement de la conscience phonémique à l'école : une approche ergonomique*. Grenoble: Université Stendhal Grenoble 3. Thèse en Didactique et Linguistique.

[25] Marzano, R. J. (2009). The Art and Science of Teaching / Teaching with Interactive Whiteboards. *Educational leadership* , 80-82.

[26] McIntyre-Brown, C. (2011). *Understanding the Next Wave of Technology Innovation in Education*: UK. Bedfordshire.

[27] Miller, D., Glover, D., & Averis, D. (2005b). *Motivation : the contribution of interactive whiteboards to teaching and learning in mathematics.* Consulté le 10 juin 2014, sur http://try.iprase.tn.it/en/activities/studyAndResearch/download/03_Interactive_whiteboard_and_mathematics.pdf

[28] Miller, D., Glover, D., & Averis, D. (2005a). Presentation and pedagogy: the effective use of interactive whiteboards in mathematics lessons. Dans D. Hewitt, & A. Noyes, *Proceedings of the sixth British Congress of Mathematics Education* (pp. 105-112). Coventry, Royaume-Uni: University of Warwick.

[29] Peraya, D. (1999). Médiation et médiatisation : le campus virtuel. *Hermès, La Revue , 25*, 153-167.

[30] Régnier, J.-C., & Gras, R. (2005). Statistique de rangs et analyse implicative. *Revue de statistique appliquée , 1,* 5-38.

[31] Régnier, J.-C., & Gras, R. (2013). Statistique de rangs et Analyse Statistique Implicative. Dans R. Gras, J.-C. Régnier, C. Marinica, & F. Guillet, *L'analyse statistique implicative. Méthode exploratoire et confirmatoire à la recherche de causalité* (pp. 187-206). Toulouse, France: Cépaduès.

[32] Slay, H., Sieborger, I., & Hodgkinson-Williams, C. (2008). Interactive whiteboards: Real beauty or just ‘‘lipstick”? *Computers & Education , 51,* 1321–1341.

[33] Smith, F., Hardman, F., & Higgins, S. (2006). The impact of interactive whiteboards on teacher–pupil interaction in the National Literacy and Numeracy Strategies. *British Educational Research Journal , 32* (3), 443-457.

[34] Smith, H. J., Higgins, S., Wall, K., & Miller, J. (2005). Interactive whiteboards: boon or bandwagon? A critical review of the literature. *Journal of Computer Assisted Learning , 21,* 91–101.

[35] Somekh, B., Haldane, M., Jones, K., Lewin, C., Steadman, S., Scrimshaw, P., et al. (2007). *Evaluation of the Primary Schools Whiteboard Expansion Project - summary report.* Education & Social Research Institute, Centre for ICT, Pedagogy and Learning. Manchester: Metropolitan University, July 2007. Rapport proposé au Becta.

[36] Swan, K., Schenker, J., & Kratcoski, A. (2008). *The Effects of the Use of Interactive Whiteboards on Student Achievement.* Consulté le 2 juin 2014, sur http://edtech2.boisestate.edu/spechtp/551/07_The_Effects_of_the_Use_of_Interactive_Whiteboards_on_Student_Achievement.pdf

[37] Tataroglu, B., & Erduran, A. (2010). Examining students’ attitudes and views towards usage an interactive whiteboard in mathematics lessons. *Procedia Social and Behavioral Sciences , 2,* 2533-2538.

[38] Torff, B., & Tirotta, R. (2010). Interactive whiteboards produce small gains in elementary students’ self-reported motivation in mathematics. *Computers & Education , 54,* 379-383.

[39] Toussaint, I. (2010). *Le tableau blanc interactif : Un pas vers le changement?* Guide annuel 2009-2010 • 500 sites Web pour réussir à l’école: Article en ligne à l'adresse : http://eduportfoliov3.weebly.com/uploads/1/0/7/3/10733334/cahier_tbi.pdf, consulté le 20/05/2014.

[40] Türel, Y. K., & Johnson, T. E. (2012). Teachers’ Belief and Use of Interactive Whiteboards for Teaching and Learning. *Educational Technology & Society , 15* (1), 381-394.

[41] Wall, K., Higgins, S., & Smith, H. (2005). The visual helps me understand the complicated things :pupil views of teaching and learning with interactive whiteboards. *British Journal of Educational Technology , 36* (5), 851–867.

[42] Zittle, F. J. (2004). Enhancing Native American Mathematics Learning: The Use of Smartboard-generated Virtual Manipulatives for Conceptual Understanding. Consulté le 2 juin 2014, sur http://downloads01.smarttech.com/media/research/international_research/usa/cerzittle.pdf

MAPEAMENTO DE CONHECIMENTOS DE PROFESSORES SOBRE TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO E SEUS USOS DIDÁTICO-PEDAGÓGICOS

Saddo AG ALMOULOUD[1], Cileda de QUEIROZ E SILVA COUTINHO[2], Maria José FERREIRA DA SILVA[3]

CARTOGRAPHIE DES CONNAISSANCES DES PROFESSEURS SUR LES TIC ET LEURS USAGES DIDACTICO-PEDAGOGIQUES

RESUMO

O objetivo de texto é apresentar resultados parciais de um projeto de pesquisa sobre os processos de construção de conhecimentos de geometria mediada por Tecnologias de Informação e Comunicação. As questões norteadoras do projeto foram: quais fatores influenciam o processo de ensino e de aprendizagem da Geometria com a utilização de recursos alternativos como ambientes computacionais? Como os processos de aprendizagem se caracterizam nos ambientes computacionais? Quais são as alternativas metodológicas para investigarmos os processos de aprendizagem nesses ambientes? Neste artigo, analisamos os dados oriundos de um questionário diagnóstico, que foi respondido por professores participantes do projeto, como um dos instrumentos para coleta de dados referentes às concepções sobre utilização das TIC nas aulas de Matemática. Descrevemos os aspectos metodológicos e apresentamos resultados das análises dos mapas gerados pelo tratamento de dados do questionário por CHIC (Análise Hierárquica, Implicativa e Coesitiva). No que se refere à primeira questão, podemos inferir que para os professores participantes, o grau de familiaridade com os recursos tecnológicos é fator importante para sua utilização pelos professores, de onde a necessidade de um trabalho para formação, atualização das TIC nas escolas, o que geraria oportunidade para que todos os alunos tivessem acesso a tais ferramentas. Os dados obtidos por meio desse questionário não nos permitem responder à segunda questão, mas podemos inferir que no que se refere à terceira questão, uma alternativa metodológica que podemos visar é a formação para o uso das TIC com utilização de grupos colaborativos, pois a interação com outros professores pode favorecer a superação de preconcepções ou pré-conceitos sobre as TIC.

Palavras-chave: *Formação de professores. Tecnologia de Comunicação e de Informação. Classificação Hierárquica, Implicativa e Coesitiva. Concepções de professores.*

RESUME

Le but du texte est de présenter les résultats partiels d'un projet de recherche sur les processus de construction des connaissances et savoirs de géométrie médiée par les technologies de l'information et de communication. Les principales questions du projet étaient : Quels facteurs influencent les processus d'enseignement et d'apprentissage de la géométrie à l'aide d'autres ressources telles que les environnements informatiques ? Comment les processus d'apprentissage sont caractérisés dans des environnements informatiques ? Quelles sont les alternatives méthodologiques pour étudier les processus d'apprentissage dans ces milieux ? Dans cet article, nous analysons les données d'un questionnaire de diagnostic soumis à des enseignants participant au projet, comme l'un des

[1] PUC/SP saddoag@pucsp.br
[2] PUC/SP cileda@pucsp.br
[3] PUC/SP zeze@pucsp.br

instruments de collecte de données liées aux conceptions de l'utilisation des TIC dans les classes de mathématiques. Nous décrivons les aspects méthodologiques et présentons les résultats de l'analyse des cartes générées par le traitement des données du questionnaire par CHIC (Classification Hiérarchique, Implicative et Cohésitive). En ce qui concerne la première question, nous pouvons déduire que, pour les enseignants, le degré de familiarité avec les moyens technologiques est un facteur important pour leur utilisation en classe, d´où la nécessité d'un travail pour la formation des enseignants, la mise en place des TIC dans les écoles, qui générerait l´occasion pour tous les élèves à avoir accès à ces outils. Les données obtenues par le biais de ce questionnaire ne nous permettent pas de répondre à la deuxième question, mais nous donne des indices pour construire une ingénierie pour la formation de ces enseignants à l'utilisation des TIC avec l'utilisation TIC comme outil didactique pour l´enseignement des mathématiques.

Mots-clés : *Formation des enseignants. Technologie de l´information et de communication. Classification Hiérarchique, Implicative et Cohésitive. Conceptions des enseignants.*

1 Introdução

Neste artigo, apresentamos uma parte dos resultados de um projeto que trata da contribuição de tecnologias, por via de ambientes computacionais, na construção de conhecimentos de Geometria. Fazemos a hipótese de que a utilização de certos ambientes tecnológicos com características de dinamismo pode facilitar a visualização e a percepção de propriedades dos objetos matemáticos em jogo. Este projeto, desenvolvido pelo grupo de pesquisa PEA-MAT, PUC-SP, tem por objetivo analisar tanto do ponto de vista teórico, quanto do prático, questões relativas à complexidade da inserção de ferramentas tecnológicas para o ensino e aprendizagem da Geometria na Escola Básica. As questões norteadoras do projeto foram: *quais fatores influenciam o processo de ensino e de aprendizagem da Geometria com a utilização de recursos alternativos como ambientes computacionais? Como os processos de aprendizagem se caracterizam nos ambientes computacionais? Quais são as alternativas metodológicas para investigarmos os processos de aprendizagem nesses ambientes?*

Para buscar respostas a estas questões adotamos os pressupostos da Engenharia Didática e da Pesquisa-Ação, e recorreremos a instrumentos diagnósticos de questionários estruturados, com objetivo de identificar **concepções** de professores de Ensino Básico a respeito de conceitos geométricos tratados e suas representações sobre as Tecnologias de Informação e Comunicação (TIC) aplicadas à educação matemática. A partir do diagnóstico, analisamos os resultados obtidos e definimos os caminhos da formação realizada como parte fundamental no desenvolvimento do projeto. A expressão “concepções de professores” (ou representações de professores) é tomada neste trabalho como um modelo construído pelo pesquisador para analisar o que pensam os professores participantes do projeto a respeito da integração das TIC como ferramentas para ensinar e aprender matemática. Entendemos que este trabalho se justifica pela necessidade de compreender o papel das tecnologias para o ensino e a aprendizagem de Matemática, neste texto, particularmente de Geometria, procurando identificar contribuições didáticas para a utilização dessas tecnologias na elaboração e implementação de sequências de ensino para o Ensino Básico. No que trata da contribuição das TIC na construção de conhecimentos de Geometria, sabemos que o uso de alguns softwares pode facilitar a visualização e a percepção de propriedades, favorecendo a aprendizagem. Nosso intuito é conduzir os professores, em ~~sua~~ formação

continuada, à aprendizagem de conceitos geométricos e à utilização de ambientes computacionais como instrumento para construção de conjecturas para resolução de problemas, de forma a que possam conceber, com autonomia, situações de aprendizagem para seus próprios alunos. Como pesquisadores, os participantes do grupo PEA-MAT preocupam-se em estudar a integração da informática no ensino e sua complexidade que, segundo Trouche (2005), nos remete a relacionar pesquisas em diferentes contextos e a emergência de novas abordagens teóricas. Para o autor, uma abordagem proeminente consiste em situar a integração de ferramentas matemáticas em uma problemática mais geral de integração de ferramentas da atividade humana. Um de nossos objetivos é contribuir para a formação continuada dos professores participantes do projeto. Em relação a esse aspecto, Showers, Joyce e Bennet (1987) afirmam que, em uma formação continuada visando o desenvolvimento profissional, os professores trazem seus conhecimentos e habilidades, seu estilo de ensino, suas características pessoais tais como estágio de crescimento, flexibilidade conceitual, senso de eficácia e conceitos, além de percepções sobre suas necessidades e preferências por certos tipos de desenvolvimento profissional. Nessa perspectiva, Pacheco e Flores afirmam também que:

> Na prática, o que os professores pensam, fazem, escrevem, verbalizam deve-se, por um lado, a um conhecimento que é o resultado de um processo aquisitivo e, por outro, a um conhecimento que se consubstancia num discurso sobre uma prática ou um modo de ação. (PACHECO; FLORES, 1999, p. 15).

Para esses autores, a tarefa do professor é selecionar e ordenar atividades didáticas, adaptando-as a uma situação específica, incluindo a elaboração didática de novas atividades, de acordo com suas concepções, para complementar as já existentes nos materiais curriculares. Neste artigo analisamos os resultados de um questionário diagnóstico que foi respondido por professores participantes do projeto, como um dos instrumentos para coleta de dados referentes às concepções sobre utilização das TIC nas aulas de Matemática. Descrevemos os aspectos metodológicos e apresentamos resultados das análises dos mapas gerados pelo tratamento de dados do questionário por CHIC (Análise Hierárquica, Implicativa e Coesitiva).

2 Aspectos metodológicos

Para a execução do projeto firmamos um convênio com a Diretoria de Ensino Região Guarulhos-Sul (Secretaria da Educação do Estado de São Paulo) no intuito de formar 80 professores no tema “Geometria e sua aprendizagem por meio de tecnologia”**.** Os objetivos dessa formação eram:

- Apoiar o professor na construção de boas intervenções para que possa promover a aprendizagem de seus alunos em Matemática.
- Criar referenciais teórico-práticos para discutir com os professores alguns modelos de gestão de aula que favoreçam a aprendizagem de seus alunos.

- Apresentar aos professores de modo contextualizado, as situações presentes nos Cadernos do Professor e do Aluno da Secretaria de Estado da Educação (SEE-SP), a fim de favorecer a construção de conhecimentos sobre a Matemática nos seus diferentes blocos de conteúdo e ressaltar uma visão não linear que está presente no material.
- Contribuir para a compreensão do papel do professor considerando a intervenção didática e a interação entre os alunos de modo a favorecer o avanço no fazer Matemático.
- Apoiar os professores na organização de conteúdos em diferentes pontos de vista para ressaltar o desenvolvimento de sequências didáticas e favorecer uma rotina de aula produtiva.
- Discutir com os professores a validação de conhecimentos do grupo de alunos e promover interações entre eles com o intuito de dar visibilidade às novas aprendizagens sem desconstruir os saberes mobilizados pelo aluno.

O público alvo era composto de professores de Matemática e professores Coordenadores da Rede pública estadual de escolas de Ensino Fundamental II e Ensino Médio, com habilitação em Matemática, que atuam nas Unidades Escolares da Diretoria de Ensino Região Guarulhos Sul.

Os conteúdos abordados foram:

- Estudo de conceitos geométricos no plano e no espaço.
- Fundamentos teóricos dos fenômenos de ensino e aprendizagem da Geometria.
- Geometria em ambientes dinâmicos.
- Estudo de objetos científicos que se tornam objetos de ensino como, por exemplo: o Teorema de Thales, o Teorema de Pitágoras, a semelhança de figuras, a noção de áreas, os quadriláteros, os triângulos, os círculos e as circunferências etc.
- Alguns fundamentos teóricos e didáticos das transformações geométricas no plano tais como a simetria axial, a simetria central, a translação, a rotação, etc.
- Alguns fundamentos teóricos e didáticos da Geometria Espacial.

A formação foi organizada em 60 horas, das quais 32 horas presenciais e 28 horas de estudos a distância para a realização de leituras, análise de vídeos e atividades em ambientes dinâmicos (Geogebra, por exemplo) etc. Utilizamos como recursos tecnológicos e materiais de apoio o currículo (Currículo Oficial, Cadernos do Professor e Aluno do Estado de São Paulo); computador, o software Geogebra e o ambiente Moodle, disponibilizado pela Pontifícia Universidade Católica de São Paulo. Para o acompanhamento e a avaliação dos participantes e do curso foram realizadas atividades ou relatórios individuais produzidos pelos próprios professores; atividades individuais realizadas pelos professores a partir de situações para o ensino de Geometria; acompanhamento de atividades via o ambiente computacional Moodle; avaliação final da formação no último encontro.

2.1 Primeira etapa da execução do projeto: questionário

Elaboramos e aplicamos um questionário estruturado (anexo 1), no intuito de identificar concepções dos professores participantes da formação sobre o ensino e a aprendizagem da geometria e sobre o uso de ambientes computacionais em educação matemática. Esse instrumento de coleta de dados foi aplicado no início do projeto, em 2011, em formato virtual (formulário do Google-Docs), e respondido, espontaneamente, no local onde o professor desejasse. Dessa forma, não temos os registros de duração do processo de preenchimento dos itens. Com relação à coleta de informações, Almeida (2008) afirma que:

> A coleta de informações, bem como sua organização, sistematização e estabelecimento de conexões, implicam mapear as informações, estabelecer conexões entre elas, analisá-las por múltiplos pontos de vista, desenvolver a visão panorâmica e simultaneamente enxergar cada dimensão associada e suas conexões provisórias, reconhecendo a relatividade do conhecimento e a transitoriedade das informações. (p. 325)

Corroborando com essa autora, a análise dos dados assim obtidos foi feita analisados usando ASI (Análise Estatística Implicativa). Lembramos que esse tipo de análise, desenvolvida por Gras (2014) e colaboradores, destina-se a identificar e a estruturar em forma de regras, um conjunto de dados, cruzando sujeitos (ou objetos) e variáveis por meio de uma modelagem estatística de quase-implicação: *se a variável ou uma combinação de variáveis a é observada na população, então, em geral, a variável b é também observada*.

As variáveis envolvidas podem ser de vários tipos: binário, modal, numérico, intervalar etc. Os conjuntos de regras obtidos podem ser estruturados de acordo com diferentes abordagens complementares (gráfico implicativo, hierarquia orientada). A visualização dos resultados (A.S.I. e Hierarquia de Similaridade) e sua interpretação são facilitadas pelo software C.H.I.C. (Classificação Hierárquica Implicativa e Coesitiva). De acordo com Almouloud (2008), tais análises permitem constituir por meio de um critério probabilístico, partições cada vez mais finas sobre um conjunto de variáveis estatísticas. Estas partições são construídas de modo ascendente em uma árvore (dendograma no caso da similaridade e da coesitiva, grafo no caso da implicativa), possibilitando o estudo e a interpretação de classes de variáveis em termos de tipologia, semelhança (dessemelhança), coesão e implicação. Os critérios para o agrupamento geram um índice de 0 a 1 relacionado a cada um deles, indicando sua probabilidade da ocorrência. Tivemos também a preocupação de identificar as variáveis suplementares (idade, tempo de magistério, gênero, por exemplo) ou sujeitos (ou grupos de sujeitos) que são típicos das classes formadas. A tipicalidade é uma informação que colabora na interpretação dos dados pela análise das variáveis de contexto que caracterizam os tipos de respostas associadas em tais agrupamentos. Digamos que uma variável é principal toda variável considerada normalmente pelo CHIC. Como padrão, todas as variáveis são principais. As variáveis suplementares (em geral são as descritivas), são variáveis que não interferem no cálculo das contribuições das categorias. Por exemplo, se desejamos saber se uma implicação é na maior parte formada por pessoas do sexo feminino ou masculino, definimos para cada indivíduo a variável “sexo”. Em seguida, estas variáveis são consideradas apenas quando da procura da contribuição ou da tipicidade das categorias. Para definir uma variável secundária ou suplementar, acrescentamos ao

nome da variável um espaço e um "s". Assim, o atributo *Feminino* é, por exemplo, codificada *Fem* como variável principal e *Fem s* como variável suplementar.

2.2 Estrutura do questionário

O questionário é composto de duas partes. A primeira parte tem por objetivo procurar informações sobre o perfil dos professores, mas também suas opiniões sobre as TIC aplicadas ao ensino e à aprendizagem da Matemática. A segunda parte está voltada para o estudo das representações desses professores no que diz respeito aos processos de ensino e de aprendizagem da geometria no Ensino Básico e a importância das TIC nesses processos. O questionário ao qual este texto se refere foi composto por 17 questões. As questões de 1 a 14 procuram informações sobre o perfil dos professores participantes da pesquisa (gênero, idade, formação tempo de magistério). As questões de 15 a 17 têm por objetivo fazer um levantamento de dados que permitem analisar o que pensam os sujeitos da pesquisa sobre as TIC aplicadas ao ensino e aprendizagem da matemática, a necessidade da formação deles e a importância do ambiente de trabalho. Apresentamos no apêndice a lista das variáveis com seus códigos.

2.3 Perfil dos professores

Participaram da pesquisa 31 professores da rede estadual de São Paulo, diretoria de Guarulhos Sul, dos quais 10 eram efetivos e os demais trabalhavam em caráter de "professor eventual". A predominância foi do gênero feminino, representado por 19 professores. Verificou-se que 19 professores possuíam de 21 a 40 anos de idade. Dos participantes, 18 eram casados ou viviam em união estável e 29 cursaram Licenciatura Plena em Matemática (1 professor cursou Licenciatura Curta e outro cursou Bacharelado em Matemática). Quanto às informações relacionadas à atuação profissional, 11 sujeitos lecionam matemática há 6 anos ou menos, conforme Tabela 1, caracterizando um grupo com relativa experiência docente:

Tempo que leciona ou lecionou matemática	Nº de professores
Até 3 anos	8
4 a 6 anos	3
7 a 18 anos	17
19 a 30 anos	3
Total	31

Tabela 1. Distribuição dos professores em relação ao tempo que lecionam ou lecionou matemática **(Fonte**: autores deste texto)

Destes 17 professores com 7 a 18 anos que lecionam Matemática, 16 afirmam ter também entre 7 e 18 anos de magistério, caracterizando o grupo como professores prioritariamente de Matemática. Tal informação é relevante para compreendermos algumas das associações feitas na análise dos dados, que indiquem comportamentos associados ao tempo de magistério e, principalmente, tempo de magistério de Matemática, como apresentamos mais adiante no texto, no item referente às análises multidimensionais. Outra informação relevante na caracterização dos professores refere-se à rotatividade nas escolas: dos 31 respondentes, 14 afirmam estar na mesma escola há

mais de 4 anos. Para um grupo formado principalmente por professores não efetivos, essa permanência nas escolas deve ser destacada, uma vez que permite que retornem para a escola os conhecimentos construídos durante a formação, sejam estes conhecimentos relativos às práticas docentes como aos conhecimentos específicos da Matemática. Para a realização da formação, foi prevista a utilização de Ambiente Virtual de Aprendizagem (AVA), no caso a plataforma *Moodle*) e a utilização do software *Geogebra* como ferramenta computacional para elaboração e resolução de atividades. Buscou-se assim conhecer as características dos participantes quanto ao uso de TIC e à disponibilidade em relação à *internet*. Os resultados observados são apresentados na Figura 1 e constatou-se também que, dos 31 professores respondentes, apenas oito já fizeram algum curso online.

Figura 1. Distribuição dos professores quanto ao local de uso da internet e à experiência no uso de computadores, respectivamente. (**Fonte**: autores deste texto)

3 Análise quanto ao uso das tecnologias da informação e da comunicação (TIC)

Passamos a apresentar os resultados das análises dos mapas gerados pelo software CHIC quanto às associações por similaridade, por coesão e associações implicativas. As variáveis estatísticas relativas às questões 1 a 14 são consideradas variáveis suplementares, e as variáveis relativas às questões 15 e 17 foram mapeadas por árvores de similaridade e coesitiva. A classificação hierárquica de similaridade foi utilizada na análise da questão 17, a ASI (mais especificamente a hierárquica Coesitiva) para mapear as regras e metarregras oriundas das relações entre as variáveis das questões 15 e 17. Decidimos de não incluir em nossa análise a questão 16 por considerarmos que as respostas fornecidas não eram consistentes para esse tipo de análise (respostas não discriminantes em função das frequências observadas). Dessa forma, a partir das opções de resposta às questões 15 e 17 (e seus respectivos subitens) geramos 26 variáveis suplementares e 107 variáveis ativas.

3.1 Mapeamento dos dados e análise de mapas gerados por CHIC

3.1.1 Análise de similaridades

Extraímos da árvore de similaridade construída pelo software CHIC a classe que chamamos D, apresentada na Figura 2, para início de nossa análise. Essa classe é

constituída pelas variáveis que se caracterizam pela discordância total ou parcial com alguns itens da questão 17. Observa-se que essa classe é organizada em subclasses, às quais chamaremos de D11 (cinco primeiras variáveis na ordem descendente da figura 2), D21 (cinco variáveis seguintes nessa ordem decrescente) e D22 (três últimas variáveis na ordem descendente da figura 2).

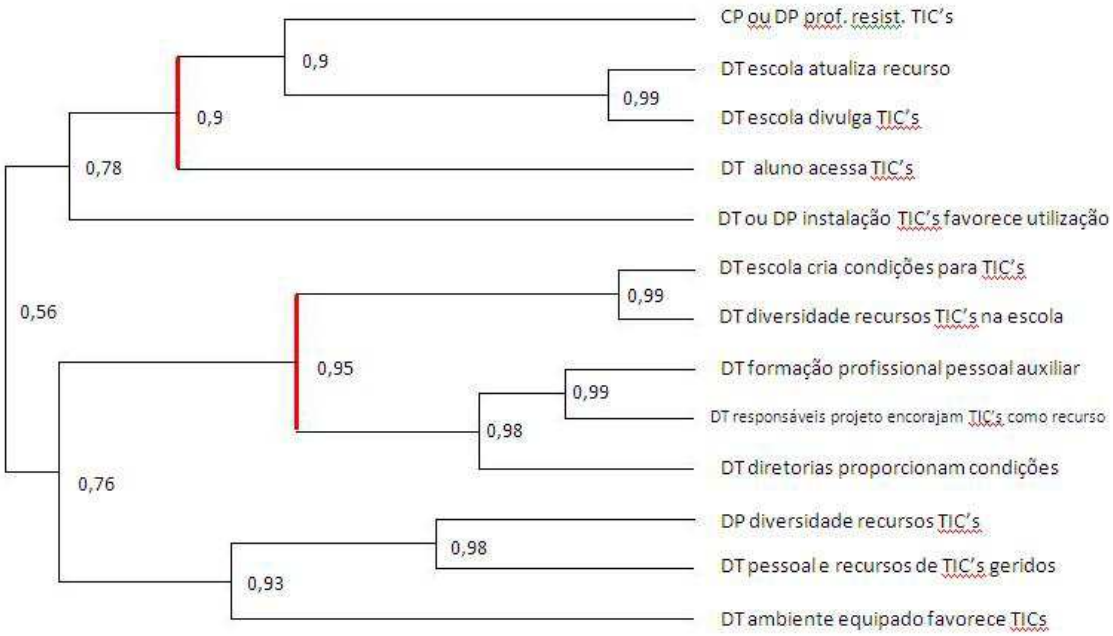


Figura 2 - Árvore de similaridades com as variáveis que constituem a classe D (**Fonte**: autores deste texto)

O software CHIC (Classificação Hierárquica Implicativa e Coesitiva) apresentou a arvore de classificação de similaridade (Figura 2) que subdividimos em subclasses de variáveis. Este tipo de análise permite, ao usuário, estudar e depois interpretar, em termos de tipologia e de semelhança (e dessemelhança) decrescente, classes de variáveis, constituídas, significativamente, em certos níveis da árvore e se opondo a outros, nestes mesmos níveis. Um critério estatístico permite saber quais são **os níveis significativos** da árvore de similaridade entre todos os níveis constituídos. São os níveis em que se formam uma partição e classes que estão mais de acordo com os indícios de similaridade iniciais (ou da hierarquia implicativa). Cada nó significativo está associado à classe obtida nesse nível. As noções de nível e de nós significativos, marcados por uma flecha vermelha, mostra para o usuário as classes que ele deve focar sua atenção, pelo fato de estarem em melhor conformidade com os indícios de similaridade iniciais.

Subclasse D11

Nessa subclasse (Figura 3), a característica principal é a discordância total das questões que se referem à atualização de recursos da escola, à divulgação das TIC para a comunidade escolar e ao acesso do aluno a elas.

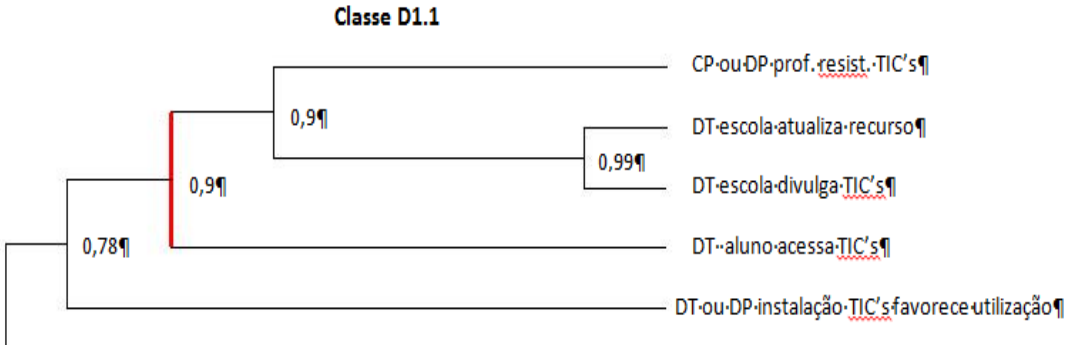


Figura 3 - Árvore de similaridades com as variáveis que constituem a subclasse D11 (**Fonte**: autores deste texto)

O nó significante do agrupamento (assinalado em vermelho) tem similaridade é igual a 0,907, enquanto que no que se refere à essa classe D11, o grau de similaridade é igual a 0,78. Tal comportamento é observado tipicamente por professores que realizaram curso *online*, com risco 0,05, ou seja, essa é a probabilidade de erro ao se fazer tal afirmação sobre a tipicidade. O valor indicado a cada agrupamento apresentado refere-se ao índice de similaridade associado a ele.

Subclasse D2.1

Nessa subclasse a característica é a discordância total das questões que se referem às condições e recursos oferecidos pela escola, à formação dos professores, ao encorajamento dos responsáveis e às condições proporcionadas pela diretoria de ensino para o uso das TIC, conforme figura 4.

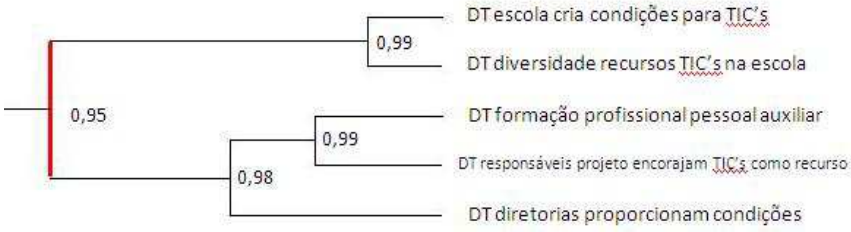


Figura 4 - Árvore de similaridades com as variáveis que constituem a subclasse D2.1 (**Fonte**: autores deste texto)

O grau de similaridade observado nessa subclasse (assinalado em vermelho na figura) é igual a 0,95, e tal comportamento é observado tipicamente por professores do gênero masculino, com risco 0,0185, ou seja, essa é probabilidade de erro ao se fazer tal afirmação sobre a tipicalidade. O valor indicado a cada agrupamento apresentado refere-se ao índice de similaridade associado a ele. Vale observar que alguns sujeitos são típicos do comportamento do conjunto da população no seguinte sentido: no estudo da similaridade (ou da hierarquia implicativa) eles atribuem ao conjunto das variáveis, valores compatíveis com as similaridades (ou a hierarquia implicativa) constituídas sobre essas variáveis, pela população. Se as variáveis suplementares foram definidas pelo usuário, obteremos, assim, a tipicalidade dessas variáveis, a partir das tipicalidades dos indivíduos que as satisfazem.

Subclasse D2.2

Nessa subclasse a característica é a discordância total das questões que se referem às diversidades dos recursos, à gestão de materiais e pessoas e o ambiente para o uso das TIC nas escolas, conforme Figura 5.

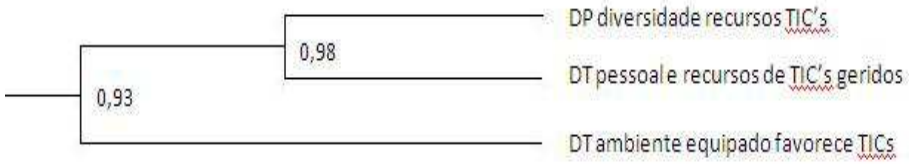


Figura 5 - Árvore de similaridades com as variáveis que constituem a subclasse D2.2 (Fonte: Autores deste texto)

O grau de similaridade observado nessa subclasse é igual a 0,93, e tal comportamento é observado tipicamente por professores do gênero feminino, com risco 0,0543, ou seja, essa é probabilidade de erro ao se fazer tal afirmação sobre a tipicalidade. O valor indicado a cada agrupamento apresentado refere-se ao índice de similaridade associado a ele.

Síntese da classe D

Na subclasse D1.1, verifica-se que os professores que realizaram cursos online pontuam que a responsabilidade pela não utilização das TIC é dos participantes do processo: professores e alunos. Na subclasse D2.1, verifica-se que os professores do gênero masculino responsabilizam a gestão escolar pela não utilização das TIC. Na subclasse D2.2, verifica-se que os professores do gênero feminino justificam a não utilização das TIC devido a um ambiente físico desfavorável. Ou seja, o foco de responsabilidade é alterado de acordo com o grupo típico de respondentes, sugerindo uma perspectiva de busca de correlação entre o grupo típico e o foco de responsabilidade. De forma geral, independentemente dos grupos característicos, a classe D é formada por docentes que responsabilizam os sujeitos (professores e alunos), a gestão e o ambiente físico pelo insucesso no uso das TIC no ambiente escolar.

3.1.2 Análise da hierarquia Coesitiva

Lembramos que, nesta parte, analisamos os mapas gerados por CHIC partir das variáveis estatísticas oriundas das questões 15 e 17. A partir ~~da arvore~~ da hierarquia Coesitiva, isolamos 25 classes de variáveis que passamos a analisar.

Análise da classe α:

Esta classe (Figura 6) é caracterizada por declarações que indicam quase nenhum ou nenhum conhecimento sobre ferramentas tecnológicas para uso em sala de aula. Em uma escala de 1 a 5, na qual 5 indicava o domínio no uso, os respondentes assinalaram 1 para os itens da questão 15 que tinham por objetivo diagnosticar tais conhecimentos.

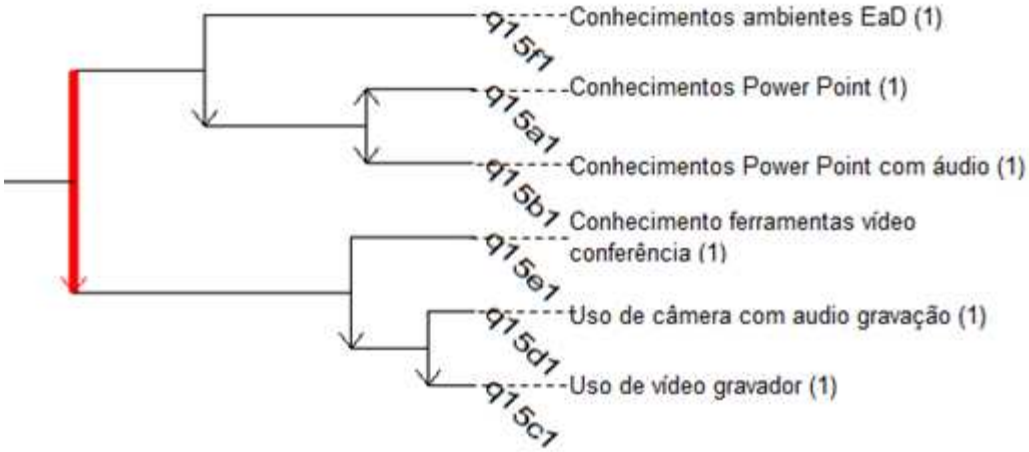


Figura 6 - Falta de conhecimentos sobre TIC e seu uso (Fonte: Autores deste texto)

Destaca-se o não conhecimento pelo uso de Power Point, usualmente utilizados por professores para apresentações as mais diversas. Essa classe evidência também o não uso de ferramentas tecnológicas, como câmara com áudio gravação e vídeo gravador, nas práticas docentes de alguns professores.

Toda essa classe tem coesão de 0,816, sendo que a variável típica é q3b (professores separados ou separados judicialmente), com risco igual a aproximadamente zero.

Análise da classe β

Apesar de CHIC não indicar as classes apresentadas na figura 7 como significativas, elas nos parecem importantes para serem analisadas pelo fato de terem índices de coesão relativamente fortes.

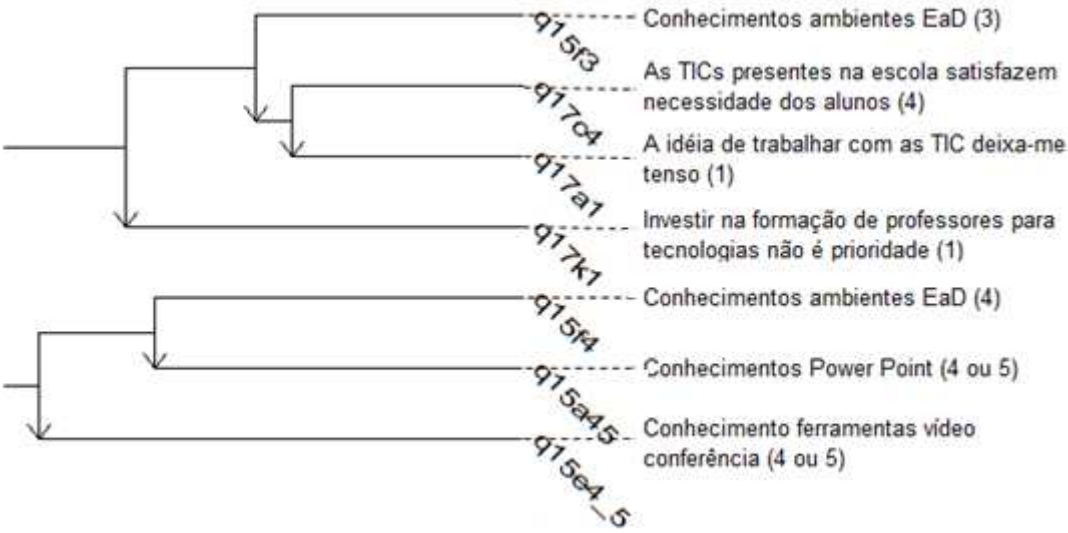


Figura 7 - Bons Conhecimentos sobre alguns aspectos de TIC e com segurança quanto ao seu uso (Fonte: Autores deste texto)

O índice de coesão da primeira classe, que retrata tanto o conhecimento do uso dos AVA (ambientes virtuais de aprendizagem) quanto à predisposição para o trabalho com

as TIC, com as variáveis ((q15f3 (q17o4, q17a1))q17k1), tem índice de coesão 0,8, enquanto que a classe ((q15f4, 15a45)q15e45) tem índice de coesão 0,62. As variáveis típicas são q9c (trabalham de 4 a 6 anos na escola atual) com um risco 0.00942 e q2a (idades entre 21 e 28 anos) com um risco de 0.00306, respectivamente. As duas classes, representadas na figura 7, caracterizam-se pela postura positiva em relação ao uso de tecnologias e à respectiva formação de professores para tal uso. Podemos inferir que a ambientação na escola, propiciada pela permanência, e a idade do professor interferem na sua postura frente ao uso das TIC.

Análise das classes A, B, C:

A classe designada por A (cf. Figura 8) indica que os professores que afirmaram possuir bons conhecimentos sobre ambientes EaD então não concordam ficar nervosos ao ouvir falar de TIC, índice de coesão 0,356.

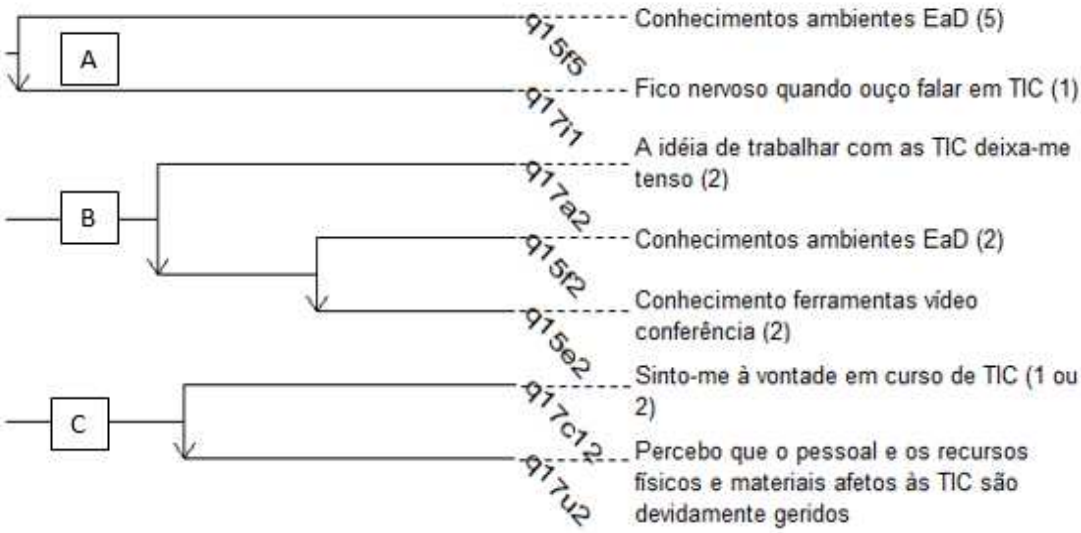


Figura 8 - Poucos conhecimentos sobre TIC e insegurança quando ao seu uso (Fonte: Autores deste texto)

Esse índice indica uma possível contradição nestas respostas dos sujeitos por ser um valor muito baixo (ou seja, indique a possibilidade de que mesmo conhecendo ambientes EaD ainda ficam nervosos com o uso das TIC), mas reflete o que alguns pesquisadores já identificaram em suas pesquisas sobre o tema. É o caso de Costa (2008, p.465), que afirma que

> O facto de a utilização dos computadores estar dependente da confiança que os professores sentem, ou não sentem, na manipulação das tecnologias, é um aspecto relevante para este conjunto de professores. Manifesta-se principalmente com referências alusivas à "sensação de medo", "receio", "insegurança" e "ansiedade" que resultam do facto de não dominarem convenientemente as tecnologias [...]

O autor também aponta que a confiança é um fator decisivo para que os professores decidam utilizar as tecnologias com os seus alunos.

A classe B, também ilustrada na mesma Figura 8, indica que quem discorda que a ideia de trabalhar com as TIC provoca tensão, então possui poucos conhecimentos sobre AVA e sobre ferramentas de vídeo conferência, sendo o índice de coesão da classe igual

a 0,803 e o grupo típico é formado por professores que possuem entre 19 e 30 anos de magistério (q5d), com risco de 0,035. Infere-se uma postura positiva em relação ao trabalho com as TIC, apesar do pouco conhecimento sobre ambientes AVA e sobre ferramentas de vídeo conferência.

A classe C (na mesma Figura 8) indica uma postura pouco favorável ao uso das TIC, embora os respondentes fossem professores participantes de um projeto para formação no uso das TIC como ferramentas para aulas de matemática. Os respondentes afirmam sentir-se pouco à vontade em curso sobre TIC e, então, percebem pouco que o pessoal e os recursos (físicos e materiais) relacionados a esse uso são geridos adequadamente. Essa relação possui índice de coesão igual a 0,825. O grupo de professores que é típico dessa relação é formado por aqueles com idades entre 21 e 28 anos (q2a), com risco 0,035.

Globalmente, essas três classes são caraterizadas por declarações que indicam nenhum conhecimento sobre TIC (ambientes EaD). Essa falta parece gerar nervosismo e desconforto quando se tem uma possibilidade de usar TIC como ferramenta didática. Na subclasse (q15f5→q17i1), a variável típica q2e (professores com mais de 50 anos de idade) com um risco igual a zero. A classe (q17a2 →(q15f2→q15e2) tem por variável típica q5d (tempo de magistério entre 19 e 30 anos) com um risco de 0.035. A variável típica à classe (q17c12→q17u2) é q2a (idade entre 21 e 28 anos) com um risco de 0.035. Podemos, portanto, inferir que a idade do professor interfere em sua postura frente ao uso das TIC.

Sobre a falta proficiência, Miranda (2007, p.44) alerta que diversas pesquisas indicam que "a maioria dos professores considera que os dois principais obstáculos ao uso das tecnologias nas práticas pedagógicas são a falta de recursos e de formação para o uso das tecnologias, mormente as computacionais".

Essa autora afirma que uma das razões da integração das TIC

> [...] prende-se com o facto da integração inovadora das tecnologias exigir um esforço de reflexão e de modificação de concepções e práticas de ensino, que grande parte dos professores não está disponível para fazer. Alterar estes aspectos não é tarefa fácil, pois é necessário esforço, persistência e empenhamento.
> O problema reside em que alguns professores têm uma concepção romântica sobre os processos que determinam a aprendizagem e a construção de conhecimento e concomitantemente do uso das tecnologias no acto de ensinar e aprender. Pensam que é suficiente colocar os computadores com algum software ligados à Internet nas salas de aula que os alunos vão aprender e as práticas se vão alterar. Sabemos que não é assim. (p.40)

Ainda de acordo com Miranda (2007, p.40),

> Os resultados mais conclusivos do imenso esforço de investigação que acompanhou a introdução em grande escala das tecnologias computacionais no ensino (sobretudo a partir dos anos 80) mostram que acrescentar estes recursos às atividades já existentes nas escolas não produz efeitos positivos visíveis na aprendizagem dos alunos, na dinâmica da classe e no empenhamento do professor [...].
> Existem mesmo autores, como Clark (1994), que consideram que os Media Educativos por si só nunca influenciarão o desempenho dos estudantes. Os efeitos positivos só se verificam quando os professores acreditam e se empenham de "corpo e alma" na sua aprendizagem e domínio e desenvolvem

> atividades desafiadoras e criativas, que explorem ao máximo as possibilidades oferecidas pelas tecnologias. E para isto é necessário que os professores as usem com os alunos: a) como novos formalismos para tratar e representar a informação; b) para apoiar os alunos a construir conhecimento significativo; c) para desenvolver projetos, integrando (e não acrescentando) criativamente as novas tecnologias no currículo.

Em síntese, os resultados observados convergem para o já apontado em pesquisas da área, no que se refere à relação entre o pouco conhecimento sobre o uso das ferramentas computacionais e a avaliação sobre sua adequação e sobre os recursos necessários para sua efetiva utilização em sala de aula.

Análise das classes D, E e F

O tratamento dos dados pelo CHIC permite destacar as classes D, E e F (Figura 9), que analisamos a seguir.

Figura 9 - Formação e razões de não uso de TIC (Fonte: Autores deste texto)

A classe D (Figura 9) é formada por dois outros agrupamentos: se concorda que a formação em tecnologia interessa apenas à algumas disciplinas, entre as quais Matemática e Ciências, então também concorda que se sente à vontade em um curso de TIC (nos dois casos, atribui 3 em uma escala crescente de concordância, variando de 1 a 5). O segundo agrupamento dessa classe indica que discorda que a escola cria condições para que os alunos adquiram habilidades com as TIC (atribuição 2 na escala), e concorda (atribuição 3 na escala) que a razão principal para o não uso das tecnologias é a falta de informação e formação. As associações feitas mostram certa coerência dentro das subclasses, mas indicam contradição na classe D, uma vez que o índice de coesão da classe D é igual a 0,438, sendo o da primeira subclasse é igual a 0,908 e da segunda subclasse igual a 0,614. O grupo típico de professores nessa classe D é q3b (separados judicialmente ou divorciados) com risco muito próximo de 0. A classe D traz à tona a necessidade da formação de professores, mais especificamente, segundo nossa análise, os conhecimentos/saberes que deve possuir um professor para a integração das TIC à sua prática docente. Elencamos alguns desses conhecimentos a seguir:

1. Os novos conhecimentos que professores devem ter sobre objetos matemáticos ensinados, do fato que eles são mediados por um ambiente de computação.

2. O comportamento das representações dos objetos matemáticos na tela do computador ou de calculadora não é idêntico ao das representações em outros contextos, em particular papel e lápis. As restrições e a opção de interface, em parte, induzem essas transformações, muitas vezes classificadas sob o termo genérico de processo de transposição informática (Balacheff, 1994).
3. Os conhecimentos dos currículos e dos seus objetivos que permitem que os professores identificar que os possíveis usos de tecnologias podem trazer uma mudança significativa na aprendizagem de alunos.
4. As competências para conceber situações-problema de um novo tipo que dá significado ao uso de TIC. Não se trata que as TIC fossem simplesmente utilizadas como permitindo fazer mais facilmente operações e ações que podem ser feitas à mão, mas que dessem origem a atividades de natureza diferente, de modo que os alunos constroem conhecimentos mais aprofundados em matemática.
5. As competências de gestão de situações novas, organização do tempo, dos tipos de atividade na classe e fora da classe que exigem uso de TIC.

Voltando para análise dos mapas apresentados, percebemos que a classe E (Figura 9) indica que se o professor concorda que a resistência ao uso das TIC deve-se à falta de formação adequada, então discorda que as diretorias de ensino proporcionam as condições para sua utilização didática, com índice de coesão igual a 0,672. O grupo típico é o de quem leciona ou lecionou Matemática entre 7 e 18 anos (q4c), risco 0,0272.

A classe F (Figura 9) indica que se o professor discorda completamente com a afirmação de que a razão principal de não usar tecnologias em aula deve-se mais à falta de informação/formação do que de equipamentos, então discorda que o uso de ferramentas didáticas deve constar de todas as formações de professores. O índice de coesão é igual a 0,045, o que permite inferir a contradição nas respostas. Podemos aqui inferir que, de uma forma geral, esta contradição é confirmada pelas demais associações feitas pelo software CHIC, indicando o desconforto dos professores em relação à utilização das TIC.

Essas três classes de variáveis apresentam várias facetes nas opiniões dos professores investigados, marcadas por contradições em relação à necessidade da formação de professores ou à necessidade de equipar as escolas em TIC. Esses professores parecem ignorar que as formas de organização e relacionamento da sociedade atual exigem cada vez mais novas habilidades e competências. Dentre elas, destacamos, a capacidade de trabalhar em grupo e assumir ações que necessitam de constante aprendizado (BELLONI, 1999).

Análise da classe G

A classe G (Figura 10) tem índice de coesão 0,529 e se caracteriza por discordar da afirmação sobre ficar nervoso ao ouvir falar sobre TIC (q17i2) e então afirmar o pouco conhecimento em relação ao uso de vídeo gravador, de Power Point simples ou com áudio. O grupo de professores mais típico desse agrupamento é formado por aqueles que têm mais de 50 anos de idade, com risco 0,00659.

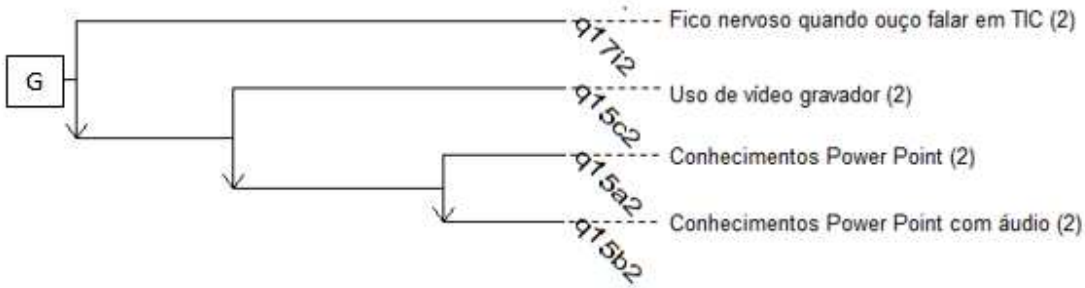


Figura 10 - Pouco uso de Power Point, vídeo gravador e alguma segurança no uso de TIC (Fonte: Autores deste texto)

Análise da classe H (Figura 11)

A classe H (coesão igual a 0,749) é composta por duas subclasses: H1 (coesão igual a 0,908) e H2 (coesão igual a 0,869). A classe H1 se caracteriza por concordâncias 2 e 3, ou seja, posturas de discordância parcial (em uma escala de 1 a 5), em relação ao uso das TIC e das condições de infraestrutura para seu uso. O grupo típico é formado por professores com idades entre 34 e 40 anos, risco 0,000858.

A classe H2 se caracteriza por posturas mais positivas quanto à concordância com os itens apresentados (respostas 3 a 5 na escala) em itens que se referem ao conhecimento sobre ferramentas tecnológicas (vídeo conferência, Power Point, vídeo gravador, áudio gravação) e à algumas condições didáticas para uso das TIC (falta de formação/informação, encorajamento pelos responsáveis, oportunidade de acesso pelos alunos). O grupo de professores mais típico é formado por aqueles que têm entre 4 e 6 anos de tempo de docência em Matemática, com risco 0,000742.

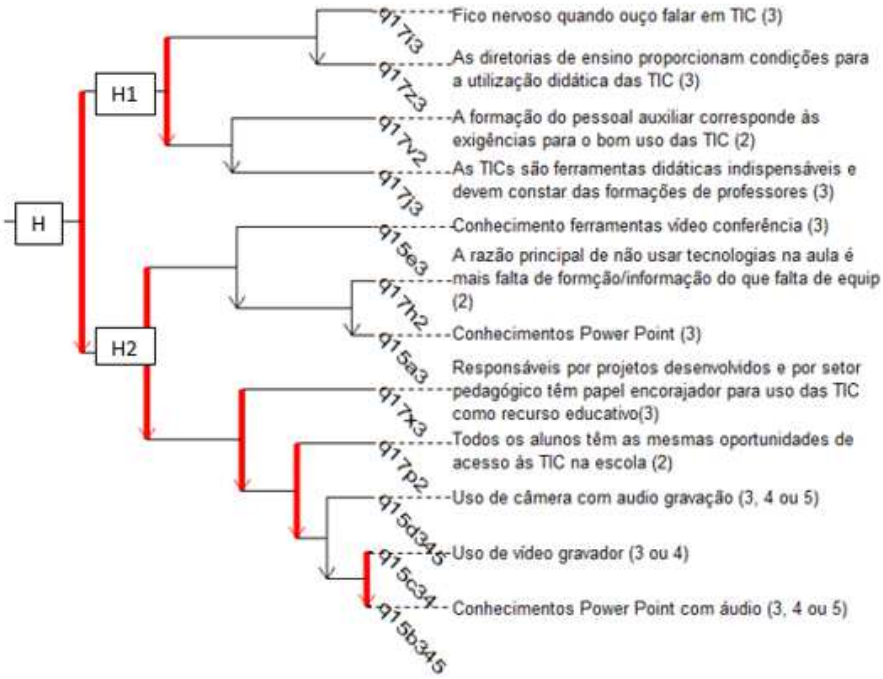


Figura 11 – Classe H (Fonte: Autores deste texto)

A classe H traz a implicação: se H1 então H2, e o grupo típico é o mesmo de H2, com risco 0,00317. Ou seja, se discordância parcial (em uma escala de 1 a 5), em relação ao uso das TIC e das condições de infraestrutura para seu uso, então posturas mais positivas quanto à concordância em relação aos itens que se referem ao conhecimento sobre ferramentas tecnológicas e algumas condições práticas que favorecem o uso de TIC como ferramentas pedagógico-didáticas para ensinar a matemática.

As tecnologias educativas são elementos organizadores da ação formadora e das práticas de aprendizagem. Portanto, o estudo das práticas de ensino midiatizado deve apoiar-se na construção de um modelo global dos processos de ensino, integrando essas tecnologias no centro da dinâmica de ação. A introdução de TIC deve ser apreendida por meio de um olhar global sobre os dispositivos de formação.

Análise da classe J

A classe J (Figura 12) se caracteriza por meta-implicações sucessivas: se faz a opção 2 ou 3 para a afirmação de que trabalhar com tecnologia é divertido e estimulante então opta também em 2 ou 3 para a afirmação sobre o gosto de trabalhar com diferentes tecnologias. Da mesma forma, assinalar a opção 3 para a afirmação de que a formação em tecnologias seja acessível a todos os professores implica nas duas afirmações anteriores.

Figura 12 - Foco na importância na formação e uso divertido e estimulante (Fonte: Autores deste texto)

Se assinala 3 (não concorda nem discorda) para a afirmação segundo a qual o professor gostaria de aprender mais sobre aplicação das TIC na Educação, então valem as opções já citadas. Finalmente, fechando essa classe, se assinala 2 ou 3 para a afirmação de que não é prioritário o investimento na formação de professores para a tecnologia (o que pode indicar uma posição favorável à tal formação), então valem as demais associações citadas. O índice de coesão é de 0,697 e o grupo típico é formado por aqueles que possuem idades entre 41 e 50 anos, risco 0,031.

Análise da classe K

A classe K se caracteriza por concordância com os itens apresentados (condições para uso das TIC, falta de informação e formação para uso das TIC e necessidade de formação de professores nesse campo), assinalando também o pouco conhecimento com ferramentas de áudio gravação.

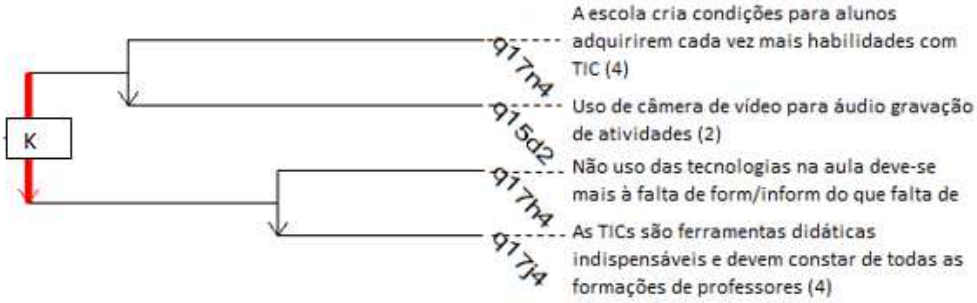


Figura 13 – Classe K (Fonte: Autores deste texto)

O índice de coesão é fraco (0,524), e a variável típica é q4b (leciona ou lecionou Matemática por 4 a 6 anos), com risco aproximadamente zero. Um dos itens avaliado positivamente (valor 4 na escala de 1 a 5) é a necessidade da formação de professores, mais especificamente a necessidade de aprender a trabalhar com TIC. Neste tipo de contexto, Costa (2008) aponta para a exigência, por par parte do professor, de sistematização e articulação entre os aspectos tecnológicos e os aspectos didático-pedagógicos. Esse mesmo autor afirma que os professores precisam ter "um conhecimento aprofundado sobre o conteúdo do software educativo disponível [...] e de procurarem estar informados relativamente à saída de novos recursos para as suas áreas disciplinares específicas" (p.452). Para Ponte (2014, p. 353) o avanço das tecnologias coloca à disposição dos professores um "manancial inesgotável de recursos", mas "falta a capacidade para identificar e selecionar os recursos pretendidos". Afirma ainda que

> Perceber quais as potencialidades das tecnologias que podem ser mobilizadas para contextos formativos e identificar modos de as usar de forma produtiva na formação inicial e contínua, tanto com professores que já usam com muita destreza estas tecnologias, como com professores que mantêm com elas uma relação incipiente, constituem aspectos importantes de uma agenda atual de investigação neste campo. (IBID, p. 354)

Análise da classe L

A classe L (Figura 14) se caracteriza pela discordância total com as afirmações feitas, que formaram duas subclasses, L1 e L2. Em L1 percebe-se a discordância com afirmações sobre as condições de uso das TIC: para este grupo de professores, se as TIC nas escolas não satisfazem a necessidade dos alunos então a escola também não cria condições para o desenvolvimento de habilidades com as TIC.

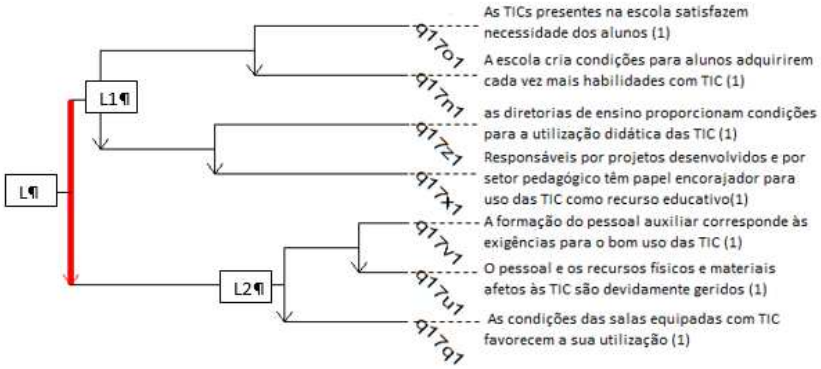


Figura 14 - Descaço da formação e das condições de trabalho dos alunos e professores (Fonte: Autores deste texto)

Na mesma direção, se as diretorias de ensino não proporcionam condições para a utilização didática das TIC, então os responsáveis por projetos e pelo setor pedagógico não têm papel encorajador para o uso das TIC. L1 mostra uma meta-implicação entre os dois grupos citados, com coesão igual a 0,732. L2 (Figura 14) é formada por dois agrupamentos consecutivos. Inicia pela meta-regra "se discorda que a formação do pessoal auxiliar corresponde às exigências para o bom uso das TIC então discorda que o pessoal e os recursos físicos são devidamente geridos", e tal meta-regra implica na discordância com a afirmação de que as condições das salas equipadas com TIC favorecem sua utilização. A coesão nessa subclasse é igual a 0,962. A coesão na classe L (se L1 então L2) é igual a 0,617 e o grupo típico é formado por professores que já fizeram curso *online*, risco 0,0455.

Análise da classe M

A classe M (Figura 15) é caracterizada pela concordância quase total com as afirmações que se referem às condições de trabalho com as TIC, tal como a presença e atualização de recursos na escola e as condições para uso didático desses recursos, com o acesso do aluno (opção 4 em uma escala de 1 a 5), com coesão igual a 0,747 e grupo típico é q4b (entre 4 e 6 anos que lecionam Matemática) com risco tendendo a zero.

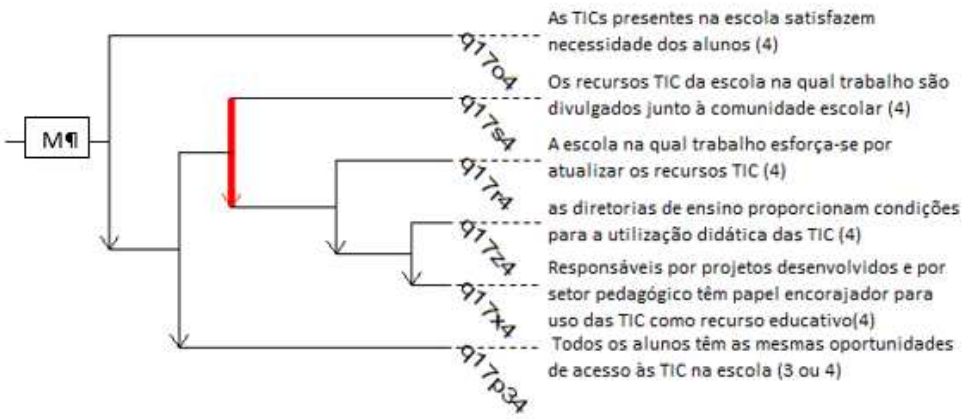


Figura 15 – Classe M (Fonte: Autores deste texto)

Observa-se que as classes L e M apresentam associações análogas, sendo L caracterizada pela discordância e M pela concordância com as afirmações.

Análise da Classe N

A classe N (Figura 16) é formada por associações sucessivas de concordância (escolha da opção 3 em uma escala de 1 a 5): se concorda com as condições de gestão dos recursos de pessoal e físicos, então concorda que a instalação das TIC em laboratórios da escola favorece sua utilização por professores e alunos.

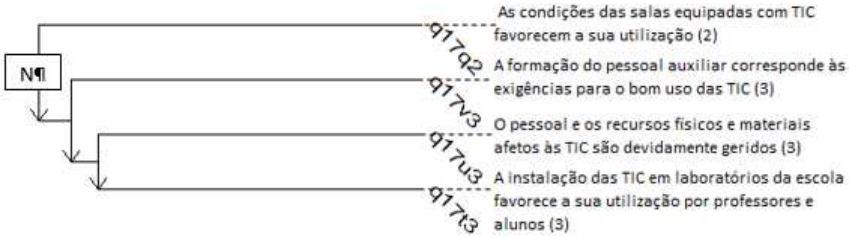


Figura 16 - Contexto escolar favorável para uso de TIC (Fonte: Autores deste texto)

Se concorda que a formação do pessoal auxiliar é adequada para o bom uso das TIC então concorda com as afirmações anteriores. Se concorda que as condições das salas equipadas com TIC favorecem sua utilização então concorda com as afirmações agrupadas anteriormente. O índice de coesão é igual a 0,534 e a variável típica é q5d (tempo total de magistério de 19 a 30 anos), com risco 0,00391.

Análise da classe O

A classe O (Figura 17) se caracteriza pela concordância com as afirmações feitas (opção 4 em uma escala de 1 a 5), tal como na classe N (se q17q então q17v).

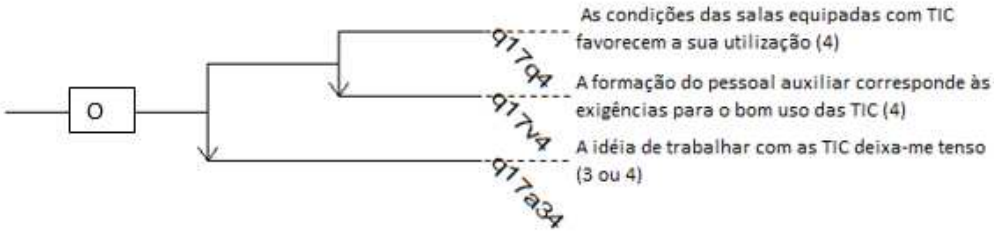


Figura 17 - Condições da escola e formação favoráveis, mas insegurança no uso de TIC (Fonte: Autores deste texto)

No entanto, a afirmação seguinte indica incoerência: se concorda que as salas equipadas favorecem utilização e que o pessoal auxiliar tem formação adequada para seu uso, então concorda que fica tenso com a ideia de sua utilização. O índice de coesão é igual a 0,88 e o grupo típico é formado por aqueles têm até 3 anos de magistério, com risco 0,0291.

Análise da classe P

A classe P (Figura 18) agrupa 11 variáveis, formando duas subclasses, P1 e P2, sendo que P2 ainda se organiza em duas subclasses, P21 e P22, na seguinte configuração: se P21 então P22 (classe P2); se P1 então P2 (classe P). O índice de coesão nessa classe P é igual a 0,739 e grupo típico q5b (4 a 6 anos de magistério), risco tendendo a zero.

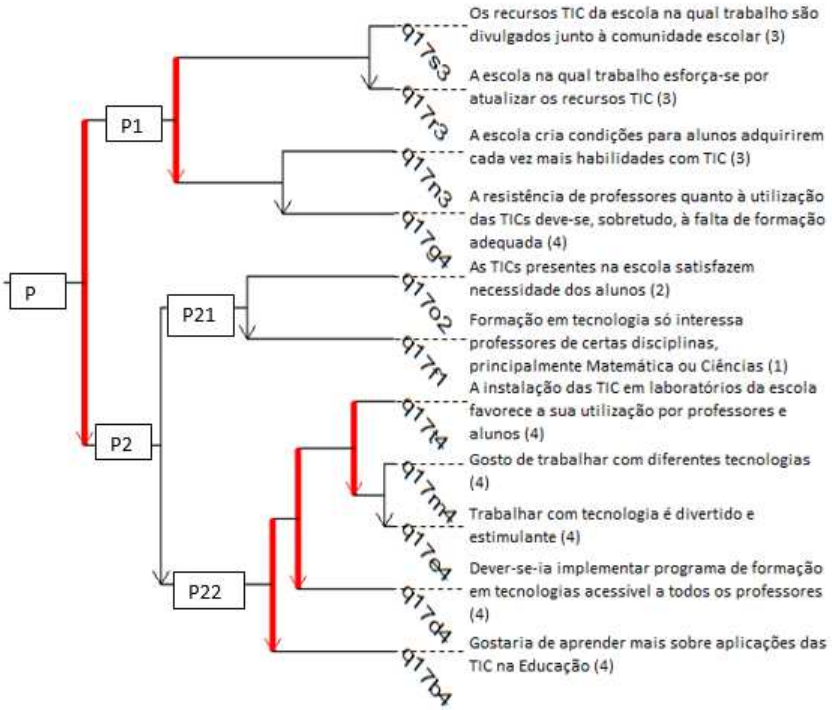


Figura 18 - Classe P (Fonte: Autores deste texto)

Analisando a classe P1: caracteriza-se pela concordância pelo menos parcial (opção 3 em uma escala de 1 a 5) com as condições de uso das TIC – divulgação dos recursos junto à comunidade e atualização desses recursos, implicando na concordância pelo menos parcial (opção 3 ou 4) com as condições para o desenvolvimento de habilidades dos alunos e a falta de formação adequada dos professores. A classe P21 caracteriza-se pela não concordância com as afirmações que as TIC presentes na escola satisfazem as necessidades dos alunos e que a formação em TIC interessa apenas à algumas disciplinas (entre as quais Matemática e Ciências). Se P21, então concordam com as afirmações (opção 4 na escala) que caracterizam conhecimento e predisposição para o trabalho com as TIC (P22): a instalação favorece sua utilização por professores e alunos, gosta de trabalhar com diferentes tecnologias, trabalhar com tecnologia é divertido e estimulante, os professores devem ter acesso à programas de formação e interesse em aprender mais sobre uso das TIC. Tal associação indica coerência na formação da classe P2, que possui índice igual a 0,858 e grupo típico q5b (4 a 6 anos de magistério), risco tendendo a zero.

Análise da classe Q

A classe Q (Figura 19) caracteriza-se pela discordância total com itens referentes à instalação e gestão das TIC na escola, com coesão igual a 0,807 e tipicidade com professores que já realizaram cursos online, risco 0,0455. A cadeia de afirmações que geraram a discordância é: se a escola se esforça para atualizar os recursos tecnológicos, então os recursos são divulgados junto à comunidade escolar.

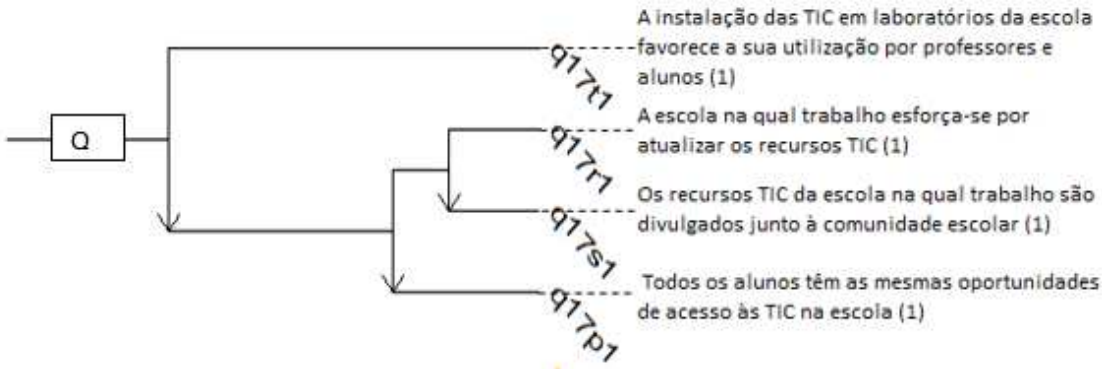


Figura 19 - Opiniões desfavoráveis em relação às condições de trabalho com TIC (Fonte: Autores deste texto)

Se discordaram com as duas afirmações, então também discorda com a afirmação de que todos os alunos têm a mesma oportunidade de acesso às TIC. Finalizando, se discorda com a afirmação de que a instalação das TIC nos laboratórios da escola favorece sua utilização por professores e alunos então também discorda com as afirmações anteriores.

Análise da classe R

A classe R (Figura 20) se caracteriza pela discordância (opção 2 em escala de 1 a 5) com itens referentes às condições de uso das TIC. A coesão é igual a 0,595. Considerando a subclasse composta pelas variáveis q17t2, q17g2, q17x2, q17r2, q17s2, q17o3, o índice de coesão aumenta para 0,781.

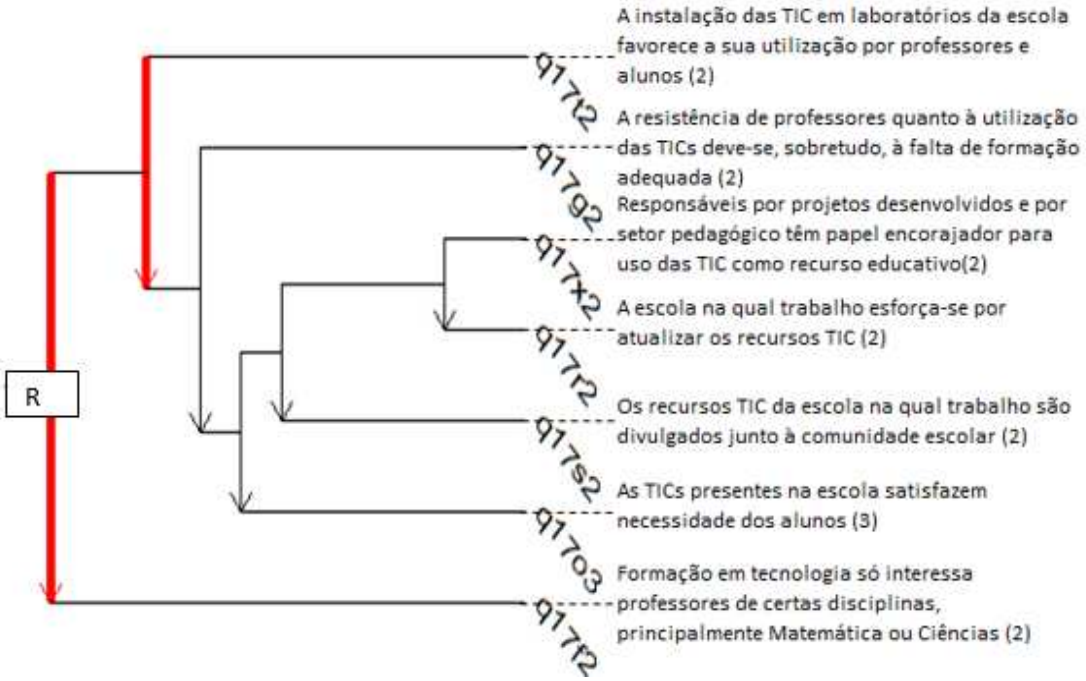


Figura 20 - Classe R (Fonte: Autores deste texto)

Observa-se a coerência nas associações estabelecidas: se os responsáveis pelos projetos não têm papel encorajador para uso das TIC, então a escola não faz esforço para atualizá-las. Se tal associação ocorre, então os recursos não são divulgados para a

comunidade escolar. Considerando-se isso, as TIC não satisfazem às necessidades dos alunos. Se a instalação das TIC nos laboratórios não favorece sua utilização, então ocorrem as associações citadas. O grupo típico dessa associação é formado por professores com tempo de magistério entre 19 e 30 anos (q5d), com risco igual a 0,00124.

Análise da classe S

A classe S (Figura 21) associa apenas duas variáveis: se concorda com a gestão de pessoal e recursos físicos relacionados às TIC, então concorda que as salas equipadas com essas ferramentas favorecem sua utilização.

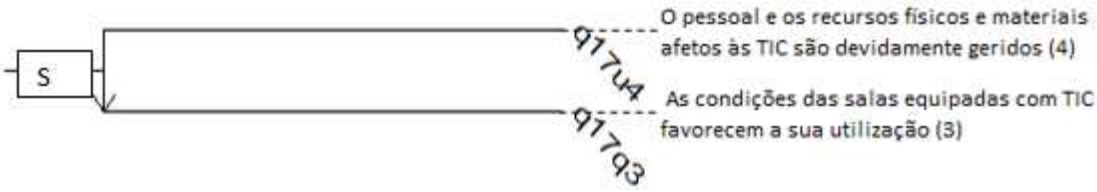


Figura 21 - Classe S (Fonte: Autores deste texto)

Com índice de coesão igual a 0,724, o grupo típico é formado por professores com idades entre 29 e 33 anos (q2b), risco igual a 0,0129.

Retomando os principais resultados dessa análise

Globalmente, nossos resultados vêm ao encontro de algumas das conclusões de Costa (2008, p.456), quando afirma que

> Ao nível das atitudes e embora denotando uma perspectiva favorável relativamente ao uso das tecnologias com os alunos, são diferentes os factores que os professores apresentam para justificar essa atitude favorável. Para alguns isso depende do gosto pessoal pelas tecnologias. Para outros isso depende, sobretudo, de um imperativo profissional relacionado com a atitude face à mudança em geral e com o reconhecimento que um professor, enquanto profissional, tem de fazer novas aprendizagens para acompanhar as mudanças. Para outros ainda, a atitude favorável depende do reconhecimento da necessidade de aprender a trabalhar com os computadores, mas sobretudo da forma como essa aprendizagem é realizada, sublinhando, aliás, a importância que aí assume a realização de experiências concretas e o apoio dado por colegas na aquisição e consolidação da confiança.

As TIC lançam, enormes desafios para os atores do sistema educativo. Sua própria lógica desestabiliza as lógicas de transmissão hierárquica do saber.

A integração das TIC traz a tona problemas para os quais devemos procurar respostas. Entre tantos, citamos os seguintes: A eficácia de novas formas de trabalhar está relacionada ao seu custo em tempo, trabalho, alterações psicológicas? Como avaliar o trabalho coletivo? Qual apoio institucional o professor recebe para facilitar essas adaptações profundas?

O uso das TIC nos processos de aprendizagem provoca mudanças de hábitos do professor no que tange o trabalho e ensino de matemática. O professor passa de um interlocutor único do aluno para um convive com vários interlocutores (sites, equipes de estudantes no mesmo trabalho etc.). A aprendizagem nesses ambientes requer uma mobilização pessoal importante, muito mais do que em um ambiente tradicional.

4 Considerações e perspectivas

O estudo do perfil dos sujeitos da pesquisa, nos permite inferir que muito há para conhecer quanto aos processos de aprendizagem nos ambientes computacionais, principalmente, no âmbito educacional. Por exemplo, *como os processos de aprendizagem se caracterizam nos ambientes computacionais? Quais são as alternativas metodológicas para investigarmos os processos de aprendizagem mediados pela tecnologia?*

De acordo com Kawasaki,

> [..] de que forma novas tecnologias alteram o conhecimento produzido? Para Balacheff e Kaput (1996), estas tecnologias têm contribuído para a produção de um realismo, jamais visto, nos objetos matemáticos e são os recursos interpretativos – nesse caso, a tecnologia computacional – que definem essa nova forma de realismo. Objetos virtuais matemáticos em uma tela de computador nos dão a sensação de sua existência material, dada a possibilidade que temos de manipulá-los dinâmica e continuamente. Assim, para esses autores, o impacto principal da introdução das TICs no sistema educacional é de caráter epistemológico e cognitivo. (KAWASAKI, 2008, p.47)

Concordamos com Kawasaki (2008, p.47), quando afirma que "A atividade humana mediada pelo computador altera de forma qualitativa a estrutura da atividade intelectual humana, reorganizando a memória, as formas com que passamos a armazenar a informação e com que organizamos a sua busca. "

Borba (2011, p.2) aponta que um tópico discutido no debate que gira em torno da questão: "como que um determinado software pode contribuir para que estudantes tentem chegar a uma justificativa matemática e façam a ligação entre a exploração indutiva e o desenvolvimento do raciocínio dedutivo? "

Esse mesmo autor afirma que

> as possibilidades experimentais dessas mídias podem ser exploradas, podendo-se chegar a elaboração de conjecturas bem como a sua verificação. Desse modo, é possível estabelecer uma importante discussão acerca das possibilidades da inclusão de softwares no contexto educacional em seus diferentes níveis. (Borba, 2011, pp2-3).

Completa ainda:

> Os softwares educacionais têm a capacidade de realçar o componente visual da matemática atribuindo um papel importante à visualização na educação matemática, pois ela alcança uma nova dimensão se for considerado o ambiente de aprendizagem com computadores como um particular coletivo pensante (Lévy, 1993), onde professores, alunos, mídia e conteúdos matemáticos residem juntos e, mais que isso, pensam juntos. Neste coletivo a mídia adquire outro status, isto é, vai além de mostrar uma imagem. Mais especificamente, é possível dizer que o software se torna ator no processo de fazer matemática. (BORBA, 2011, p.3)

O software dinâmico permite visualizar relações entre elementos de uma figura ou de um gráfico, propriedades matemáticas, que sem ele, poderiam ser difíceis de ser exploradas na resolução de problemas. Este ponto de vista é corroborado por Borba (2011, p.3), que identificou cinco particularidades do aspecto visual, em educação

matemática, proporcionada pelas tecnologias computacionais. Destacamos os itens apontados por Borba e Villareal, também citados por Borba (2011):

> • Visualização constitui um meio alternativo de acesso ao conhecimento matemático.
> • A compreensão de conceitos matemáticos requer múltiplas representações, e representações visuais podem transformar o entendimento deles.
> • Visualização e parte da atividade matemática e uma maneira de resolver problemas.
> • Tecnologias com poderosas interfaces visuais estão presentes nas escolas, e a sua utilização para o ensino e aprendizagem da matemática exige a compreensão dos processos visuais.
> • Se o conteúdo de matemática pode mudar devido aos computadores, (...) e claro neste ponto que a matemática nas escolas passarão por pelo menos algum tipo de mudança (...) (BORBA & VILLAREAL, 2005, apud BORBA, 2011).

Quando pensamos em formação continuada de professores fazendo uso de TIC, duas ideias distintas surgem: formar um docente para o uso de novas tecnologias e formar-se por meio das novas tecnologias. Apesar de intrinsecamente relacionados, neste estudo, estamos interessados na formação de professores por meio das novas tecnologias. Para Ponte (2014, p. 356) desenvolver dispositivos de formação implica também em perceber "o que está e o que não está ao seu alcance", pois há aprendizagens que ocorrem em um dia, mas outros que requerem meses ou anos de trabalho.

O grupo de professores participantes do processo de formação continuada desenvolvido no seio deste projeto de pesquisa foi bem retratado quanto aos conhecimentos sobre as TIC, quanto à sua predisposição para o uso dessas TIC, sempre considerando seu discurso sobre esses conhecimentos. Outra fase do projeto permitiu a observação destes professores em situação de resolução de problemas no campo da Geometria, com utilização do software Geogebra, e pudemos constatar o pouco domínio da ferramenta, agravado pelo pouco domínio de conteúdo.

As análises realizadas e aqui apresentadas nos permitem responder, ao menos parcialmente, nossas questões de pesquisa: *quais fatores influenciam o processo de ensino e de aprendizagem da Geometria com a utilização de recursos alternativos como ambientes computacionais? Como os processos de aprendizagem se caracterizam nos ambientes computacionais? Quais são as alternativas metodológicas para investigarmos os processos de aprendizagem nesses ambientes?*

No que se refere à primeira questão, podemos inferir que para os professores participantes, o grau de familiaridade com os recursos tecnológicos é fator importante para sua utilização pelos professores, de onde a necessidade de um trabalho para formação, atualização das TIC nas escolas, o que geraria oportunidade para que todos os alunos tivessem acesso a tais ferramentas.

Os dados obtidos por meio desse questionário não nos permitem responder à segunda questão, mas podemos inferir que no que se refere à terceira questão, uma alternativa metodológica que podemos visar é a formação para o uso das TIC com utilização de grupos colaborativos, pois a interação com outros professores pode favorecer a superação de preconcepções ou pré-conceitos sobre as TIC.

Referências

[1] Almeida, M. E. B.(2008) de. Mapeando percepções de docentes no CHIC para análise da prática pedagógica. In: OKADA, A. (Org.). *Cartografia cognitiva: mapas do conhecimento para pesquisa, aprendizagem e formação docente*. Cuiabá: KCM. p. 331-341.

[2] Almouloud, S. (2008) Análise e mapeamento estatístico de fenômenos didáticos com CHIC , in Okada (org.), *Cartografia cognitiva: mapas do conhecimento para pesquisa, aprendizagem e formação docente*. Cuiabá: KCM editora, p.301-324,

[3] Balacheff, N.(1994). Didactique et intelligence artificiel. *Recherches en didactique des mathématiques*. Grenoble: La Pensée Sauvage, VOL. 14/1.2, P.9-42,

[4] Belloni, M. L. (1999) *Educação a Distância*. Campinas, SP: Associados.

[5] Borba, M. de C. (2001) *Educação matemática a distância online: balanço e perspectivas.* (In http://www.rc.unesp.br/gpimem/downloads/artigos/borba/xiiiciem-edmatonline-balepersp.pdf , acessado em 21/08/2012)

[6] Costa, F. A. A. (2008) *A utilização das TIC em contexto educativo. Representações e práticas de professores*, Doutoramento em ciências da educação Desenvolvimento Curricular e Avaliação em Educação. Universidade de Lisboa Faculdade de Psicologia e de Ciências da Educação, 2008 (in http://aprendercom.org/comtic/wp-content/uploads/2013/01/TeseCostaF2008TICemContextoEducativo.pdf, acesso, 14/06/2015)

[7] Gras, R. (2014) Genèse et développement de l'analyse statistique implicative : rétrospective historique. *Revista de Educação Matemática Pesquisa*, v. 14.3, p. 645-661,

[8] Kawasaki, T. F.(2008) *Tecnologias na sala de aula de matemática: resistência e mudanças na formação continuada de professores*. Tese de doutorado em Educação Matemática – UNESP, (in http://www.bibliotecadigital.ufmg.br/dspace/bitstream/1843/FAEC-84XH59/1/teresinhakawasakitese.pdf, acessado em 21/08/2012)

[9] Miranda G. L. (2007). Limites e possibilidades das TIC na educação. *Revista de Ciências da Educação*, 03, pp. 41-50, 2007. (in http://portaldoprofessor.mec.gov.br/storage/materiais/0000012617.pdf, acesso, 12/06/2015)

[10] Pacheco, J., & Flores, M. (1999). *Formação e avaliação de professores*. Porto: Porto Editora.

[11] Ponte, J. P. (2014) Formação do professor de Matemática: perspectivas atuais. In: Ponte, J. P. (org). *Práticas Profissionais dos Professores de Matemática*. 1.ed., 2014, p. 343-360.

[12] Showers, B., Joyce, B., Bennett, B. (1987). Synthesis of research on staff development: *A framework for future study and a state-of-the-art analysis. Educational Leadership*, 45 (3), 1987, p. 77–87.

[13] Trouche, L. (2005) Construction et conduite des instruments dans les apprentissages mathématiques : nécessité des orchestrations. *Recherches en Didactique des Mathématiques. v*.25/1, n.73, p. 91-138.

Apêndice (Variáveis com suas legendas)

VARIÁVEIS SUPLEMENTARES	código
1) Gênero	
Masculino	q1M
Feminino	q1f
2) Idade	
21 a 28 anos	q2a
29 a 33 anos	q2b
34 a 40 anos	q2c
41 a 50 anos	q2d
mais de 50	q2e
3) Estado civil	
Casado(a) ou em união estável	q3a
Separado(a) judicialmente ou divorciado	q3b
Solteiro(a)	q3c
4) Tempo que leciona, ou lecionou, Matemática	
Até 3 anos	q4a
4 a 6 anos	q4b
7 a 18 anos	q4c
19 a 30 anos	q4d
5) Tempo total de magistério	
Até 3 anos	q5a
4 a 6 anos	q5b
7 a 18 anos	q5c
19 a 30 anos	q5d
6) a) Informações sobre sua formação no Ensino Fundamental	
6) b) Informações sobre sua formação no Ensino Médio	
6) c) Informações sobre sua formação na Graduação	
7) Pós graduação	
8) Há quanto tempo você é professor efetivo da rede pública?	
9) Há quanto tempo você trabalha na escola atual?	
Menos de 1 ano	q9a
1 a 3 anos	q9b
4 a 6 anos	q9c
7 a 18 anos	q9d
19 a 30 anos	q9e
Pergunta sem título	
Carga horária semanal nos últimos três anos	
10) Qual sua carga horária semanal na escola neste ano? Na rede pública e/ou na rede privada	
11) Qual sua disponibilidade para participar da parte do projeto que se desenvolverá via Internet?	
Somente em casa	q11a
Somente na escola	q11b
Em casa e na escola	q11c
Nem em casa, nem na escola	q11d
12) Indique os programas de computador com os quais tem experiência.	
13) Em geral, como você classifica seu domínio no uso de computadores?	
Principiante	q13a
Bom utilizador	q13b
Pouco experiente	q13c
14) Já fez curso online?	
sim	q14a
não	q14b

VARIÁVEIS ATIVAS

15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [Tipo]				
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [a) Elaboração de apresentações em Power Point]	q15a1	q15a2	q15a3	q15a4
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [b) Associação de apresentações em Power Point com áudio]	q15b1	q15b2	q15b3	q15b4
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [c) Uso de videogravador - gravação de imagens da tela do computador]	q15c1	q15c2	q15c3	q15c4
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [d) Uso de câmera de vídeo para áudio gravação de atividades]	q15d1	q15d2	q15d3	q15d4
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [e) Computador com uso de ferramentas de vídeo conferência (Skype/MSN/GMail/Facebook)]	q15e1	q15e2	q15e3	q15e4
15) Como classificaria os seus conhecimentos como utilizador de Tecnologias de Informação e Comunicação, em uma escala de 1 a 5? [f) Computador com uso de ferramentas para aprendizagem a distância (ambientes do tipo Moodle)]	q15f1	q15f2	q15f3	q15f4
16) Os conhecimentos que possui foram obtidos				
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [A idéia de trabalhar com as TIC deixa-me tenso.]	q17a1	q17a2	q17a3	q17a4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Gostaria de aprender mais sobre aplicações das TIC na Educação.]	q17b1	q17b2	q17b3	q17b4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Sinto-me à vontade em curso de TIC]	q17c1	q17c2	q17c3	q17c4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Dever-se-ia implementar um programa de formação em tecnologias acessível a todos os professores]	q17d1	q17d2	q17d3	q17d4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Acredito que trabalhar com tecnologia é divertido e estimulante]	q17e1	q17e2	q17e3	q17e4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso que a formação em tecnologia só poderá interessar para professores de certas disciplinas, principalmente Matemática ou Ciências]	q17f1	q17f2	q17f3	q17f4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso que a resistência de muitos professores quanto à utilização das tecnologias na sala de aula deve-se, sobretudo, à falta de formação adequada]	q17g1	q17g2	q17g3	q17g4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso que a razão principal para os professores não utilizarem as tecnologias na aula está mais na falta de formação e informação do que na falta de equipamento]	q17h1	q17h2	q17h3	q17h4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Fico nervoso(a) quando ouço falar em TIC]	q17i1	q17i2	q17i3	q17i4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso que as tecnologias são ferramentas didáticas indispensáveis à atividade docente e devem constar de todos os planos de formação de professores]	q17j1	q17j2	q17j3	q17j4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso não ser prioritário investir na formação de professores na área das tecnologias.]	q17k1	q17k2	q17k3	q17k4

17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Gosto de trabalhar com diferentes tecnologias]	q17m1	q17m2	q17m3	q17m4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Penso que a escola cria condições para que os alunos adquiram cada vez mais habilidades no domínio das TIC]	q17n1	q17n2	q17n3	q17n4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que a diversidade de recursos TIC existente na escola vai ao encontro das necessidades de professores e alunos]	q17o1	q17o2	q17o3	q17o4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que todos os alunos têm as mesmas oportunidades de acesso às TIC na escola]	q17p1	q17p2	q17p3	q17p4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que as condições ambientais (iluminação, climatização, distribuição de espaços) das salas equipadas com TIC favorecem a sua utilização]	q17q1	q17q2	q17q3	q17q4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que a escola na qual trabalho esforça-se por atualizar os recursos TIC.]	q17r1	q17r2	q17r3	q17r4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que os recursos TIC da escola na qual trabalho são divulgados junto à comunidade escolar]	q17s1	q17s2	q17s3	q17s4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que a instalação das TIC em laboratórios da escola favorece a sua utilização por professores e alunos]	q17t1	q17t2	q17t3	q17t4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que o pessoal e os recursos físicos e materiais afetos às TIC são devidamente geridos]	q17u1	q17u2	q17u3	q17u4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que a formação profissional do pessoal auxiliar corresponde às exigências para o bom uso das TIC.]	q17v1	q17v2	q17v3	q17v4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que os responsáveis pelos projetos desenvolvidos na escola e pelo setor pedagógico têm um papel encorajador junto aos colegas para usarem as TIC como recurso educativo]	q17x1	q17x2	q17x3	q17x4
17) Você encontrará abaixo uma série de afirmações que se referem ao tema TIC e, ao lado, uma escala que estabelece a seguinte escala de concordância: [Percebo que as diretorias de ensino proporcionam condições para a utilização didática das TIC]	q17z1	q17z2	q17z3	q17z4

DIVERS MOTIFS DANS LES REPRÉSENTATIONS DE L'ASCENSION DU CHRIST EN OCCIDENT ENTRE LE IXe ET LE XIIIe SIECLE. ÉTUDE DE LIENS QUASI IMPLICATIFS

Magali GUENOT[1], Jean-Claude REGNIER[2]

SEVERAL MOTIVES IN THE PICTURES OF THE ASCENSION OF CHRIST IN OCCIDENT BETWEEN THE 9TH AND THE 13TH CENTURY. STUDY OF QUASI IMPLICATIVE LINKS.

RÉSUMÉ

L'analyse sérielle en histoire de l'art porte sur une série d'images ayant un sujet, une technique, ou une production commune. Elle a pour objectif de déterminer les permanences, mais aussi les écarts d'une même série, en utilisant les statistiques. Dans le cadre de l'iconographie de l'Ascension du Christ en Occident entre le IXe et le XIIIe siècle, plusieurs motifs sont apparus : il s'agit désormais de vérifier s'il existe des liens de cause à effet entre eux, puis, en fonction des résultats, d'en proposer une interprétation. L'analyse statistique implicative est alors utilisée pour observer les connexions existant entre les motifs, pour en faciliter ensuite la lecture.

Mots-clés : *iconographie, Moyen Âge, statistique implicative.*

ABSTRACT

In art history, serial analysis consists on the study of a series of pictures which share the same subject, technic or production. Its main objective is to determine the permanences and the differences within the same series, by using statistics. As part of a work about the iconography of the Ascension of Christ in Occident between the 9th and the 13th century, many motives appeared, between which we have to check causal relationship and then, according to the results, to propose an interpretation. Statistic implicative analysis is used to observe connections between motives, to ease the reading.

Keywords : *iconography, Middle Age, implicative statistic.*

1 Introduction

L'analyse des représentations de l'Ascension du Christ[3] en Occident entre le IXe et le XIIIe siècle recense quelques 320 œuvres figurant le sujet, sur des supports divers, comme une lettre historiée de quelques centimètres carrés ou un portail d'église de plusieurs mètres carrés. Les images s'organisent souvent de la même manière : le Christ s'élève, parfois dans une mandorle[4], sous le regard des témoins. Des anges peuvent assister à la scène. Les représentations du Christ avaient déjà fait l'objet d'un premier article, qui visait à dresser un portrait global de l'iconographie du Christ dans les

[1] École doctorale 483 Sciences Sociales, Université Lumière Lyon 2, 86 rue Pasteur, 69007 Lyon ; magali.guenot@gmail.com

[2] Université de Lyon – Lyon2 UMR 5191 ICAR ; jean-claude.regnier@univ-lyon2.fr

[3] L'Ascension du Christ est un épisode biblique, qui relate la montée au ciel du Christ quarante jours après la Résurrection. La référence principale se lit dans les *Actes des apôtres* 1, 2-9.

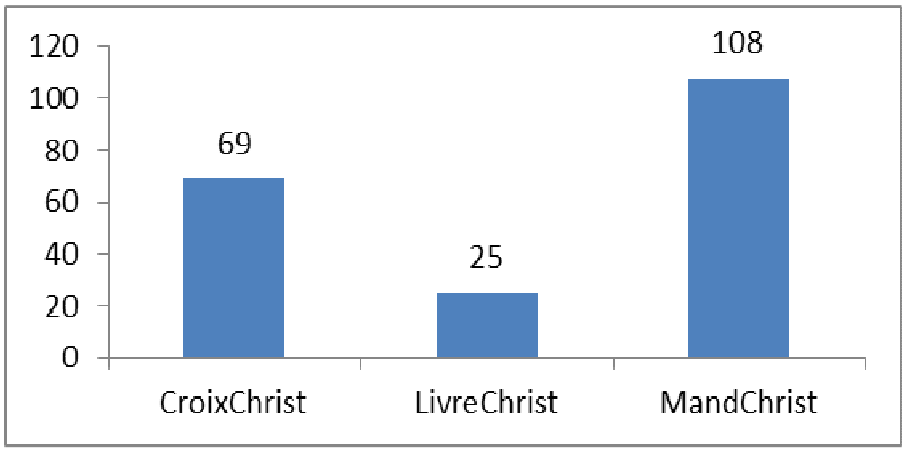
[4] Forme d'amande qui entoure parfois le Christ, dès lors qu'il monte et siège au ciel.

représentations de l'Ascension, en ayant recours à l'analyse statistique (Guénot, Régnier, 2009). Plus tard, nous devions aller plus avant en nous penchant plus spécifiquement sur les répercussions que pouvaient avoir la mise en image des motifs dans la structure iconique de l'image (Guénot, Régnier, 2010). Il était apparu que la quasi-implication au sens de (Gras & al, 2009, 2013) de certains motifs induisait une représentation particulière, de laquelle découlait une interprétation spécifique de l'image. Devant les résultats encourageants de ces études, nous tentions d'étudier des motifs rarissimes en employant non plus le graphe implicatif, mais l'arbre des similarités (Guénot, Régnier, 2013).

Notre intérêt se porte désormais sur les motifs spécifiques, plus ou moins récurrents. Ainsi, le Christ peut porter une croix ou un livre, ou être ceint d'une mandorle. De même, parmi les témoins, certains peuvent être identifiables : la Vierge, saint Pierre, saint Paul. Enfin, certains motifs sont plus ou moins fréquents, comme la présence d'un objet à l'aplomb du Christ, sur le mont des Oliviers. Ces motifs apparaissent-ils de manière indépendante, ou sont-ils en lien avec les autres ? C'est alors qu'intervient l'analyse statistique implicative, qui permet de vérifier la présence ou l'absence de connexion entre les variables et donc, dans le cas présent, entre les motifs. À travers trois études de cas s'intéressant aux attributs du Christ, à l'identification et la relation des témoins entre eux, et à la représentation de l'objet à l'aplomb du Christ, les chemins observés sont-ils conformes aux attentes, et vers quels résultats orientent-ils la réflexion ?

2 . Attributs du Christ

Le Christ est régulièrement entouré d'une mandorle, et porte une croix à hampe ou un support écrit (livre surtout, phylactère ou *rotulus*). Ces motifs ne sont pas systématiques, et connaissent une faveur diverse selon les périodes et les régions.



Graphique 1 : Répartition des attributs du Christ.

Fig. 1 : Psautier d'Odbert, 990-1007. © Boulogne, Bibliothèque municipale

Fig. 2 : Évangéliaire, 1100-1124. © Londres, British Library

Enveloppant le Christ montant au ciel, la mandorle est représentée sur un tiers des images. En forme d'amande le plus souvent, circulaire parfois, elle peut être figurée seule autour du Christ. Elle peut aussi être portée par deux anges, voire quatre (Fig. 1). Elle n'est pas toujours complètement fermée, comme sur l'évangéliaire réalisé en Souabe (Fig. 2), qui représente la partie inférieure de la mandorle ouverte.

Fig. 3 : Montceaux-l'Étoile, église Saint-Pierre, 12e siècle © Magali Guénot

Fig. 4 : Sacramentaire d'Ottobeuren, vers 1175. © Londres, British Library

Sur près de 22% des images, le Christ monte au ciel avec une croix. Il la présente, comme sur le portail de Montceaux-l'Étoile (Fig. 3), ou peut la porter sur son épaule pendant qu'il monte au ciel. La croix peut aussi être à bannière, comme sur le sacramentaire d'Ottobeuren (Fig. 4). Enfin, sur une plaque d'émail, le Christ porte un globe surmonté d'une croix.

La croix comme le livre valorisent un des aspects de l'Ascension : par la croix, la voie du Salut est à nouveau ouverte aux hommes tandis que le livre montre que le Christ détient seul le mystère des Écritures, qu'il emporte avec lui au ciel.

Fig. 5 : Bible en images de Sancho VII, 1197. © Amiens, Bibliothèque municipale

Fig. 6 : Missel de Stammheim, vers 1170. © Malibu, Jean-Paul Getty's museum

Fig. 7 : Sacramentaire de Cologne, 983-996. © Paris, Bibliothèque nationale de France

Les représentations où le Christ monte avec un support écrit dépassent à peine 8%. En majorité, il s'agit d'un livre, qu'il tient ouvert, ou fermé (Fig. 5). Le fils de Dieu déroule un phylactère sur le missel de Stammheim (Fig. 6), mais il peut aussi tenir un rotulus (Fig. 7). Nous souhaitons dès lors vérifier si la croix et le livre sont figurés ensemble sur les images, et s'il existe un lien entre leur présence et la position, la présentation et l'action du Christ (Fig. 8).

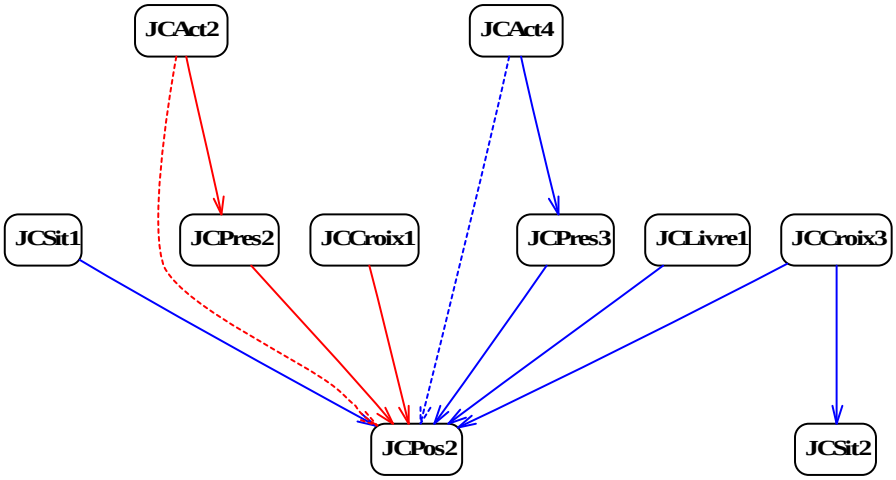


Fig. 8 : Graphe implicatif concernant le livre, la croix, la position, la présentation, l'action et la situation du Christ (1-α = 0,99 et 0,95). Le livre, la croix tenue par le Christ et la croix à bannière impliquent, lorsqu'ils sont figurés, la représentation en pieds du Christ.

La position en pieds (JCPos2) apparaît comme le nœud fédérateur : hormis la situation du Christ encore sur terre (JCSit2), tous les nœuds convergent vers la position en pieds du Christ. Ainsi, la croix et le livre sont représentés quand le Christ est en pieds. En revanche, à un niveau de quasi implication de 0,95, aucun attribut n'est lié à la présentation de profil ou de trois quarts du Christ, ni à son action de monter ou de

gravir, alors qu'au moins pour la croix, nous nous attendions à ce que cet attribut soit rattaché à ce type de présentation et d'action, d'après nos observations.

Nous décidons d'ajouter la variable « mandorle » au graphe implicatif précédent, en laissant l'indice de quasi implication à 0,99.

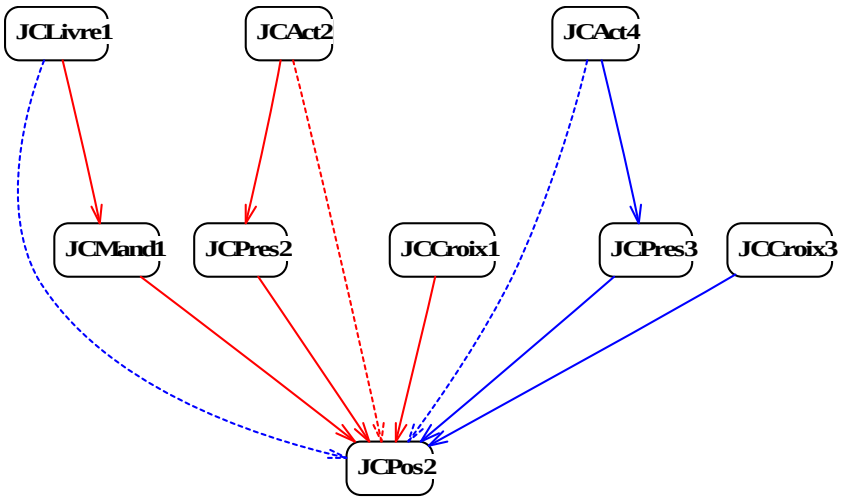


Fig. 9 : Graphe implicatif mandorle-livre-croix-présentation/action/position/situation du Christ (1-α=0,99 et 0,95). Le Christ porte un livre essentiellement quand il est dans une mandorle. La mandorle portée par dans anges implique la représentation du Christ dans le ciel.

Fig. 10 : Évangéliaire de Prüm, 1100-1133.
© Paris, Bibliothèque nationale de France

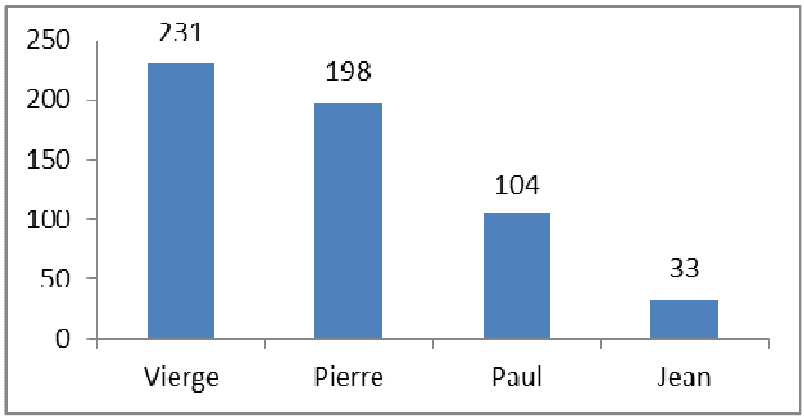
Lorsque le Christ porte un livre (JCLivre1), il y a de fortes chances qu'il soit dans une mandorle (JCMand1) ; mais aucun chemin n'existe entre la croix portée par le Christ et la mandorle. Lorsque la mandorle est portée par des anges (JCMand2), le Christ est dans le ciel (JCSit1) (Fig. 9). Pour autant, dans ce dernier cas de figure, la mandorle portée par des anges n'induit pas que le Christ est essentiellement figuré en pied (JCPos2). On remarque aussi qu'à ce degré d'implication maximum, le Christ n'est

pas principalement figuré de face quand il est dans la mandorle, même en abaissant le niveau d'implication à 0,9. Cela suppose premièrement que le Christ de face n'est pas toujours représenté en mandorle, loin s'en faut. La position de face suffirait alors à manifester sa divinité. Deuxièmement, la mandorle peut aussi entourer le Christ lorsqu'il est de profil ou de trois quart, comme sur l'évangéliaire de Prüm (Fig. 10). Elle deviendrait alors une manière de manifester le fait que le Christ s'élève seul, sans aide extérieure, et assurerait le rôle de moyen de transport dont parle François Boespflug. Et elle pourrait aussi témoigner de la double nature du Christ : de profil ou de trois quart, le Fils de Dieu est dans une action humaine. La mandorle le caractérise uniquement dans les manifestations de sa double nature : elle serait ici une expression de la nature également divine du Christ. De fait, que la présence de la croix n'implique pas celle de la mandorle peut orienter la réflexion sur la mise en avant de la nature humaine du Christ lorsqu'il porte la croix : en effet, si la mandorle met l'accent sur la nature divine du Christ, la présence de la croix, instrument de la mort du Christ, insisterait sur la nature humaine du Christ. Aussi, leur présence simultanée dans les images en brouillerait la lecture, ce qui expliquerait l'absence de lien entre les deux.

Alors que nous nous attendions à voir des liens forts entre les attributs entre eux, mais aussi avec le Christ, l'analyse statistique implicative a mis en avant le caractère mouvant de ces motifs.

3 Les témoins comme représentation de l'Église

Grâce aux conventions iconographiques utilisées, il est aisé de reconnaître les personnages phares figurés lors de l'Ascension. La Vierge, saint Pierre, saint Paul et, dans une moindre mesure, saint Jean, sont souvent reconnaissables. La surprise vient surtout de la figuration de la Vierge, dont les textes canoniques ne font pas cas, et saint Paul, converti plus tard. La répartition du graphique ci-dessous indique que leur présence n'est pas anecdotique :



Graphique 2 : Identification des témoins principaux

Fig. 11 : Anzy-le-Duc, 1120-1124 © Magali Guénot

Fig. 12 : manuscrit réalisé à Helmarshausen, vers 1160. © Gniezno, archives du diocèse

Voilée, vêtue, la Vierge apparaît 231 fois, soit sur plus de 72% des représentations, alors qu'elle n'est pas mentionnée dans les *Actes des apôtres*, et que les textes exégétiques se gardent de l'y associer. D'un point de vue strictement numéraire, elle est surtout figurée sur les lettres historiées (84/321). Or, si l'on s'attache à la proportion, elle est représentée sur près de 71% des lettres, mais sur 90% des pleines pages. De même, quelle que soit la période prise en compte, 72 à 75% des œuvres la font assister à l'Ascension.

Elle peut être située à l'aplomb du Christ, comme sur le portail d'Anzy-le-Duc, où elle croise les mains sur la poitrine, entourée des témoins dont saint Pierre, à sa droite, qui porte les clefs du Paradis. Sinon, elle est à la tête d'un groupe de témoins, si possible à la droite du Christ, à l'instar du manuscrit conservé à Gniezno (Fig. 12), où ses gestes restent mesurés. De rares exemples la relèguent au second plan, dans une posture exubérante, comme le portail de Montceaux-l'Étoile, où, derrière saint Jean, elle lève exagérément le bras en direction de son fils (Fig. 3).

Fig. 13 : Lectionnaire de Trond, 1165-1180. © New York, Pierpont Morgan Library

Fig. 14 : Bible, vers 1240-1260. ©Besançon, Bibliothèque municipale

Porteur de clefs, barbe courte, Pierre est identifiable avec certitude 198 fois, soit sur près de 62% des représentations : sa personnification est donc moins systématique que celle de la Vierge. Si la mère du Christ est au centre, il est situé le plus souvent à sa droite, comme le montre le lectionnaire de Trond (Fig. 13). Si la Vierge est représentée à droite du Christ, saint Pierre sera à sa gauche. Et s'il s'agit d'une image où la Vierge est absente, le prince des apôtres sera alors le plus souvent à droite du Christ, avec saint Paul en pendant, comme sur la lettre historiée de la Bible conservée à Besançon (Fig. 14).

Un tiers des images identifie saint Paul parmi les témoins, alors que sa conversion est postérieure[5]. Il est reconnaissable à sa barbe longue et sa calvitie. Le plus souvent, il est à la tête d'un groupe de témoins, à gauche du Christ, en pendant de saint Pierre. La Vierge peut être un motif séparateur : la pleine page du sacramentaire d'Ottobeuren la figure au centre, entourée des deux princes des apôtres (Fig. 4).

Fig. 15 : Psautier, 1230-1255. © New York, Pierpont Morgan Library

Enfin, un peu plus de 10% des images identifient saint Jean, imberbe, parfois un livre à la main, comme le psautier de Trèves (Fig. 15) : il est représenté derrière saint Pierre, à l'extrême gauche de l'image. Il est rarement à la place d'honneur, comme au portail de Montceaux-l'Étoile, où il s'entretient avec un homme en blanc, livre contre lui, reléguant la Vierge au second plan.

En premier lieu, les personnages permettent à eux seuls de figurer l'Église en tant qu'institution. Le Christ, les apôtres et la Vierge offrent une vision du corps de l'Église : le Christ en est la tête, les apôtres le clergé et la Vierge, la communauté des fidèles. Mais dans quelle mesure la présence de la Vierge induit celle de Pierre et/ou de Paul ?

[5] Paul a participé au martyre de saint Etienne. De retour, sur le chemin de Damas, il est aveuglé et une voix l'admoneste « Saül ! Saül ! Pourquoi me persécutes-tu ? ». Recouvrant la vue trois jours plus tard, le soldat romain se convertit au christianisme et devient son plus fervent prédicateur auprès des Gentils (*Ac* 9, 1-2).

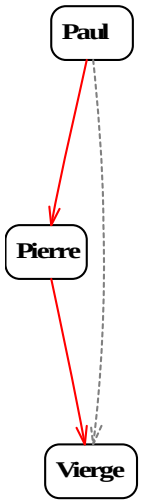


Fig. 16 : Graphe implicatif Vierge-Pierre-Paul (1-α 0,99 et 0,88). La présence de Paul implique celle de Pierre, qui implique celle de la Vierge, selon un indice de quasi implication de 0,99.

Le graphe (Fig. 16) montre la forte implication des trois personnages dans les représentations de l'Ascension : la présence de Paul inclut celle de Pierre, qui inclut celle de la Vierge. Une fermeture transitive permet aussi d'observer une quasi implication forte entre la présence de Paul et celle de la Vierge.

La mise en pendant de saint Pierre et saint Paul dans une scène d'Ascension valorise les deux figures de proue de l'Église. La présence de la Vierge complète cette vision de l'Église : cette dernière prend corps sous les yeux du fidèle grâce à la représentation des différents protagonistes ayant participé à sa naissance, qu'elle soit spirituelle ou institutionnelle. Certaines figures sont parfois particulièrement valorisées : par exemple, le sacramentaire d'Ottobeuren entoure la Vierge d'un voile tenu par saint Pierre et saint Paul (Fig. 4), mettant en avant le dogme de l'Incarnation. Le portail de Montceaux-l'Etoile présente saint Pierre ployant sous le poids des clefs (Fig. 3) : son rôle sacerdotal est à l'honneur.

4 Un objet à l'aplomb du Christ

Quinze œuvres figurent un objet de forme parallélépipédique sous le Christ montant. Les angles sont parfois adoucis, cassés, ou de forme circulaire. Cet objet est au sommet du mont, et le cas échéant, sert d'appui au Christ (Fig. 10). Andrea Worm note que ce détail apparaît en Basse Saxe entre le 12e et la première moitié du 13e siècle (Worm 2003, p. 313). L'auteur identifie ce motif à la pierre de l'église de l'Ascension. En effet, selon elle, la recrudescence de ce détail est le reflet des récits de pèlerinage de Croisés, notamment de Pierre Tudebold et de Daniel abbé russe, qui en 1106, mentionnent tous les deux la pierre de l'église de l'Ascension. Lorsque les Croisés arrivèrent en Terre sainte, l'église de l'Ascension était très abîmée, et ils décidèrent de remplacer la pierre portant les empreintes des pieds du Christ par une autre, sans empreinte celle-ci, et qui tient lieu et place de vénération. Worm précise que d'après les écrits de Tudebold et de l'abbé Daniel, la pierre était ronde, emprise dans un autel en marbre ne laissant voir que

la partie supérieure de la pierre : c'est elle que les fidèles embrassent. Selon Worm, l'iconographie reprend les récits des Croisés : l'objet est un bloc de pierre[6], qu'elle assimile d'ailleurs souvent et sans équivoque à un bloc de marbre[7] faisant référence à cette nouvelle pierre, et pour sa noblesse. Les empreintes de pieds apparaissent plus tardivement, tout d'abord parce qu'elles font référence à l'humanité du Christ alors que selon elle, l'iconographie s'axe davantage sur la filiation divine du Christ :

À partir de 1140, les textes mentionnent les traces de pieds[8], détail que l'iconographie reprendra entre 1208 et 1213 : il s'agit de rendre le site plus reconnaissable pour ceux qui ne seraient pas allés sur les Lieux saints. Nous rejoignons Worm dans l'idée que ce bloc est une pierre et qu'il peut s'agir d'une traduction des récits des Croisés renvoyant à l'église de l'Ascension.

Cela dit, si nous ne rejetons pas l'hypothèse selon laquelle ce bloc est la pierre de l'église de l'Ascension, nous souhaitons explorer une autre piste, en rapport avec la signification ecclésiologique liée à l'Ascension.

Observons les formes de cette pierre en elle-même. Deux plats de reliure[9] dessinent une pierre circulaire. Sur les autres œuvres, treize représentent des pierres parallélépipédiques, dont deux sont sculptées, évoquant un autel. Nous comptons également deux pierres brutes. Les pierres non taillées peuvent être une référence directe à Lothaire de Segni, le futur pape Innocent III qui, reprenant *Ex*ode 20, 25, insiste sur la nécessité de ne pas tailler la pierre de l'autel[10]. La pierre circulaire des plats de reliure peut être une référence directe au témoignage de l'abbé Daniel qui, comme le souligne Worm, mentionne une pierre ronde.

Pour les autres formes, nos réflexions s'orientent vers la lecture typologique assimilant le Christ à la pierre angulaire dont *Isaïe* fait état (« Voici que je [Dieu] vais poser en Sion une pierre, une pierre de granit, pierre angulaire, précieuse, pierre de fondation bien assise : celui qui s'y fie ne sera pas ébranlé » - *Is*. 28, 16) ou à la pierre rejetée devenue pierre d'angle dans le *Ps*. 118 (117), 22 (« La pierre qu'ont rejetée ceux qui bâtissaient est devenue pierre d'angle »). À plusieurs reprises dans le Nouveau Testament, la pierre angulaire est le Christ, comme en témoigne *Mc* 21, 42, qui reprend le texte vétérotestamentaire, ou saint Paul dans *Eph*.2, 20 :

> « Car la construction que vous êtes a pour fondation les apôtres et prophètes, et pour pierre d'angle le Christ Jésus lui-même. »

Sur les représentations de l'Ascension, la pierre à l'aplomb du Christ pourrait alors être la pierre angulaire de l'Église tandis que les témoins, premiers fidèles, en seraient

---

[6] « Steinblock » selon la désignation de Worm (2003, p. 299)

[7] Par exemple, sur l'Évangéliaire de Prüm (Fig. 10) : « mit deutlicher Marmoräderung gÉzéchiel eichneter Quader gestaltet », Worm 2003, p. 299.

[8] Gauthier de Saint-Victor écrit *Unde ex quo fuit perfecte humiliatus, statim ascendere coepit. Primo de inferno ad sepulchrum, deinde de sepulchro in mundum, de mundo in coelum, praebens nobis exemplum ut sequamur vestigia eus* (CCCM 30, p.131), ainsi qu'Honorius Augustodunensis, plus tard : *Vestigium namque quod ascendens harenae inpressit, adhuc locus ille retinet, et cum terra cottidie a fidelibus inde tollatur, vestigium deleri non potest* (PL 172, col. 958C).

[9] Un plat de reliure est une couverture de manuscrit. Les plus prestigieux sont en ivoire et/ou en orfèvrerie, et proposent un programme iconographique, c'est-à-dire plusieurs images réunies.

[10] Dans *Ex* 20, 25, Dieu s'adresse à Moïse en ces termes : « Si tu [Moïse] m'élèves un autel de pierre, tu ne le bâtiras point en pierres taillées; car en passant ton ciseau sur la pierre, tu la profanerais ». Innocent III développe cette idée : *PL* 217, col. 203.

les pierres vivantes « dans le cadre d'une construction sociale pensée comme une construction spatiale », pour reprendre les mots de Dominique Iogna-Prat (2006, p.313). Cette pierre marquerait alors la naissance de l'Église lors de l'Ascension :

- métaphoriquement, en montant au ciel, le Christ est la pierre angulaire de l'Église en tant qu'institution ;
- elle symbolise aussi l'église en rappelant le matériau de construction.

Cette identification du motif à la pierre angulaire, dans une optique ecclésiologique, est d'autant plus envisageable que le schéma suivant (Fig. 17) montre les relations entre l'objet en question, saint Pierre et saint Paul. Nous ajoutons également la variable « sym » (symétrie) pour mieux définir une possible séparation significative dans la partie basse de l'image. La rareté de l'objet « cube », en comparaison avec la Vierge par exemple, nécessite de baisser le niveau d'implication à 0,84, qui reste toutefois élevé.

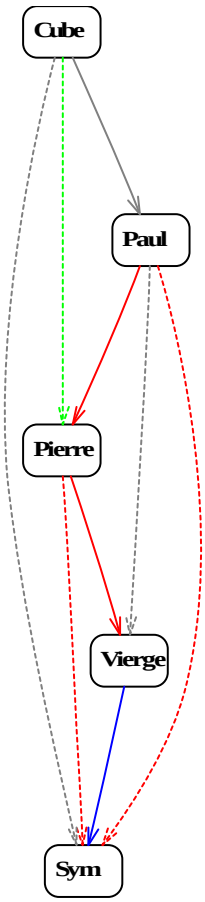


Fig. 17 : graphe implicatif (1-α= 0,99, 0,95, 0,90 et 0,84). Le peu de fréquence de l'objet a nécessité l'abaissement de l'indice de quasi implication pour faire apparaître les connexions possibles entre les motifs. La présence de l'objet implique celle de Pierre selon un indice de 0,9, ainsi que celle de Paul selon un indice de 0,84.

La représentation de l'objet implique celle de Paul d'après une quasi implication de 0,84, celle de Paul induisant également celle de Pierre selon une quasi implication très forte, de 0,99. Mais si Paul n'est pas représenté dans l'image où est figuré le cube, il y a toutefois de fortes probabilités pour que Pierre figure. Les deux apôtres les plus importants dans la fondation et l'épanouissement de l'Église sont ainsi associés à

l'objet. Ils renforcent l'hypothèse de la pierre angulaire comme métaphore de l'Église, d'autant qu'ils ont eux-mêmes interprété la pierre angulaire comme métaphore ecclésiale. De plus, dans Mt 16, 18-19, le Christ dit à Pierre « Tu es Pierre et sur cette pierre, je bâtirai mon église ». L'implication entre la présence de l'objet et celle du premier apôtre est très forte (0,99) : il peut y avoir un jeu de correspondance entre le Christ fondateur de l'Église, Pierre représentant de l'Église institutionnelle, et le cube symbolisant la communauté des fidèles mais aussi le lieu de culte. Par ailleurs, nous relevons que tous ces motifs créent une symétrie dans l'image, marquant une séparation en deux groupes dans la partie basse de l'image, l'un avec saint Pierre à sa tête, le second avec la Vierge, et manifestant ainsi la communauté ecclésiale.

Autel, pierre d'angle et Christ sont associés. Ce lien semble d'autant plus fort avec les images où cette pierre apparaît que le Christ porte une croix comme sur l'évangéliaire de Prüm (Fig. 10), et que deux lettres enluminées représentent une pierre sculptée dont le décor évoque les autels. Nous pouvons donc interpréter cette pierre comme aussi une métaphore du Christ et de la communauté ecclésiale, dont le Fils de Dieu est la pierre angulaire, et saint Pierre et saint Paul les principaux chantres.

5 Conclusion

À travers ces trois études de cas, nous souhaitions démontrer les différentes applications de l'analyse statistique implicative en iconographie médiévale. Le premier graphe a permis de mettre en avant les différentes manières de représenter le Christ. A partir de cet exemple, nous avons décidé de pousser notre investigation en nous penchant sur les attributs, afin de déterminer une possible correspondance entre la façon de figurer le Fils de Dieu, et ses attributs. Mais le croisement des données n'a pas toujours le résultat escompté : nous pensions en effet voir apparaître des liens plus explicites entre une action du Christ et un objet qu'il porterait, comme la croix, ce qui orienterait une réflexion, mais ce résultat ne fut pas probant. Néanmoins, cette absence de lien peut aussi être significative, et guider le chercheur vers une interprétation. En revanche, les mises en lien peuvent être fructueuses. Ainsi, nous avons voulu étudier la liaison entre les témoins, et leur relation avec l'objet. Les liens établis grâce à l'ASI ont permis d'émettre des hypothèses par rapport à l'ecclésiologie. Au terme de cette étude, l'utilisation de l'ASI dans des analyses sérielles en histoire de l'art se révèle intéressante : il est possible de déterminer des liens entre les motifs étudiés, et selon les implications observées, de proposer de nouvelles hypothèses d'interprétation du sujet étudié.

Références

[1] Baschet J. (1996). Inventivité et sérialité des images médiévales. Pour une approche iconographique élargie. *Annales HSS*, 51/1, 93-133.

[2] Chenu, M.D. (1957). *La Théologie au XIIe siècle*, Vrin, Paris.

[3] Congar, Y.-M., o.p. (1954), Marie et l'Église dans la pensée patristique, *Revue des sciences philosophiques et théologiques*, tome XXXVIII, 3-38.

[4] Congar, Y.-M., op. (1968), *L'Ecclésiologie du haut Moyen Âge*, Paris, Cerf.

[5] Gauthier de Saint-Victor, *Sermo XV De Ascensione Domine, Galteri a Sancto Victore et quorumdam aliorum*, Brepols, Turnhout, *CCCM 30*, 1975, 129-135.

[6] Gras, R., Régnier, J.-C., Guillet, F. (2009), *Analyse statistique implicative: Une méthode d'analyse de données pour la recherche de causalités*, Éditions Cépaduès, Toulouse.

[7] Gras, R., Régnier, J.-C., Marinica, C., Guillet, F. (2013), *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.*, Éditions Cépaduès, Toulouse.

[8] Guénot, M., Régnier J.-C. (2009), Iconographie médiévale en histoire de l'art et Analyse Statistique Implicative, In R. Gras, J.-C. Régnier, F. Guillet. *Analyse statistique implicative: Une méthode d'analyse de données pour la recherche de causalités*, Éditions Cépaduès, Toulouse, 471-492.

[9] Guénot, M., Régnier J.-C. (2010), L'histoire de l'art à l'épreuve de l'analyse statistique implicative : l'exemple de la structure iconique de l'image médiévale, *"Quaderni di Ricerca in Didattica (Mathematics)"*, n°20 suppl 1, 497-522.

[10] Guénot, M., Régnier J.-C. (2013), Comment l'analyse statistique implicative participe-t-elle à l'interprétation d'une œuvre d'art ? In R. Gras, J.C. Régnier, C. Marinica, F. Guillet (2013), *L'analyse statistique implicative Méthode exploratoire et confirmatoire à la recherche de causalités*. Cépaduès, Toulouse, 497-506.

[11] Iogna-Prat, D. (2006), *La Maison Dieu, Une histoire monumentale de l'Église au Moyen Age (v.800-v.1200)*, Seuil, Paris.

[12] Innocent III, Sermo XIV. *In Solemnitate annuntiationis et gloriosissimae semper Virginis Mariae, PL 217*, Amboise, J.-P. Migne, 1855, col.521B-526C.

[13] Worm, A. (2003), Steine und Fußspuren Christi auf dem Ölberg Zu zwei ungewöhnlichen Motiven bei Darstellungen der Himmelfahrt Christi, *Zeitschrift für Kunstgeschichte*, 66 Bd., H. 3, 297-320.

A FORMAÇÃO HUMANA DOS EDUCADORES ATRAVÉS DAS COMUNICAÇÕES NOS COLÓQUIOS INTERNACIONAIS SOBRE ANÁLISE ESTATÍSTICA IMPLICATIVA (A.S.I.): casos dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7

Djailton PEREIRA DA CUNHA[1], Nadja Maria ACIOLY-REGNIER[2] et Aurino LIMA FERREIRA[3]

LA FORMATION HUMAINE DES EDUCATEURS A TRAVERS LES COMMUNICATIONS DANS LES COLLOQUES INTERNATIONAUX D'ANALYSE STATISTIQUE IMPLICATIVE (A.S. I): CAS DES COLLOQUES A.S.I.5, A.S.I.6 ET A.S.I.7.

THE HUMAN TRAINING OF EDUCATORS BY CONTRIBUTIONS FROM INTERNATIONAL MEETING STATISTICAL IMPLICATIVE ANALYSIS (S.I.A.): CASE MEETING S.I.A.5, S.I.A.6 AND S.I.A.7.

RESUMO

A educação e o sujeito contemporâneo ora situados em um contexto sócio-econômico-cultural que por vezes privilegia a fragmentação do ser, reforçando o racionalismo e o pensamento reducionista-utilitarista no aspecto educacional e social, tem nos levado a refletir sobre o "lugar-modelo" da Formação Humana do educador e da sua prática educacional. Neste sentido, o presente estudo tem o propósito de identificar como os artigos apresentados nos colóquios internacionais da Análise Estatística Implicativa (A.S.I.) podem contribuir para uma reflexão sobre o processo de Formação Humana que contemple uma visão multidimensional do ser podendo assim eventualmente influenciar a prática educacional dos educadores-pesquisadores. Para tanto, foi realizada uma pesquisa do tipo descritiva, utilizando os métodos quantitativo e qualitativo de análise. Como instrumento para a coleta de dados foi adotado um roteiro estruturado de investigação, aplicado aos 68 artigos que contemplam os colóquios A.S.I.5, A.S.I.6 e A.S.I.7. Para as análises foram empregadas a técnica de Análise de Conteúdo (Bardin, 1977) e o quadro teórico da A.S.I. (Gras *et al*., 2009, 2013) com a utilização do CHIC (Couturier *et al*., 2012). Os resultados possibilitaram identificar que os colóquios têm contribuído para a formação dos pesquisadores, principalmente no que se refere a formação acadêmica-profissional. Por outro lado, também sugerem uma compreensão de que a Formação Humana tem sido escopo dos colóquios da A.S.I, o que é evidenciado com o número crescente de publicações e de pesquisadores de diversas disciplinas e de diferentes culturas que são incitados a trabalhar de forma intercultural e interdisciplinar nos contextos desses coloquios

Mots-clés : *Formação Humana; Educação, Análise Estatística Implicativa, CHIC, Interculturalidade*

[1]Doutorando em cotutela pelo Programa de Pós-Graduação em Educação da Universidade Federal de Pernambuco (pesquisa realizada com apoio da CAPES) e em Ciências da Educação pela escola doutoral ED485 EPIC [Éducation, Psychologie, Information & Communication] da Universidade Lumière Lyon 2, e-mail : djailtoncunha@uol.com.br

[2] ESPE Université Lyon 1. EAM-SIS-HCL 4128. e-mail: acioly.regnier@wanadoo.fr

[3] Programa de Pós-Graduação em Educação da Universidade Federal de Pernambuco (UFPE), Recife, Brasil, e-mail: aurinolima@gmail.com

RÉSUMÉ

L'éducation et le sujet contemporain se trouvent maintenant dans un contexte socio-économique et culturel qui favorise parfois la fragmentation de l'être, renforçant le rationalisme et la pensée réductionniste-utilitariste autant dans le domaine éducatif que social. Ceci nous a amené à réfléchir sur le "place-modèle «de la formation humaine des éducateurs et de leur pratique éducative. En ce sens, cette étude vise à déterminer comment les communications présentées lors des conférences internationales sur Implicative Analyse Statistique (ASI) peuvent contribuer à une réflexion sur le processus de formation humaine qui couvre une vision multidimensionnelle de l'être et influence par conséquent la pratique éducative des enseignants-chercheurs. Pour cela, une enquête de type descriptive a été réalisée en utilisant des méthodes quantitatives et qualitatives d'analyse. La construction des données a été réalisée par une étude des 68 résumés des communications publiées dans colloques A.S.I.5, A.S.I.6 et A.S.I.7. Pour l'analyse, nous avons utilisé la technique de l'analyse de contenu (Bardin, 1977), le cadre théorique de l'A.S.I. (Gras *et al*., 2009, 2013); et les données produites par le logiciel chic (Couturier *et al*., 2012). Les résultats ont permis d'identifier que les colloques ont contribué à la formation des chercheurs, notamment en matière de formation académique et professionnelle. D'autre part, ils suggèrent aussi une contribution forte pour la formation humaine comme en témoigne le nombre croissant de publications et des chercheurs de différents champs disciplinaires et de cultures distinctes dont la rencontre « oblige » un travail interdisciplinaire et interculturelle pour l'avancement conceptuel.

Mots-clés : *Formation Humaine; Education, Analyse Statistique Implicative, CHIC , Interculturalité.*

ABSTRACT

Education and the contemporary subject located in a socio-economic and cultural context that sometimes favors the fragmentation of the reinforcing rationalism and the reductionist-utilitarian thought in the educational and social aspect, has led us to reflect on the "place-model "Human Formation of educators and their educational practice. In this sense, this study aims to identify how the papers presented at international conferences on Statistical Implicative Analysis (S.I.A.) can contribute to a reflection on the process of Human Formation that covers a multidimensional view of being and consequently influence the educational practice the teacher-researchers. For this, a survey of descriptive type was carried out using quantitative and qualitative methods of analysis. As a tool for data collection was carried out a structured script research, applied to 68 articles that include the meeting S.I.A.5, S.I.A.6 and S.I.A.7. For the analysis we used the content analysis technique (Bardin, 1977), approach S.I.A. (Gras *et al*., 2009, 2013) and CHIC (Couturier *et al*., 2012). The results allowed identifying that the talks have contributed to training of researchers, particularly with regard to academic and professional training. On the other hand, also suggests an understanding that the Human Formation is also scope of S.I.A conferences, as evidenced by the growing number of publications and researchers from different areas that contribute to the advancement of studies aimed at Human Formation educators.

Keywords : *Human Formation; Education, Statistical Analysis Implicative, CHIC, Intercultural.*

1 Introdução

A educação e o sujeito contemporâneo ora situados em um contexto sócio-econômico-cultural que por vezes privilegia a fragmentação do ser, reforçando o racionalismo e o pensamento reducionista-utilitarista no aspecto educacional e social, tem nos levado a refletir sobre o "lugar-modelo" da Formação Humana do educador e

da sua prática educacional. Neste sentido, o presente estudo tem o propósito de identificar como os artigos apresentados nos colóquios internacionais sobre Análise Estatística Implicativa (A.S.I.) podem contribuir para uma reflexão do processo de Formação Humana que contemple uma visão multidimensional do ser e como consequência seja susceptível de influenciar a prática educacional dos educadores-pesquisadores.

Para tanto serão abordados nesse trabalho: (1) a apresentação geral do artigo; (2) os aspectos históricos-conceituais da Formação Humana; (3) uma (re)leitura da prática educacional como uma visão multidimensional da educação e do educador; (4) os colóquios internacionais sobre análise estatística implicativa (A.S.I.): os casos dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7; (5) os fundamentos e procedimentos metodológicos; (6) Análise e discussão de resultados e (7) as considerações finais.

2 Aspectos históricos-conceituais da Formação Humana: Um encontro epistêmico-ontológico

Nessa seção temos a intenção de apresentar e refletir sobre a visão histórica-conceitual da Formação Humana. Mesmo sabendo que se trata de um recorte e que não temos a pretensão de esgotar todas as informações relativas a esses “marcos” na Grécia Antiga e na Alemanha da Modernidade, pretendemos explorar aspectos que influenciaram as pesquisas voltadas para essa temática na contemporaneidade.

Iniciamos nosso percurso rumo à base conceitual, tomando como referência, estudos que apresentam essa revisão sistemática do processo de formação. Utilizaremos, para tanto, principalmente os trabalhos de Jaerger (2003), Freitas (2010, 2005) e Gadamer (2008). Outros autores participarão dessa discussão, norteando a trajetória que pretendemos desenvolver com esse estudo, em busca de uma compreensão epistêmico-ontológica de-para a Formação Humana.

A Grécia antiga foi a pedra fundamental para o surgimento da filosofia e da educação na perspectiva de formar o homem com uma visão de mundo e de ser distinta e abrangente. Na compreensão dos antigos gregos, a ideia de educação estava assentada no esforço humano para o desenvolvimento de sua natureza corporal e espiritual. Nesse sentido, a concepção de educação perpassa a noção de Paidéia. Esse termo grego, que pode ter significado distinto face às expressões modernas como cultura, civilização, tradição, literatura, educação, assumia à época o sentido de cultivo, voltado à formação de um elevado tipo de Homem (Jaeger, 2003). Nesse sentido, corroboramos o pensamento de que a Paidéia e a Educação estariam imbricadas numa perspectiva de desenvolvimento humano, social e cultural, visando um caráter virtuoso (Pagni & Silva, 2007).

Nesse contexto, podemos caracterizar a Paidéia Grega como a educação que envolvia aspectos da cultura, da nobreza, pólis, voltada para guerreiros, heróis, governantes, autoformação, busca do divino. A educação, nesse cenário, não era algo pertencente ao indivíduo, era por natureza, de domínio da sociedade, atuando na vida e no crescimento da mesma. Isso levou a uma amplitude do campo de atuação da educação. Desse modo, a Paidéia Grega inaugurou, de forma original, a visão de formação humana da antiguidade ocidental, defendendo pelo exercício prático, um

caminho de desenvolvimento dos aspectos intrínsecos (subjetivos-morais) e extrínsecos (sócio-culturais) do ser.

Nessa configuração, alguns autores defendem que o termo grego Paidéia se aproxima da palavra alemã Bildung (Jaeger, 2003; Pagni & Silva, 2007; Freitas, 2005).

> A palavra alemã Bildung (formação, configuração) é a que designa do mais intuitivo a essência da educação no sentido grego e platônico. Contém ao mesmo tempo a configuração artística e plástica, e a imagem, "ideia", ou "tipo" normativo que se descobre na intimidade do artista. Em todo lugar onde esta ideia reaparece mais tarde na História, ela é uma herança dos Gregos, e aparece sempre que o espírito humano abandona a ideia de um adestramento em função de fins exteriores e reflete na essência própria da educação (Jaeger, 2003, p. 10).

Nesse aspecto, a concepção de educação e de formação, no âmbito do projeto iluminista, assume compreensões distintas. Enquanto educação tem foco na preparação do indivíduo para uma função social, referindo-se ao ensino como uma instituição social; a formação remete a perspectiva de Bildung. Essa concepção de formação-Bildung, por ser importante para a modernidade, articula concepções como Filosofia da Cultura, Pedagogia e Literatura ou romances de formação (Freitas, 2010). Percebemos com isso, que a compreensão de Bildung, assim como fora a Paidéia, envolve uma visão de formação humana que extrapola o conceito de uma educação com foco exclusivista no cognitivo por meio institucional.

Para Gadamer (2008), o conceito de formação (Bildung) demonstra uma mudança espiritual que nos aproxima do século XIX e nos distancia intensamente da época barroca. No entendimento desse autor, a formação envolve uma compreensão histórica, e esse aspecto histórico é o que vale para o entendimento das ciências humanas. Nas palavras do autor:

> [...] formação significa mais que cultura, ou seja, aperfeiçoamento de faculdades e de talentos. A ascensão da palavra "formação" desperta, antes, a antiga tradição mística, segundo a qual o homem traz em sua alma a imagem de Deus, segundo a qual foi criado, e que deve reconstruir em si mesmo. [...] No conceito "formação" (Bildung) encontra-se a palavra "imagem" (Bild). O conceito de forma retrocede para aquém da misteriosa duplicidade da "imagem" (Bild), que abrange tanto o significado de "cópia" (Nachbild) quanto o de "modelo" (Vorbild) (Gadamer, 2008, p.46).

Reforçando esse entendimento de Bildung, na tradição alemã, Röhr (2010) destaca que "[...] a diferenciação do conceito de Educação (Erziehung) e Formação (Bildung) é bem mais complexa. [...] Desta forma, podemos falar desses conceitos somente ligados a autores ou grupo de autores com pensamento bem próximo" (p.209). O ponto principal dessa reflexão é observar como tais conceitos fazem interface com as temáticas emergentes da contemporaneidade face à crise de sentido que afeta sistemas, processos, instituições e pessoas.

Esse entendimento quando relacionado ao processo de desenvolvimento do ser, pode causar implicações e efeitos distintos. Isso porque, conforme Ferreira (2010), "a busca para compreensão do humano, a partir de modelos de desenvolvimento, remonta fontes pré-modernas, modernas e pós-modernas" (p. 113). Mais do que isso, remete a uma visão multidimensional de educação, uma (re)leitura da prática educacional como veremos na próxima seção.

3 (Re)leitura da Prática Educacional: Uma visão multidimensional da educação e do educador

Diversas são as definições e compreensões sobre a prática educacional. Para esse estudo, a concepção de prática educacional está apoiada na compreensão de Educação como Formação Humana. Nesse contexto, a abordagem a ser empregada busca extrapolar as várias dimensões organizacionais, institucionais ou curriculares que podem configurar a formação e atuação do educador.

Em consonância com o escopo dessa pesquisa, vamos apresentar, sem pretensão de completude, a multidimensionalidade na formação (e na prática) do educador. Para tanto, seguiremos os passos de Rohr (1999), quando elege as dimensões axiológica, epistemológica e ética da Educação com suas respectivas idéias regulativas de integralidade, verdade e liberdade.

Ao mesmo tempo, para o conceito de integralidade do ser humano, adotaremos modelo multidimensional do ser humano (Rohr, 2010). Desse modo, faremos o reconhecimento da importância de cada uma das cinco dimensões básicas (física, sensorial, emocional, intelectual e espiritual) proposta pelo referido autor, destacando a centralidade da dimensão espiritual para a Formação Humana.

A dimensão axiológica na formação do educador compreende a visão da totalidade do ato pedagógico e da importância de decidir frente a ela (Rohr, 1999). Nesse sentido, podemos pensar que essa totalidade enquanto integralidade, caracterizada como idéia regulativa, não promove a orientação da prática pedagógica. De certo modo, isso é fato quando consideramos a perspectiva de que ela não aponta para uma práxis pedagógica pré-estabelecida e desenvolvida mecanicamente. Desse modo, como afirma Rohr (1999), é possível "interpretar a história das teorias e práticas educacionais até seu estado de discussão atual, como trajetória de negações e aproximações da idéia regulativa na formação da integralidade do ser humano. Tarefa que acreditamos ser indispensável na formação do educador" (p.103)

No que diz respeito à dimensão epistemológica do pedagógico, podemos compreender pelo que de fato impele a práxis, enquanto verdade, do educador. Dentro desse escopo,

> A questão da validade da ideia da integralidade do ser humano enquanto meta do pedagógico é no fundo a questão da verdade dela. [...] Cada um tem sua verdade, o que resta é encontrar um caminho de administrar essa situação. O relativismo absoluto, porém, não tem nem condições de se autoafirmar enquanto verdadeiro. [...] A verdade, nesse sentido, é uma idéia regulativa. Mesmo não a encontrando como fato empírico e pleno na nossa realidade, precisamos dela enquanto orientação da nossa busca (Rohr, 1999, p.103).

Já a dimensão ética do pedagógico pressupõe a aceitação e o compromisso integral com o ético. Nesse contexto, entendemos como Rohr (1999), que "a idéia regulativa do ético aparece na questão pedagógica e conseqüentemente na formação do educador" (p.105). Para tanto, reforçamos essa concepção, destacando que a prática educacional exige um ato ético com o educando, consigo mesmo e com o socioambiental; transcende, portanto, ao pedagógico. Tais aspectos corroboram a importância da dimensão ética, apontada no esquema multidimensional do ser, enquanto fator de promoção de Formação Humana (Rohr, 2010).

A partir desse cenário compreendemos que a formação e atuação do educador está além de abordagens técnica e profissional, sendo indispensável o foco para a dimensão humana. A integralidade das dimensões imanentes (física, sensorial, emocional, intelectual) e transcendente (espiritual), bem como, entre atitude e princípios é uma condição fundamental para o desenvolvimento dessa formação, o que refletirá no exercício da função do educador.

Nessa perspectiva, concordamos com Santiago (2010) quando afirma que "O verdadeiro trabalho formativo é conduzir o homem a uma vivência autêntica, que o possibilite a interligação com o mundo, face ao qual se compreende a defesa incondicional da solidariedade, do vínculo entre os humanos, através de uma vida em comunidade"(p.196). Isso assegura uma visão integral e multidimensional de Educação. Assim, podemos destacar que a educação como formação humana é práxis e não teoria. Essa práxis educacional é desenvolvida a partir de uma formação integral, e não somente de pressupostos técnico-metodológicos. Um processo infindável de possibilidades de vir-a-ser, que requer comprometimento com a realidade, consigo e com tudo que acontece a sua volta.

Neste caso, podemos presumir que educador e educando são os mesmos no encontro existencial em busca da integralidade do ser, que se configura com a Formação Humana. Seguindo essa linha de raciocínio, podemos entender que há possíveis contribuições do(s) artigos elaborados e apresentados em colóquios sobre análise estatística implicativa para Formação Humana dos educadores?

4 Casos dos colóquios internacionais A.S.I.5, A.S.I.6 e A.S.I.7

O colóquio internacional sobre análise estatística implicativa (A.S.I.) é um encontro de pesquisadores, educadores, profissionais e estudantes ligados a diversos campos do saber que buscam contribuir para o desenvolvimento e refinamento de estudos voltados para a A.S.I. Em cada colóquio realizado, há uma extensa produção de conhecimento científico disponibilizado através das conferências, cursos, oficinas e comunicações dos trabalhos efetuados, que exploram aspectos teóricos e práticos relativos à A.S.I.

Do ponto de vista histórico foram realizados 7 colóquios internacionais A.S.I. O primeiro colóquio (A.S.I.1) ocorreu na França, foi promovido pelo Instituto de Formação de Professores de Caen, nos dias 23 e 24 de junho de 2000, e organizado pelos pesquisadores Marc Bailleul e Régis Gras. O tema central desse 1º Colóquio A.S.I. teve como foco principal a "busca nos dados pelo Método da Estatística Implicativa". Isto correspondeu a um direcionamento promissor quando se considera o desenvolvimento e as atividades da Associação EGC (Extração e Gestão de Conhecimentos). Nesse colóquio foram apresentadas duas conferências e 14 comunicações orais.

O segundo colóquio (A.S.I.2) aconteceu no Brasil, sediado pela Pontifícia Universidade Católica de São Paulo, de 09 a 11 de julho de 2003, e organizado pelo pesquisador Saddo Ag Almouloud. A temática central desse encontro abordou "O Método Estatístico Implicativo utilizado em estudos qualitativos de associação de Régis Gras: contribuição para a pesquisa em Educação". A programação desse colóquio envolveu a realização de um mini-curso e de apresentações de seis estudos-comunicações orais.

O terceiro colóquio (A.S.I.3) teve lugar na Itália, conduzido pela Universidade de Palermo, de 6 a 8 de outubro de 2005, sob a organização do pesquisador Filippo Spagnolo. Na oportunidade desse colóquio foi apresentado o panorama de desenvolvimento da A.S.I. através de situações fundadoras por Regis Gras, além de outras 26 comunicações de trabalhos científicos, envolvendo 41 pesquisadores de diferentes nacionalidades.

O quarto colóquio (A.S.I.4) foi sediado na Espanha e promovido pela Universidade Jaume I de Castellón de la Plana, de 18 a 21 de outubro de 2007. Esse encontro foi organizado pelos pesquisadores Pilar Orus e Pablo Gregori. O objetivo desse colóquio foi oferecer: (1) elementos de formação para a A.S.I., (2) avanços teóricos no método de análise de dados não-simétricos, a análise Implicativa Estatística (ASI); (3) um número significativo de aplicações concretas em vários campos do saber, tais como, a Ciências da Educação, Didática da Matemática, Estatística, Psicologia, Economia, Biologia, etc. Na oportunidade foram desenvolvidas quatro temáticas centrais: (1) A análise quantitativa de dados, um exemplo: Análise Implicativa Estatística (A.S.I.). Aspectos teóricos e práticos sobre A.S.I.; (2) Formação em A.S.I. Primeiras aplicações significativas em várias disciplinas; (3) Tratamentos de situações em Ciências da Educação, Ciências Sociais e Psicologia pela A.S.I.; (4) Alguns aportes teóricos sobre o desenvolvimento e a extensão da A.S.I.; (5) Análise de conceitos ensinados nas aulas de matemática em diferentes níveis de ensino através da A.S.I.

O quinto, sexto e sétimo colóquios internacional A.S.I. seguiram a mesma lógica científica dos precedentes. Estes encontros são os casos estudados nesse trabalho, e por isso serão apresentados de forma mais detalhada nas próximas seções.

4.1 Caso 1 - 5° Colóquio Internacional sobre Análise Estatística Implicativa

O quinto colóquio (A.S.I.5[4]) teve lugar novamente na Itália, em Palermo, de 5 a 7 de novembro de 2010. O presidente da comissão científica foi o pesquisador Jean Claude Regnier e o presidente da comissão organizadora foi o pesquisador Filippo Spagnolo. O tema central desse encontro envolveu a "Análise estatística implicativa como um objeto de pesquisa e de formação em análise de dados, um instrumento para a pesquisa multidisciplinar".

Esse quinto colóquio teve como eixos temáticos:

- Conceitos fundamentais em A.S.I.: modelos estatísticos, tipos de variáveis, variáveis principais e suplementares
- Novos avanços, índices de estabilidade, extensão a novos tipos de variáveis, regras de exceção, dualidade espaço dos sujeitos - espaço das regras, estrutura métrica e topologia do espaço dos sujeitos induzidas por sua contribuição ou sua tipicalidade, análise vetorial.
- Comparação crítica das abordagens, dos modelos, das representações e dos resultados da A.S.I. com os dos outros métodos de análise de dados (treillis de Galois, redes bayesianas, árvores de indução, análise fatorial, etc.)
- Prática do software CHIC, desenvolvimentos atuais e esperados

[4] http://sites.univ-lyon2.fr/asi5/

- Aplicações realizadas através do tratamento ASI, comparadas as de outros métodos nos seguintes dominios: didática, ciências da educação, psicologia, sociologia, economia, história da arte, biologia, medicina, arqueologia, etc.
- Apresentações gráficas e numéricas dos resultados, auxilio à interpretação desses resultados, papéis respectivos e críticas dos tipos de variáveis, das variáveis principais e suplementares escolhidas
- Especificidades da formação à A.S.I.: uso do software CHIC, interpretação das representações gráficas (grafo implicativo; árvore da hierarquia coesitiva)
- Problemáticas da didática A.S.I.

A conferência de abertura - "*Spécificités de l'A.S.I. par rapport à d'autres mesures de qualité de règles d'association*" - foi proferida por Regis Gras, presidente de honra da A.S.I. 5. Na sequência do evento, foram realizadas três oficinas sobre "*CHIC_1/2/3 - Formation à l'A.S.I. par l'usage du logiciel CHIC*", conduzidas por Raphaël Couturier e Jean-Claude Oriol, e apresentadas 30 comunicações aprovadas pelo comitê científico entre aquelas que haviam sido submetidos. Essas comunicações envolveram 54 pesquisadores de diferentes instituições e nacionalidades como a alemã, brasileira, chilena, cipriota, cubana, espanhola, francesa, gabonesa, grega, italiana e japonesa. Reforçando o caráter internacional do colóquio, as comunicações puderam ser escritas em um dos seguintes idiomas: espanhol, francês, inglês, italiano e português. Uma característica encontrada na estrutura dos artigos publicados desse encontro é que apenas três comunicações possuem as palavras-chave.

4.2 Caso 2 - 6° Colóquio Internacional sobre Análise Estatística Implicativa

O colóquio A.S.I.6[5] aconteceu novamente em Caen, França, de 7 a 10 de novembro de 2012. O presidente da comissão científica foi o pesquisador Jean Claude Regnier e o presidente da comissão organizadora foi o pesquisador Marc Bailleul. O tema central desse encontro abordou "A análise estatística implicativa: de uma postura exploratória para uma postura confirmatória".

Este encontro manteve os eixos temáticos estabelecidos no A.S.I. 5. A conferência de abertura "*De l'observation clinique à la macro didactique : les aléas de l'analyse statistique et les contributions de l'analyse statistique implicative*", foi proferida por Guy Brousseau em homenagem a Filippo Spagnolo. Na sequência do evento, foi realizada uma oficina sobre a "*Formation à l'A.S.I. par l'usage du logiciel CHIC : prise en main et nouvelles fonctionnalités*", conduzida por Raphaël Couturier e Saddo Ag Almouloud, depois foram apresentadas e discutidas as 19 comunicações orais de trabalhos científicos, e foi feita uma conferência-demonstração sobre o tema "*Pour un parcours complet des étapes d'une recherche par une analyse de données sous l'approche A.S.I. avec le logiciel CHIC : du questionnement aux résultats, en passant par le questionnaire et les graphes*", por Marc Bailleul.

Neste evento A.S.I.6, as 19 comunicações apresentadas envolveram 31 pesquisadores de diferentes instituições e nacionalidades como a brasileira, chilena, espanhola, eslovaca, francesa, gabonesa, italiana, mexicana e uruguaia. Além disso, foi

[5] http://sites.univ-lyon2.fr/asi6/

mantida a dimensão internacional, iniciada no A.S.I.5, com a introdução das cinco línguas de trabalho: espanhol, francês, inglês, italiano e português.

4.3 Caso 3 - 7° Colóquio Internacional sobre Análise Estatística Implicativa

O sétimo colóquio (A.S.I.7[6]), ocorreu em São Paulo, Brasil, de 27 a 30 de novembro de 2013, 10 anos depois do segundo colóquio (A.S.I.2) também realizado em São Paulo. Nas duas ocasiões houve intensa apresentação de comunicações e "calorosos" debates acerca dessa temática. Em 2013, o tema central do encontro foi: Análise estatística implicativa: marco teórico e aplicativo para exploração semântica e não simétrica dos dados. O presidente da comissão científica foi o pesquisador Jean Claude Regnier e o presidente da comissão organizadora foi o pesquisador Saddo Ag Almouloud.

Este encontro manteve os principais eixos temáticos dos A.S.I 5 e A.S.I.6. Antes do colóquio, foi realizado um minicurso por Jean-Claude Régnier e Nadja Acioly-Regnier. A conferência de abertura "*Genèse et développement de l'analyse statistique implicative. rétrospective intuiti*", foi proferida por Régis Gras, presidente de honra do A.S.I.7. Na sequência do evento, duas oficinas sobre a "formação para o A.S.I. pelo uso do software CHIC: iniciação e novos recursos" foram lideradas por Raphael Couturier e Saddo Ag Almouloud ; bem como, 20 comunicações orais de trabalhos científicos foram apresentadas e discutidas.

Neste A.S.I.7, as 20 comunicações apresentadas envolveram 34 pesquisadores de diferentes instituições e nacionalidades como a argentina, brasileira, espanhola, francesa, gabonesa, inglesa, mexicana e uruguai. Além disso, foi mantida a dimensão internacional, iniciada no A.S.I.5, com a introdução das cinco línguas de trabalho: espanhol, francês, inglês, italiano e português.

5 Fundamentos e procedimentos metodológicos

Nesta seção apresentaremos a sistemática metodológica adotada no presente estudo. Para isso, abordaremos os fundamentos metodológicos da pesquisa, que compreendem a concepção filosófica, a estratégia de investigação e o método da pesquisa; bem como, os procedimentos metodológicos que envolvem a fonte, a coleta e a análise de dados.

Toda pesquisa científica está fundamentada em uma rede de pressupostos ontológicos, epistemológicos e metodológicos que definem a visão de mundo e da natureza humana do pesquisador (Richardson et al., 1999). Para tanto é importante refletir que essa visão está diretamente associada a uma das três abordagens adotadas em um estudo: qualitativa, quantitativa e método misto (qualitativo-quantitativo). Neste sentido, a concepção filosófica da pesquisa compreende a orientação geral sobre o mundo e a natureza da investigação defendida pelo autor. Por sua vez, a estratégia de investigação, também conhecida como abordagens da investigação ou metodologias de pesquisa, corresponde aos modelos de métodos qualitativo, quantitativo e misto. A partir dessa estratégia de investigação há um direcionamento dos procedimentos metodológicos da pesquisa (Creswell, 2010). Dentro dessa perspectiva, esta pesquisa é do tipo descritiva e emprega uma abordagem mista, utilizando os métodos de

[6] http://sites.univ-lyon2.fr/asi7/

investigação quantitativo e qualitativo, com o intuito de responder as questões e os objetivos propostos.

A fonte de dados, o *corpus* desta pesquisa, compreende as comunicações orais apresentadas nos colóquios A.S.I.5, A.S.I.6 e A.S.I.7, que foram publicadas posteriormente aos encontros e estão descritos neste trabalho como "casos" nas seções 4.1. 4.2 e 4.3. respectivamente. A escolha desses colóquios como "casos" foi decorrente da sistematização dos artigos como uma autêntica base de dados, disponibilizada na íntegra para consulta.

Como instrumento para a coleta de dados foi adotado um roteiro estruturado de investigação, aplicado aos 68 artigos publicados a partir dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7. Este roteiro está dividido em três etapas e está apresentado no anexo 1.

A primeira parte desse roteiro foi o levantamento dos dados demográficos (sexo, nacionalidade e instituição filiada) do primeiro autor de cada artigo pesquisado. Essa parte corresponde as variáveis suplementares do estudo.

A segunda fase foi a compreensão do resumo de cada artigo pesquisado, o que consistiu em definir o idioma que foi redigida a comunicação; se o título da publicação faz referência a formação profissional, formação humana, interculturalidade, cultura local, A.S.I., CHIC; se o resumo do artigo possui uma estrutura que contempla visão introdutória com o propósito/objetivo do estudo, referência teórica com explicitação de autor(es), procedimento metodológico, resultados e palavras-chave); se as referências fazem menção aos colóquios precedentes e aos trabalhos científicos voltados para a formação humana, e se o artigo foi publicado no colóquio A.S.I.5, A.S.I.6 ou A.S.I.7. Essa parte diz respeito à sistematização das variáveis principais.

Por fim, a terceira etapa desse roteiro de investigação envolve a caracterização dos colóquios estudados com a descrição da quantidade de artigos produzidos; quantidade de autores que apresentaram trabalhos e a quantidade de páginas produzidas. Desse modo, construímos 8 variáveis conforme quadro abaixo.

Variável	Tipo	Código	Descrição da Variável Codificada
Sexo	Suplementar	Masc s	Sexo Masculino
		Femi s	Sexo Feminino
Nacionalidade	Suplementar	Nac_Br s	Nacionalidade brasileira
		Nac_Es s	Nacionalidade espanhola
		Nac_Fr s	Nacionalidade francesa
		Nac_It s	Nacionalidade italiana
		Nac_Ou s	Outra Nacionalidade
Instituição	Suplementar	Inst_Br s	Instituição filiada brasileira
		Inst_Es s	Instituição filiada espanhola
		Inst_Fr s	Instituição filiada francesa
		Inst_It s	Instituição filiada italiana
		Inst_Ou s	Instituição filiada de outro pais
Língua	Principal	Ling_Es	Artigo escrito na língua espanhola
		Ling_Fr	Artigo escrito na língua francesa
		Ling_In	Artigo escrito na língua inglesa
		Ling_It	Artigo escrito na língua italiana
		Ling_Pt	Artigo escrito na língua portuguesa
Título	Principal	Tit_FP	Título do artigo faz referência à Formação Profissional.
		Tit_FH	Título do artigo faz menção à Formação Humana
		Tit_INT	Título do artigo faz referência à Interculturalidade
		Tit_CLO	Título do artigo remete à Cultura Local
		Tit_ASI	Título do artigo faz referência à A.S.I.
		Tit_CHIC	Título do artigo faz referência ao CHIC
Resumo	Principal	Res_Int	Resumo do artigo contém visão introdutória com o objetivo do estudo
		Res_Teo	Resumo do artigo contém referência teórica com explicitação de autor(es)
		Res_Met	Resumo do artigo contém procedimentos metodológicos adotados no estudo
		Res_Res	Resumo do artigo contém resultados
		Res_Mot	Resumo do artigo contém palavras-chave
Referências	Principal	Ref_ASI	Referências do artigo faz menção a colóquios precedentes
		Ref_FH	Referências do artigo faz menção a estudos relativos à Formação Humana
Colóquio	Principal	ASI_5	Artigo publicado no colóquio A.S.I.5
		ASI_6	Artigo publicado no colóquio A.S.I.6
		ASI_7	Artigo publicado no colóquio A.S.I.7

Quadro 1 – Apresentação dos dados

Para as análises foram empregadas a técnica de Análise de Conteúdo (Bardin, 1977) e o quadro teórico da A.S.I. (Gras *et al*., 2009, 2013) com a utilização do software CHIC (Couturier *et al*., 2012). A análise de conteúdo foi adotada visando mapear as categorias-chave dos resumos dos artigos pesquisados. Já a A.S.I. foi aplicada na perspectiva do conceito de implicação estatística ou mais especificamente sobre o conceito de quase implicação, diferenciando do entendimento de implicação lógica de natureza matemática e lógica. Nesse sentido, o CHIC foi utilizado para melhor caracterização da análise estatística implicativa através dos recursos oferecidos por esse software.

Para análise de conteúdo, seguimos as fases estabelecidas por Bardin (1977), que compreende: 1) pré-análise, 2) exploração do material e 3) tratamento dos resultados, inferência e interpretação. Na pré-análise, organizamos o material que foi analisado com o intuito de sistematização das ideias iniciais. Isso implicou em: (a) leitura dos resumos dos artigos selecionados; (b) escolha do item que foi analisado (as palavras-chave); (c) levantamento dos objetivos com a análise de conteúdo; (d) estabelecimento de indicadores através dos textos sob avaliação. Na exploração do material, definimos as categorias, as unidades de registro e as unidades de contexto nos documentos. Nesta fase efetuamos a codificação, a classificação e a categorização das unidades estudadas. A última etapa desse processo de análise de conteúdo consistiu no tratamento dos dados obtidos com a codificação-classificação-categorização realizadas na fase anterior, possibilitando com isso a inferência e a interpretação dos resultados. Como destaca Bardin (1977), essa fase é dedicada a avaliação crítica e reflexiva, o que envolve intuição e razão.

Para a A.S.I, adotamos como referência os trabalhos desenvolvidos por Gras *et al*. (2009, 2013) e utilizamos o CHIC para operacionalização da A.S.I. Como afirma Couturier *et al* (2012) "O software CHIC tem por funções essenciais extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecer um índice de qualidade de associação e de representar uma estruturação das variáveis obtida por meio destas regras"(p.1).

Nesse contexto, foram analisados os artigos publicados nos colóquios A.S.I.5, A.S.I.6 e A.S.I.7 que abordam as variáveis objeto de estudo dessa investigação. Em cada artigo selecionado foi analisado o resumo. Observamos neste resumo a presença (1) ou ausência (0) de cada variável estudada. Com base nos resultados, criamos uma tabela que possibilitou uma avaliação preliminar dos itens presentes nos artigos selecionados. Na sequência, já com os dados dispostos na forma binária, fizemos uso do software CHIC para tratamento e disponibilização dos elementos-chave para as análises e discussões dos resultados.

6 Análises e discussões dos resultados

Nesta seção vamos apresentar as principais considerações realizadas a partir da análise de conteúdo e da A.S.I. desenvolvidas, conforme procedimento descrito no item anterior.

6.1 Discussão dos resultados a partir da análise de conteúdo

No processo de análise conteúdo identificamos 211 palavras-chaves (PC) nos artigos pesquisados dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7. Estratificando esse resultado por colóquio realizado, constatamos que apenas 3 artigos do A.S.I.5 tiveram PC em seus respectivos resumos, perfazendo um total de 13 palavras. Essa performance é completamente diferente quando avaliamos os demais colóquios. No A.S.I.6 tivemos 100% dos artigos com a inclusão de PC, totalizando 96 palavras. Já no A.S.I.7, obtivemos 95% das comunicações com PC, totalizando 102 palavras. Nesse colóquio, apenas uma publicação não tinha o registro das PC. O quadro 2 ilustra essa análise.

Colóquio	Artigos publicados	Artigos com PC	Qtd de PC no colóquio	% Artigos com PC
A.S.I.5	29	3	13	10%
A.S.I.6	19	19	96	100%
A.S.I.7	20	19	102	95%
Total	**68**	**41**	**211**	**60%**

Quadro 2 – Estratificação das Palavras-Chave (PC) por colóquio A.S.I.

A partir desses dados, iniciamos a sistemática de classificação-codificação e mapeamos 12 classes: Formação de professores, Funções, Formação profissional, A.S.I., Variáveis, Formação humana, Implicações, Regras, Método de pesquisa, CHIC, Validação e Instrumento, conforme ilustra o gráfico abaixo.

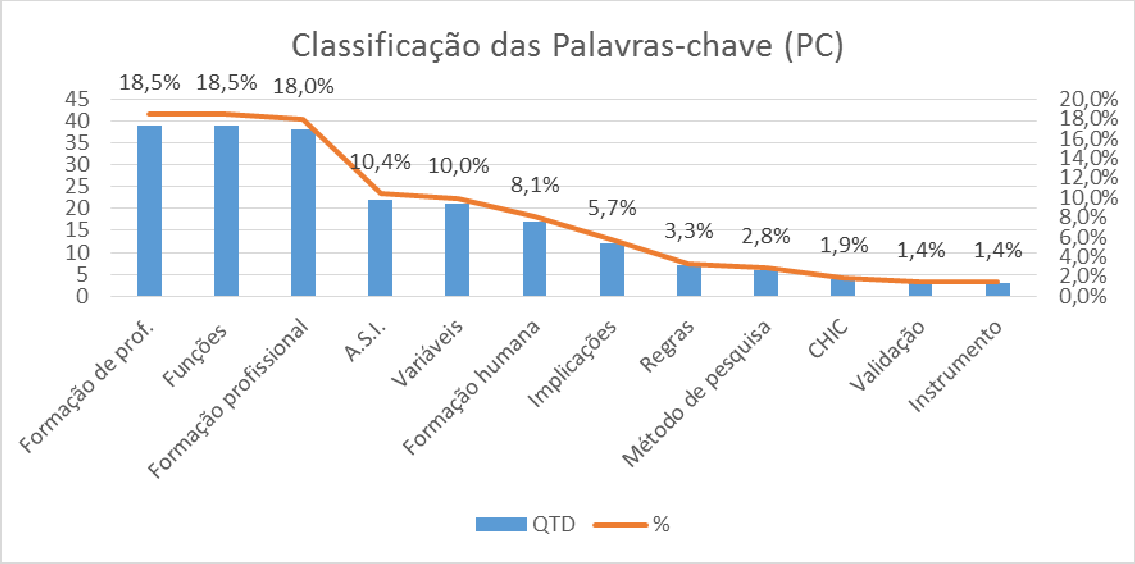


Figura 1- Gráfico de Classificação das Palavras-chave (PC)

Essas 12 classes, quando retomamos os artigos para uma nova leitura-reflexiva, percebemos que elas podem ser reorganizadas e assim chegarmos as três grandes categorias que retratam os principais interesses dos trabalhos apresentados nos colóquios analisados. De forma sucinta, podemos compreender que o foco das pesquisas apresentadas nos A.S.I.5, A.S.I.6 e A.S.I.7 está concentrado nas grandes-categorias formação, metodologia de análise e teorização matemática-estatística.

Cerca de 44,6% das PC codificadas estão relacionadas à formação, seja a formação dos professores (18,5%), formação de outros profissionais como estatísticos, matemáticos, psicólogos, médicos, etc (18,0%), ou ainda a formação humana (8,1%), entendendo essa última como um conjunto de atitudes, competências e conhecimentos adquiridos que extrapolam o aspecto cognitivo-racional.

No que se refere a metodologia de análise, percebemos que 28,0% das PC codificadas estão diretamente associadas a essa categoria. Isso porque agrupamos como metodologia as classes voltadas para a A.S.I (10,4%), variáveis (10,0%), estratégia de investigação (2,8) CHIC (1,9%), instrumentos (1,4%) e validação (1,0%)

A última categoria diz respeito aos aspectos teóricos da matemática-estatística, representando 27,5% das PC codificadas. Essa categoria envolveu as classes funções (18,5%), implicações (5,7%) e regras (3,3%).

Com base nesses resultados, podemos identificar que os colóquios têm sido fortemente usados para intensificar as discussões sobre a formação dos pesquisadores, as metodologias de pesquisa principalmente reforçando e/ou comparando a A.S.I com outras técnicas, contribuindo também para a expansão das abordagens teóricas acerca da matemática-estatística no que se refere mais especificamente às funções, regras e (quase) implicações chave.

6.2 Discussão dos resultados a partir da A.S.I.

O processo de análise da estatística implicativa foi iniciado com a preparação da tabela de dados, conforme pode ser vista no anexo 2. Percebemos, a partir de uma análise ainda preliminar que a variável Res_Int está presente em todos os artigos, enquanto que a variável Ling_It só está presente em um artigo. Isso nos aponta que todos os artigos possuem uma visão geral introdutória em seus resumos, com uma contextualização e apresentação do próposto a ser alcançado, por outro lado, nos indica que apesar de termos vários pesquisadores de nacionalidade italiana, apenas um utilizou a língua materna para redigir seu artigo.

No que se refere a formação humana, ainda de forma incipiente, pela análise da tabela, podemos observar que são utilizadas mais referências bibliográficas voltada para essa temática (Ref_FH) do que os títulos e consequemente os artigos sejam voltados para a Formação Humana (Tit_FM).

Esses resultados iniciais indicam limitações da análise. Com intuito de reverter essa situação, adotaremos em seguida a abordagem da Analise Estatística Implicativa – A.S.I. utilizando o software CHIC.

Estudamos 33 variáveis binárias (12 variáveis suplementares e 21 variavéis principais), a partir da amostra de 68 resumos dos colóquios ASI.5, ASI.6 e ASI7. Para desenvolvimento da análise estastítica implicativa, vamos empregar as análises de similaridade, classificação baseada no modelo de Lerman (1981), e análise implicativa a partir do estudo do gráfico de implicação e as contribuições das variáveis suplementares.

6.2.1 Análise de similaridade

A figura 2 – Grafíco de Similaridade – expressa a árvore de similaridade, uma classificação estabelecida baseada no modelo de Lerman (1981), onde são empregados índices de probabilidade. Constatamos a separação das variáveis binárias em uma classe que se divide em outras duas subclasses.

Encontramos 20 níveis de similaridade entre 0.999638 e 1.10259e-07. Estabelecemos um índice de similaridade mínimo de 0,75 para desenvolvermos nossa análise e constatamos oito níveis. Dessas interações, temos nós significativos nos níveis: 1, 6, 8, 10, 12, e 16. O nó mais significativo está no nível 1

No primeiro nível de similaridade, encontramos uma relação de similaridade entre as variáveis Tit_FH e Ref_FH com um índice de similaridade de 0.999638, o que indica

uma chance alta do título do artigo fazer menção a Formação Humana e destacar publicações relativas a essa temática nas referências.

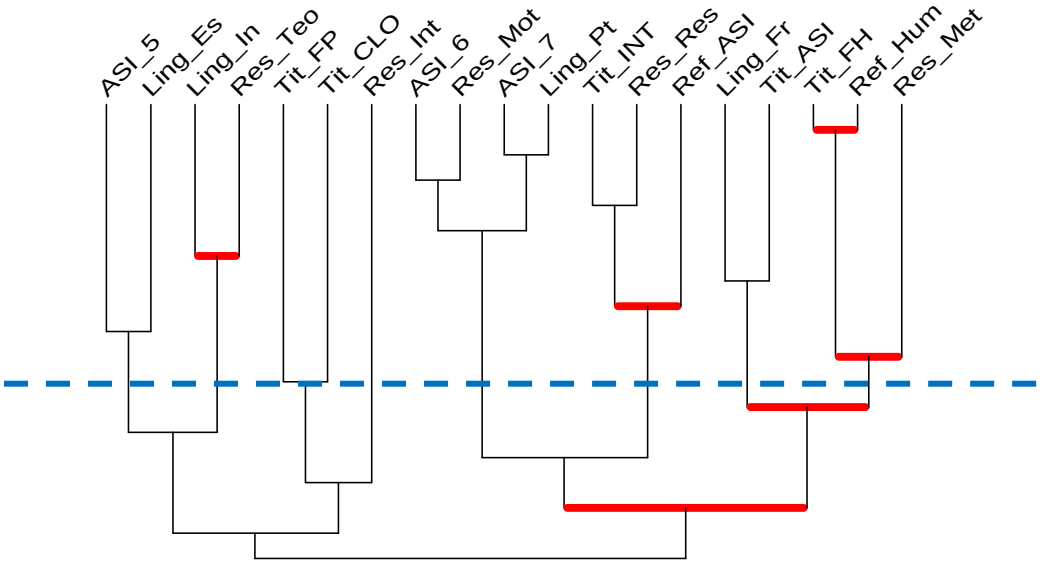


Figura 2 – Gráfico de similaridade

No segundo nível de similaridade, com índice de 0.999499, temos as variáveis Ling_Pt ASI_7. Isso reforça a condição de que o colóquio realizado no Brasil possibilitou uma maior adesão de pesquisadores brasileiros, haja visto que os artigos redigidos em português estão mais associados ao colóquio A.S.I.7, ocorrido em São Paulo-Brasil. Isso denota também que esses pesquisadores preferiram utilizar a língua materna para a escrita de suas comunicações.

Destacamos a seguir as contribuições da varíaveis suplementares para a classe Tit-FH e Ref_Hum por apresentarem o maior nível de similaridade e por envolverem variáveis que estão diretamente ligadas ao principal objeto desse trabalho, que é o processo de formação humana.

Contribuição à classe : Tit_FH,Ref_Hum (1)

Grupo ótima : a65 a29 a53 a16 a47 a14 a41 a42 a13 (card GO 9, p 0.132, 1-p 0.868)

A variável Masc contribui a esta classe com um risco de : 0.872 (intersection avec le groupe optimal 2)

A variável Femi contribui a esta classe com um risco de : 0.0528 (intersection avec le groupe optimal 7)

A variável Nac_Br contribui a esta classe com um risco de : 0.0143 (intersection avec le groupe optimal 4)

A variável Nac_Fr contribui a esta classe com um risco de : 0.457 (intersection avec le groupe optimal 3)

A variável Nac_It contribui a esta classe com um risco de : 0.433 (intersection avec le groupe optimal 0)

A variável Nac_Es contribui a esta classe com um risco de : 0.679 (intersection avec le groupe optimal 0)

A variável Nac_Ou contribui a esta classe com um risco de : 0.505 (intersection avec le groupe optimal 2)

A variável Inst_Br contribui a esta classe com um risco de : 0.0631 (intersection avec le groupe optimal 3)

A variável Inst_Fr contribui a esta classe com um risco de : 0.0528 (intersection avec le groupe optimal 7)

A variável Inst_It contribui a esta classe com um risco de : 0.433 (intersection avec le groupe optimal 0)

A variável Inst_Es contribui a esta classe com um risco de : 0.573 (intersection avec le groupe optimal 0)

A variável Inst_Ou contribui a esta classe com um risco de : 0.737 (intersection avec le groupe optimal 1)

Percebemos que as variavéis complementares que mais contribuem para esta classe são : Fem, Nac_Br, Inst_Fr e a variável que contribui mais a esta classe é Nac_Br com um risco de 0.0143.

6.2.2 Análise de implicação

Os índices de implicações, segundo a teoria clássica, calculados usando a lei binomial, são apresentados na figura abaixo. Exploramos o gráfico implicativo (Figura 3), construído a partir de 33 variáveis binárias (12 suplementares e 21 principais) envolvendo uma amostra total de 68 artigos dos cóloquios ASI_5, ASI_6 e ASI_7.

Para o nível de confiança de 0,85, identificamos 14 variáveis em 18 trajetos implicativos. Encontamos apenas três trajetos (4, 7 e 16) que contemplaram três váriáveis, os demais caminhos envolveram somente duas variáveis, conforme ilustra o quadro 3.

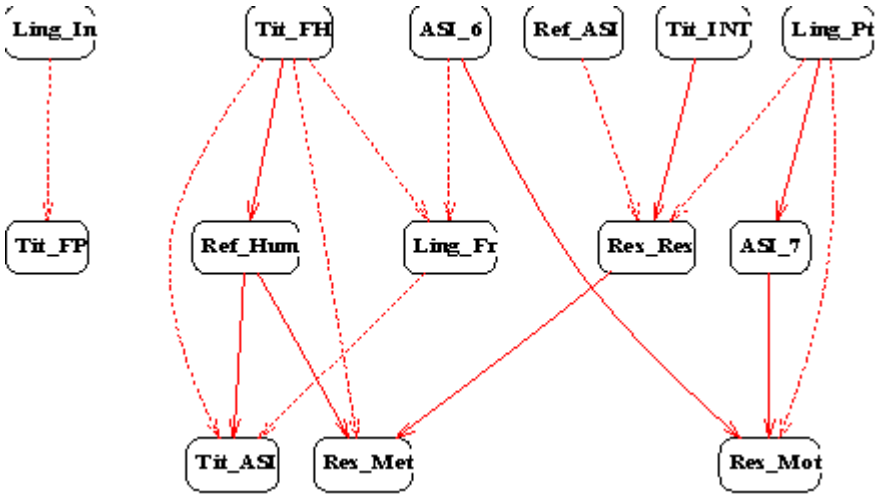


Figura 3- Gráfico implicativo

Destacamos baixo os caminhos 4 e 7 porque apresentam o percurso das variávies que se referem à Formação Humana.

O caminho 4, que envolve as variáveis principais Tit_FH ; Ref_Hum et Res_Met é inflenciado pelas variáveis suplementares : Nac_Br (risco : 0.0143) ; Femi (risco : 0.0528) e Inst_Fr (risco : 0.0528), embora a variável Inst_BR apresenta um fator de rico muito próximo, de 0.0631. A variável Nac_Br é a que mais contribui neste percurso. Isso sugere que os artigos relativos à formação humana apresentam uma referência bibliográfica voltada para essa temática (FH), destacando nos resumos os procedimentos metodológicos utilisados. Ao mesmo tempo, tais artigos indicam que as pesquisas empregando a ASI e o CHIC nas temáticas relacionadas à formação humana tem sido desenvolvidas com maior contribuição por pesquisadores de nacionalidade brasileira, de sexo feminino, vinculados as instituições francesas e brasileiras. Encontramos nisso a parceria das universidades francesas e brasileiras, realizando trabalhos de cotutela e pesquisa conjunta nos campos da educação como processo de desenvolvimento humano e integral.

O caminho 7, que envolve as variáveis principais Tit_FH ; Ling_Fr et Tit_ASI é inflenciado pelas variáveis suplementares Nac_Br (risco : 0.0436) ; Femi (risco : 0.0747) e Inst_Fr (risco : 0.0747). A variável Nac_Br é a que mais contribui neste percurso. Isso sugere que os artigos relativos à formação humana são escritos predominantemente em francês e têm em seus títulos uma indicação direta da ASI. Ao mesmo tempo, assim como no caminho 4, tais artigos indicam que as pesquisas empregando a ASI e o CHIC nas temáticas relacionadas à formação humana tem sido desenvolvidas com maior contribuição por pesquisadores de nacionalidade brasileira, de sexo feminino, vinculados as instituições francesas. Encontramos nisso a parceria das universidades francesas e brasileiras, realizando trabalhos de cotutela e pesquisa conjunta nos campos da educação como processo de desenvolvimento humano e integral.

Item	Caminho	Variável que mais contribui	Risco
1	Ref_Hum-Res_Met	Nac_BR	0,00592
2	Ref_Hum-Tit_ASI	Inst_Fr	0,00923
3	Tit_FH-Res_Met	Inst_Fr	0,0256
4	Tit_FH-Ref_Hum-Res_Met	Nac_BR	0,0143
5	Tit_FH-Ref_Hum-Tit_ASI	Nac_BR	0,0436
6	Tit_FH-Ref_Hum	Nac_BR	0,0143
7	Tit_FH-Ling_Fr-Tit_ASI	Inst_Fr	0,02
8	Tit_FH-Ling_Fr	Inst_Fr	0,02
9	Tit_FH-Tit_ASI	Inst_Ou	0,0369
10	Ling_In-Tit_FP	Inst_Fr	0.000404
11	ASI_6-Ling_Fr	Inst_Fr	0,0078
12	ASI_6-Res_Mot	Nac_Fr	0,0363
13	Ref_ASI-Res_Res	Nac_Ou	0,0158
14	Tit_INT-Res_Res	Inst_Br	0,00873
15	Ling_Pt-Res_Res	Inst_Ou	0,00487
16	Ling_Pt-ASI_7-Res_Mot	Inst_Br	0,00971
17	Ling_Pt-ASI_7	Inst_Br	0,00127
18	Ling_Pt-Res_Mot	Nac_Br	0

Quadro 3- Caminhos de implicação

7 Considerações Finais

Os resultados possibilitaram identificar que os colóquios têm contribuído a formação dos pesquisadores, principalmente no que se refere a formação acadêmica-profissional, voltada para as metodologias de pesquisa, expandindo e refinando o uso da abordagem da A.S.I a partir de estudos teóricos mais avançadados nas áreas da matemática e estatística.

Por outro lado, percebemos através da análise estatística implicativa e com o auxílio do software CHIC, as relações hierárquicas implicativas existentes entre variáveis Tit_FH e Ref_FH, o que nos sugere uma compreensão de que a Formação Humana também é escopo dos colóquios da A.S.I, o que é evidencindo com o número crescente

de publicações e de pesquisadores das diversas áreas que contribuem para o avanço dos estudos voltados para a Formação Humana dos educadores.

Mais especificamente, podemos perceber que os artigos relativos à formação humana são escritos predominantemente em francês, têm em seus títulos uma indicação direta da ASI, apresentam uma referência bibliográfica voltada para essa temática (FH), destacando nos resumos os procedimentos metodológicos utilisados.

Isso sugere que tais artigos indicam que as pesquisas empregando a ASI e o CHIC nas temáticas relacionadas à formação humana têm sido desenvolvidas com maior contribuição por pesquisadores de nacionalidade brasileira, de sexo feminino, vinculados as instituições francesas e brasileiras. Encontramos nisso a parceria das universidades desses dois paises, realizando trabalhos de cotutela e pesquisa conjunta nos campos da educação como processo de desenvolvimento humano e integral.

Agradecimentos

Agradeço ao Professor Jean Claude Regnier por ter me apresentado o marco teórico e aplicativo da Análise Estatística Implicativa, através de aulas e seminários no Brasil e na França, integrando-me ao grupo de pesquisa da ASI Franco-Brasileiro com apoio do programa Ciência sem fronteiras da CAPES que convidou o Professor enquanto PVE – Pesquisador Visitante Especial. Isso me proporcionou inserir essa abordagem na metodologia da minha tese de doutorado.

Referências

[1] Bardin, L. (1977). *Análise de Conteúdo*. Lisboa : Edições 70, LDA.

[2] Couturier, R., Bodin, A., & Gras, R. (2012). A classificação hierárquica implicativa e coesiva. *Manual do CHIC versão, 6*.

[3] Creswell, J. W. (2010). *Projeto de Pesquisa: métodos qualitativo, quantitativo e misto*. Porto Alegre : Artmed.

[4] Ferreira, A. L. (2010). Espiritualidade e Educação: um diálogo sobre quão reto é o caminho da formação humana. In: RÖHR, F. (Org.). *Diálogos em Educação e Espiritualidade*. Recife: Editora Universitária da UFPE, 109-159

[5] Freitas, A. S. (2005). *Fundamentos para uma Sociologia Crítica da Formação Humana*. Tese de Doutorado em Ciências Sociais. Universidade Federal de Pernambuco.

[6] Freitas, A. S. (2010). O “cuidado de si” como articulador de uma nova relação entre educação e espiritualidade. In: Röhr, F. (Org.). *Diálogos em Educação e Espiritualidade*. Recife: Editora Universitária da UFPE, 53-80

[7] Gadamer, H. G. (2008). *Verdade e Método I: Traços fundamentais de uma hermenêutica filosófica*. 10 ed. Petrópolis : Vozes.

[8] Gras, R., Régnier J.-C. & Guillet F. (Eds). (2009). *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*. RNTI-E-16 Toulouse : Cépaduès

[9] Gras, R., Régnier, J.-C., Marinica, C. & Guillet, F. (2013). *L'analyse statistique implicative*. Toulouse : Cépaduès-Éditions.

[10] Jaeger, W. (2003). *Paidéia: A formação do homem grego*. Tradução Artur M. Parreira.4ª. ed. São Paulo : Martins Fontes

[11] Lerman, I.-C. (1981). *Classification et analyse ordinale des données*. Paris : Dunod

[12] Pagni, P. A. & Silva, D. J. (2007). *Introdução à Filosofia da Educação: Temas Contemporâneos e História*. São Paulo: Avercamp

[13] Régnier, J-C., Spagnolo, F., Di Paola, B. & Gras, R. (Eds) (2010). *Analyse statistique implicative: Objet de recherche et de formation en analyse de donnés outil pour la recherche multidisciplinaire. V Colloque International – V International Conference A.S.I. Analyse Statistique Implicative – Statistical Implicative Analysis*. Palermo (Itália).

[14] Régnier, J-C. Bailleul, M. & Gras, R. (Eds) (2012). *L'analyse statistique implicative: de l'exploratoire au confirmatoire. VI Colloque International – VI International Conference A.S.I. Analyse Statistique Implicative – Statistical Implicative Analysis*. Caen (France).

[15] Régnier, J-C., Almouloud, S. & Gras, R. (Eds) (2014). *Analyse statistique implicative: Cadre théorique et applicatif por l'exploration semantique et non symetrique des donnés.. VII Colloque International – VII International Conference A.S.I. Analyse Statistique Implicative – Statistical Implicative Analysis*. São Paulo (Brasil).

[16] Richardson, R. J. *et al*. (1999). *Pesquisa social: método e técnicas*. São Paulo: Atlas.

[17] Röhr, F. (1999). A multidimensionalidade na formação do educador. In: *Revista da Educação*. ACE. Ano 28, n.110, jan/mar, 1999, 100-108.

[18] Röhr, F. (Org). (2010). *Diálogos em Educação e Espiritualidade*. Recife: Ed. Universitária da UFPE.

[19] Santiago, M. B. N. (2010). Diálogo e Transcendência na visão educativa de Martin Buber. In: RÖHR, F. (Org.). *Diálogos em Educação e Espiritualidade.* Recife: Editora Universitária da UFPE, 161-198.

Anexo

Anexo 1 – Grade de análise e modelagem das variáveis

1ª Parte: Dados demográficos do 1º autor do artigo (variáveis suplementares)

1) Sexo: ( )Masculino ( )Feminino

2) Nacionalidade: ( )Brasileira ( )Francesa ( )Espanhola ( ) Italiana ( )Outra

3) Instituição filiada: ( )Brasileira ( )Francesa ( )Espanhola ( ) Italiana ( )Outra

2a Parte: Caracterização do artigo (varáveis principais)

4) O título desse artigo remete a:

( )Formação Profissional ( )Formação Humana ( )Interculturalidade ( )Cultura local

( )A.S.I. ( )C.H.I.C.

5) Quais são as palavras-chaves desse artigo?

6) O resumo desse artigo contempla:

( )Visão introdutória ( )Ref. Teórico ( )Metodologia ( )Resultados ( )Palavras-Chave

7) As referências bibliográficas desse artigo contemplam artigos de ASIs anteriores?:

8) As referências bibliográficas desse artigo contemplam a palavra formação humana?:

3ª parte: Colóquio (variáveis secundárias)

9) A quantidade de artigos produzidos

10) A quantidade de páginas produzidas

Anexo 2 – Tabela de dados

	VARIAVÉIS																																
ARTIGOS	Masc s	Femi s	Nac_Br s	Nac_Fr s	Nac_It s	Nac_Es s	Nac_Ou s	Inst_Br s	Inst_Fr s	Inst_It s	Inst_Es s	Inst_Ou s	Ling_Pt	Ling_Fr	Ling_It	Ling_Es	Ling_In	Tit_FP	Tit_FH	Tit_INT	Tit_CLO	Tit_ASI	Tit_CHIC	Res_Int	Res_Teo	Res_Met	Res_Ref	Res_PCh	Ref_ASI	Ref_FH	ASI 5	ASI 6	ASI 7
a01	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	0	1	0	1	1	0	1	0	1	0	0
a02	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	0	1	1	0	1	0	1	0	0
a03	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	0	0	0	1	0	0
a04	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	1	0	0
a05	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0
a06	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	0	1	1	0	0	0	1	0	0
a07	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0
a08	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	1	1	0	1	0	1	0	0
a09	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1	1	1	1	0	0	0	1	0	0
a10	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	1	0	0	0	0	1	0	0
a11	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	1	1	1	0	0	0	1	0	0
a12	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	1	1	1	0	0	1	1	0	0
a13	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	1	0	1	0	1	0	1	1	0	1	1	1	0	0
a14	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	0	0	1	1	0	0
a15	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0
a16	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	1	0	0	0	1	1	0	0
a17	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	1	1	1	0	0	0	1	0	0
a18	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0
a19	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0
a20	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	0	0	1	1	0	0
a21	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0
a22	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	1	0	1	0	0	0	1	1	0	0
a23	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0	1	0	1	0	0
a24	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0
a25	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0
a26	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	0	1	0	1	1	0	0
a27	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0
a28	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0
a29	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1	1	0	1	0	1	1	0	0
a30	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	0	1	0
a31	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	1	0	0	1	0
a32	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	1	0	0	0	1	0
a33	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	1	1	1	1	0	0	0	1	0
a34	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0
a35	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0
a36	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	1	1	1	1	0	0	0	1	0
a37	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0
a38	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	1	0	1	0	0	0	1	0
a39	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	0	0	1	0
a40	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1	0	1	1	1	0	0	0	1	0
a41	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0
a42	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	1	0
a43	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	0	1	1	0	0	1	0
a44	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0
a45	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0
a46	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	0	1	1	1	1	1	0	0	0	1	0
a47	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	0
a48	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	0	0	1	0
a49	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	1	1	0	0	0	0	1
a50	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1
a51	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	0	1	0	0	1
a52	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	1	0	1	0	0	1
a53	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	1	1	1	1	1	1	0	0	1
a54	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1
a55	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	0	0	0	1	0	0	0	0	1
a56	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	1	0	1	1	1	0	1	0	0	0	0	1
a57	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1	0	0	0	1
a58	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	1	1	1	1	1	0	0	1
a59	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0	0	0	1
a60	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	1	1	1	1	1	0	0	0	0	1
a61	0	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	1	1	1	1	0	0	0	1
a62	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	1	1	1	0	0	0	0	1
a63	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	1	0	0	1	0	0	0	0	1
a64	1	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	1	0	1	0	1	0	0	0	0	1
a65	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	1	1	0	1	0	1	1	1	0	1	0	0	1
a66	1	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	0	0	0	0	1
a67	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1
a68	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	1	1	0	1	0	0	0	0	1

EXPECTATIVAS DE INSERÇÃO NO MERCADO DE TRABALHO DOS ESTUDANTES DA UNIVERSIDADE FEDERAL DA BAHIA-BRASIL: UM ESTUDO NO QUADRO DA ANALISE ESTATISTICA IMPLICATIVA.

Andréia CARDOSO SILVEIRA,[1] Jean-Claude REGNIER,[2] Nadja Maria ACIOLY-REGNIER,[3] e Robinson Moreira TENÓRIO,[4]

ATTENTES D'INSERTION DANS LE MARCHÉ DU TRAVAIL DES ÉTUDIANTS À L'UNIVERSITÉ FÉDÉRALE DE BAHIA - BRÉSIL: UNE ÉTUDE DANS LE CADRE DE L'ANALYSE ESTATISTIVA IMPLICATIVE

EXPECTATIONS OF INSERTION IN THE LABOR MARKET OF THE STUDENTS OF THE FEDERAL UNIVERSITY OF BAHIA-BRAZIL: A STUDY IN THE FRAMEWORK OF ANALYSIS ESTATISTIVA IMPLICATIVE

RESUMO

Este artigo objetiva discutir questões referentes às expectativas de inserção no mercado de trabalho dos estudantes, cotistas e não cotistas, de cursos de alto e baixo prestigio social, da Universidade Federal da Bahia (UFBA), Salvador - Brasil. Para isso, apresenta e analisa parte dos resultados de um estudo de doutorado, em andamento, intitulado "Expectativa e estratégias de inserção no mercado de trabalho dos estudantes cotistas e não cotistas da UFBA", cujos dados têm sido tratados por meio do software Classificação Hierárquica, Implicativa e Coesitiva (CHIC). Como resultado parcial, é possível observar que as diferenças de expectativas profissionais dos estudantes não apresentam relação com a modalidade de ingresso na universidade (cotas ou sistema tradicional), mas com o prestígio social dos cursos, pois os estudantes dos cursos de alto prestígio mostraram maiores expectativas em relação àqueles inseridos nas formações de baixo prestígio.

Palavras chaves: *Estudantes, Expectativas, Universidade, Mercado de trabalho.*

RESUME

Cet article a pour but de discuter des questions relatives aux attentes liées à insertion dans le marché du travail des étudiants, qui relèvent ou non de la politique des quota, dans les formations à bas ou haut prestige social au sein de l'Université Fédérale de la Bahia (UFBA), Salvador – Brésil. Pour cela, nous présentons et analysons une partie des résultats d'une étude en cours dans le cadre d'un doctorat sur le thème *Attentes et stratégies d'insertion dans le marché du travail des étudiants « quotistes » ou « non-quotistes » de l'UFBA*, dont les données ont été traitées au moyen du logiciel CHIC. Comme résultats partiel, il est possible d'observer que les différentes attentes professionnelles des étudiants

[1] Universidade Federal da Bahia; Brasil; Membro do Projeto "Determinantes da Equidade no Ensino Superior" financiado pelo Programa OBEDUC/CAPES; Universidade Lumière Lyon 2 ; França; andreia-car@hotmail.com

[2] Pesquisador Visitante Especial PVE/CAPES no PPGEC/UFRPE (Programa de Pós-Graduação em Ensino das Ciências da Universidade Federal Rural de Pernambuco) – UMR5191 – ICAR Université Lumiere Lyon 2 (FRA) jean-claude.regnier@univ-lyon2.f

[3] ESPE - Université Claude Bernard Lyon 1; France; nadja.acioly-regnier@univ-lyon1.fr

[4] Universidade Federal da Bahia; Brasil ; Coordenador do Projeto "Determinantes da Equidade no Ensino Superior" financiado pelo Programa OBEDUC/CAPES ; robinson.tenorio@uol.com.br

ne présentent pas de lien avec la modalité d'entrée à l'Université, mais avec le prestige social des formations, car ceux des filières prestigieuses manifestent les attentes les plus élevées par rapport aux étudiants des formations moins prestigieuses.

Mots clés: *Étudiants, Attentes Université, Marché du travail*

1 Introdução

As representações elaboradas sobre o negro, na cultura brasileira, não se diferenciam muito daquelas construídas em sociedades que se desenvolveram economicamente por meio da mão de obra escrava. Nessas sociedades, o negro, sujeito imprescindível para o alargamento e conglomeração do capital, passou a ser mão de obra barata, em objeto a ser aproveitado nos engenhos, nas minas e nas fábricas, tendo a sua importância avaliada a partir do que valia enquanto produto de troca (Fonseca, 2006).

No Brasil, a destituição do escravismo, em 1888, deixou os ex-escravos no patamar mais desprezível da hierarquia socioeconômica. Esses sujeitos, mal adaptados tanto Social quanto economicamente e despreparados para exercer suas funções enquanto homens emancipados foram sendo excluídos do mercado de trabalho (Hansebalg, 2005). Esse foi o preço que o ex-escravo teve que pagar por ter adquirido sua liberdade e por não estar preparado para trabalhar como sujeitos livres de um novo modelo de mercado que estava surgindo (Fonseca, 2006).

Em consequência desse processo, os negros ficaram assinalados por uma série de práticas racistas surgidas pelas circunstâncias em que ficaram "umbilicalmente" ligados às atividades que, na visão de Fonseca (2006, p. 97), "passaram a ser definidas, na grande maioria das vezes, pela cor de quem executa". Por isso, ao longo da formação da sociedade brasileira, a população negra permaneceu distinguida por marcas que determinam e explicam o seu comparecimento em ocupações desvalorizadas, mal recompensadas e que dificultam a sua incorporação à sociedade da qual faz parte. "A cor da pele, mesmo quando esmaecida pelo mito da harmonia social", é frequentemente recorrida a um sistema de relações em que ser negro permanece a ser constituído por formas de despersonalização edificadas pelos pressupostos escravocratas (Fonseca, 2006, p. 97).

Assim, o período escravista pode ser assinalado como uma fase que ocasionou grandes implicações para o negro que, no seu cotidiano enfrenta situações de disparidades, tanto racial como social. Segundo informações divulgadas em 2013 pelo Departamento Intersindical de Estatística e Estudos Socioeconômicos (Dieese), apesar dos avanços em relação à maior ocupação e menor desemprego dos negros (48.2%) nas regiões metropolitanas no biênio 2011 - 2012 (Belo Horizonte, Distrito Federal, Fortaleza, Porto Alegre, Recife, Salvador, São Paulo), essa população continua tendo sua força de trabalho menos valorizada. Suas remunerações por hora de trabalho, em média, ficaram restringidas a 63,9% do valor pago aos não negros.

Regiões Metropolitanas	Proporção de negros na população ocupada	Rendimento/hora (Em R$ de junho de 2013)		
		Negros	Não negros	Negros/Não negros
Belo Horizonte	64,0	7,68	10,98	69,95
Distrito Federal	68,6	11,32	17,35	65,24
Fortaleza	75,6	5,47	7,23	75,66
Porto Alegre	11,9	6,61	9,29	71,15
Recife	70,2	5,26	8,07	65,18
Salvador	89,2	5,89	9,84	59,86
São Paulo	34,0	6,96	11,40	61,05
Total	**48,2**	**6,83**	**10,69**	**63,89**

Tabela 1. Proporção de negros na população ocupada e rendimento hora dos ocupados negros e não negros. Regiões Metropolitanas[5] – Biênio 2011 – 2012 **Fonte:** Dieese /Seade, TEM/FAT e entidades regionais. PED – Pesquisa de Emprego e Desemprego.

Observa-se também que na estrutura setorial de ocupação metropolitana (segmentos: Apoio, Execução e Direção e Planejamento) a participação dos pretos e pardos esteve mais concentrada nas ocupações de Execução e "escassamente" inseridos em postos de Direção e Gestão.

Setor de Atividade	Belo Horizonte		Distrito Federal		Recife		Salvador		São Paulo	
	Negro	Não negro	Negro	Não negro	Negro	Não negro	Negro	Não negro	Negro	Não negro
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Direção e Planejamento	11,3	23,2	15,7	25,5	8,3	18,6	6,9	21,3	5,7	18,1
Execução	59,4	50,2	51,4	42,3	61,8	51,9	66,2	53,0	61,1	52,1
Apoio	23,9	21,6	31,0	30,8	23,7	25,1	22,5	20,7	24,7	23,3
Ocupações mal definidas	5,4	5,0	1,9	1,4	6,1	4,4	4,4	5,0	8,4	6,6

Tabela 2. Distribuição dos Ocupados por Cor, segundo Grupos de Ocupações Selecionados. Regiões Metropolitanas – Biênio 2011 – 2012 **Fonte:** Dieese /Seade, TEM/FAT e entidade regionais. PED – Pesquisa de Emprego e Desemprego

Nesse contexto, os dados apresentados são desafiadores e nos direcionam a refletir sobre algumas questões colocas por Osório (2008): Como tais disparidades podem permanecer numa sociedade em que os indivíduos estão convencidos "da falácia da raça"? Quando as pessoas enfatizam que não existe raça e que somos de uma mesma espécie humana, em que práticas racistas são crimes inafiançáveis? (Osório, 2008).

No Brasil, um número significativo de políticas de combate ao racismo têm sido implementados no combate as desigualdades raciais; entre elas destacam-se as Políticas Afirmativas que se constituem num conjunto de ações "públicas e privadas, de caráter compulsório (obrigatório), facultativo (não obrigatório) ou voluntário que tem como objetivo corrigir as desigualdades historicamente impostas a determinados grupos sociais e/ou étnico/raciais" (Munanga; Gomes, 2006, p. 186).

Entre as diversas ações[6] que compõem as Políticas Afirmativas, destacam-se as Ações Afirmativas de Promoção ao Ensino Superior que se constitui numa ação governamental que institui a importância das instituições públicas e privadas reservarem vagas para aqueles indivíduos menos favorecidos, como o índio, o portador de necessidades especiais, os alunos de escolas públicas, bem como para os negros. Assim, trata-se de uma ação que tem o objetivo de atenuar as dificuldades de ingresso desses

[5] Corresponde ao total das Regiões Metropolitanas de Belo Horizonte, Distrito Federal, Fortaleza, Porto Alegre, Recife, Salvador, São Paulo. Obs. Raça/cor negra = pretos e pardos; raça/cor não negra = brancos e amarelos

[6] Outras ações, também importantes, tem se constituído em ações afirmativas, como Bolsas de estudos para aqueles que desejam se preparar para concursos, curso pré-vestibular para negros e indígenas e, ainda, cotas para ingresso em instituições de trabalho públicas.

indivíduos no ensino superior, equilibrando e tornando mais "equitativo o sistema meritocrático", haja vista que no Brasil a competência intelectual e a condição financeira são critérios imprescindíveis para o acesso à universidade (Munanga; Gomes, 2006).

Vale destacar que as Políticas Afirmativas de acesso ao ensino apresentam como objetivo principal a redução das desigualdades no acesso a educação superior, porém ela carrega outro importante intuito que é o alcance da equidade no acesso ao mercado de trabalho. Trata-se de proporcionar aos negros a oportunidade de ter uma profissão, mas também de exercê-la. Todavia, infelizmente, ainda não há estudos que mostrem se tal objetivo tem sido alcançado. A maioria das pesquisas referentes ao sistema de reserva de vagas tem focado em torno do acesso e permanência desses sujeitos nas universidades.

Neste contexto, é que foi proposto a tese de doutorado "Expectativa, estratégias e alcances no acesso ao mercado de trabalho: um estudo com estudantes cotistas e não cotistas da Universidade Federal da Bahia". Essa tese busca responder algumas questões como: quais são as expectativas e estratégias de inserção profissional dos estudantes negros cotistas? Os cotistas tem se inserido no mercado de trabalho? Estes alunos têm ocupado postos de trabalhos que correspondem as suas formações? Possuem empregos estáveis? Qual é a renda desses indivíduos? Existem diferenças entre negros cotistas e brancos não cotistas no acesso ao mercado de trabalho?

Para realização dessa pesquisa, foi selecionada uma amostra de estudantes formandos cotistas e não cotistas da UFBA, inseridos em cursos de alto e baixo prestigio social para aplicação de um questionário, seis meses antes da saída da universidade e seis meses após. Os dados coletados estão sendo tratados por meio do método estatístico muldimendional, viabilizado pelo software Classificação Hierárquica, Implicativa e Coersitiva (CHIC). No momento, a coleta de dados está em andamento, porém alguns resultados sobre as expectativas profissionais já podem ser apresentados. Assim, apresenta-se o presente artigo que objetiva discutir questões concernentes às expectativas de inserção no mercado de trabalho dos formandos negros cotistas e brancos não cotistas da UFBA de cursos de alto e baixo prestigio social.

Compreende-se que as expectativas profissionais são ideias que o individuo tem a respeito do que vai acontecer em relação ao seu acesso ao mercado de trabalho. Sabe-se que todos trazem consigo expectativas que surgem antes da entrada na universidade e que vão sendo ampliadas, fortalecidas ou redefinidas durante o percurso de formação profissional. Tais expectativas, muitas vezes, decorrem das interações sociais realizadas no ambiente familiar, no grupo de amigos e, principalmente, na universidade, uma vez que está última além de ensinar, de preparar seus estudantes para a vida profissional, se constitui ainda em um espaço de criação de laços que vão interferir no processo de inserção no mercado de trabalho.

Nessa direção, o artigo foi estruturado em cinco partes, incluindo esta introdução. Na segunda parte, discorre-se sobre o combate as desigualdades raciais no Brasil, inclusive aborda-se as políticas afirmativas de acesso ao ensino superior. Em seguida, apresenta o percurso metodológico da pesquisa, resultados e discussões. E, por fim, as considerações finais.

2 O combate às desigualdades raciais no Brasil

No Brasil, embora a temática racial seja um objeto de abrangência e importância restrita, as discussões no que se refere a temática têm avançado significativamente. Nessa conjuntura, observa-se que no decorrer dos últimos 20 anos, a abordagem da questão das disparidades raciais alcançou importante espaço na agenda do debate público. O decênio de 1980 e 1990 foi assinalado por um amplo debate em torno da questão de existência ou não do racismo. Nesse período, o paradigma da democracia foi colocado em questão e o reconhecimento das desigualdades raciais, bem como suas causas foram identificadas como objeto de reflexão que precisava se materializar. Assim, de maneira evolutiva, a partir dos anos 90, o debate foi modificado e o reconhecimento das disparidades existentes entre negros e brancos se efetivou e inseriu-se no âmbito das ações de políticas públicas (Theodoro, 2008).

Sobre isso, Jaccoud (2008) destaca que entre os anos de 1980 e 2000, podem ser identificadas três gerações de empreendimentos de combate a discriminação racial. A primeira se desenvolveu no período em que o Brasil presenciava um momento de redemocratização e de mobilização social, direcionada pela luta por direitos civis, políticos e por justiça social. Faz parte desse período a reogarnização dos movimentos sociais que trouxeram para a arena política a discussão sobre a discriminação racial. O surgimento de conselhos e órgãos de assessoria em diversos municípios e estados do país (São Paulo, Minas Gerais, Rio Grande do Sul, Mato Grosso do Sul, Distrito Federal e Bahia) com o objetivo de suscitar a participação do segmento negro. E, por fim, em 1988, a criação da Fundação Cultural Palmares (FCP)[7].

A respeito da segunda geração, essa surge no final da década de 1980 e pode ser caracterizada pela intensa luta de combate a discriminação e o racismo através da sua criminalização. Nesse mesmo contexto, atendendo as solicitações do movimento negro, a Constituição de 1988[8] classifica o racismo em crime inafiançável e imprescritível. Logo, em 1989 ocorre a promulgação da Lei Caó[9] que determina como crime de racismo atitudes que inibi ou impede a entrada em lugares público-comerciais, a empregos, por origem de cor/raça. Após a implementação da referida lei, outras resoluções legais com fins de afligir crimes referentes a discriminação foram implementadas (Jaccoud, 2008).

Por volta da década de 1990, surge a terceira geração[10] de promoção da igualdade racial, assinalada pela ampliação de ações de combate a discriminação através de

[7] "Organismo Federal voltado a promoção e a preservação da influencia negra na sociedade brasileira, ligada ao Ministério da Cultura, tinha também entre seus objetivos a identificação das comunidades remanescentes de quilombo e o apoio a demarcação e titulação de suas terras" (Silva, et. al, 2009, p. 26) .

[8] "A Constituição deCorresponde ao total das Regiões Metropolitanas de Belo Horizonte, Distrito Federal, Fortaleza, Porto Alegre, RecCorresponde ao total das Regiões Metropolitanas de Belo Horizonte, Distrito Federal, Fortaleza, Porto Alegre, Recife, Salvador, São Paulo. Obs. Raça/cor negra = pretos e pardos; raça/cor não negra = brancos e amarelosife, Salvador, São Paulo. Obs. Raça/cor negra = pretos e pardos; raça/cor não negra = brancos e amarelos 1988 tratou com destaque os temas da discriminação racial, da diversidade cultural e do conhecimento dos direitos da população remanescente de quilombos [...] reconheceu o racismo como crime inafiançável e imprescindível [...] afirmou a diversidade cultural como patrimônio comum a ser valorizado e preservado" (SILVA, et. al, 2009, p. 27).

[9] Lei n° 7716 de autoria do deputado federal Carlos Alberto Oliveira.

[10] Em 2003 é criado a Secretaria Especial de Políticas de Promoção da Igualdade Racial - SEPPIR, responsável pela construção e direção das políticas de combate a discriminação racial.

políticas públicas. Sobre essa geração, destacam-se 4 (quatro) ações implementadas no âmbito federal, a saber: 1. Ações de implementação da Lei nº 10.639/2003, que institui a obrigatoriedade da inserção no currículo do Ensino Básico do estudo da História e da Cultura Afro-brasileira; 2. Programa de combate ao Racismo Institucional - PCRI; 3. Programa de Promoção da Igualdade de Oportunidade para Todos, sob administração do Ministério Público do Trabalho – MPT; 4. Ações Afirmativas de ingresso ao Ensino Superior (Silva, et. al 2009).

Tendo em vista o foco de discussão do presente artigo, far-se-á uma breve apresentação das 3 (três) primeiras ações e, posteriormente, como maior ênfase será abordada as Ações Afirmativas de acesso a Universidade.

A primeira ação foi a implementação da Lei 10.639/2003[11]. A citada Lei instituiu a obrigatoriedade do estudo da História e da Cultura Afro-Brasileira no domínio de todo currículo da Educação Básica, especificamente nas disciplinas de Literatura, História e Arte. Trata-se de uma ação que objetiva cooperar para a exclusão de práticas discriminativas, preconceituosas, auxiliar na produção e dinamização do conhecimento que engloba a história e as lutas do processo de formação do nosso país e na valorização da diversidade cultural (Jaccoud, 2008).

A segunda ação apresentada por Jaccoud (2008, p. 142) foi o Programa de Combate ao Racismo Institucional - PCRI [12]. Estruturado para funcionar durante o período de 2005 a 2006, o referido programa teve dois objetivos: "fortalecer a capacidade do setor público na identificação e prevenção do racismo institucional e fomentar a participação das organizações da sociedade civil organizada no diálogo sobre políticas públicas". Na esfera Federal, o programa foi direcionado pelo Ministério da Saúde e intencionou reduzir as disparidades raciais existentes nesse setor e promover a equidade no Sistema Único de Saúde (SUS).

Outra ação apresentada por Jaccoud (2008, p. 142) foi o Programa de Promoção da Igualdade de Oportunidades para Todos, do Ministério do Trabalho - MPT, lançado em 2005. O referido programa objetivou suprimir a discriminação de raça e gênero nas relações de trabalho. Através desse programa busca-se a sensibilização das empresas que exercem atitudes discriminativas a tomarem medidas que possam extinguir tais comportamentos e implementar ações afirmativas de transformação das desigualdades (Jaccoud, 2008). Até o momento, as contribuições do citado programa foi a fiscalização de práticas discriminativas em alguns setores da economia como bancos e supermercados.

E, por fim, a quarta ação foi as Políticas Afirmativas de Promoção ao Ensino Superior. Sobre isso, apresenta-se o seguinte sub-tópico.

---

[11] "Fruto do Projeto de Lei (PL) N. 259/1999, de autoria da então deputada Esther Grossi, altera a Lei de Diretrizes e Bases (LDB), visando efetivar e garantir o cumprimento dos artigos constitucionais 206 e 2010, que determinam que o ensino no Brasil deva se pautar pelo pluralismo e respeito aos valores culturais do país" (Silva, et.al, 2009).

[12] Formulado por duas instituições internacionais. Programa das Nações Unidas para o Desenvolvimento – PNUD; Agência de Cooperação Técnica do Ministério Britânico para o Desenvolvimento Internacional e Redução da Pobreza – DFI (Jaccoud, 2008, p. 143).

2.1 Políticas Afirmativas e Reserva de Vagas nas Universidades Brasileira.

A discussão em torno da realização de Políticas Afirmativas no espaço educacional é, conforme Silva (et.al 2009), bastante remota. Já desde os anos de 1930 que os grupos do movimento social negro assinalavam sobre a importância de implementação de políticas públicas que garantisse o acesso dos negros a educação. Nesse período, tendo como pauta a educação, a Frente Negra Brasileira criou escolas comunitárias para crianças em suas sedes, bem como em algumas regiões do Brasil. No período de 1940 e 1950, logo após o término da ditadura do Estado Novo de Vargas que tornou ilegal as ações do movimento social, esse último reviu sua pauta de discussão e outra vez trouxe para o debate o tema educacional. De 1960 a 1970, observou-se a ampliação das discussões em torno do referido tema, mas só a partir do final dos anos 1990 que tais trabalhos se materializaram no espaço brasileiro, inclusive com a efetivação das Políticas Afirmativas (Silva, et.al. 2009).

Sobre Políticas Afirmativas, Silva (et.al. 2009), alerta que com frequência ocorrem confusões entre o conceito de ações afirmativas e cotas raciais ou sociais. As Ações Afirmativas constituem-se em "ações públicas ou privadas, de caráter compulsório (obrigatório), facultativo (não obrigatório), voluntário que têm como objetivo corrigir as desigualdades historicamente impostas a determinados grupos sociais e/ou étnico/raciais com um histórico comprovado" de marginalização (Munanga e Gomes, 2006, p. 186). Para a implementação das Políticas Afirmativas, estabelecem um "tratamento diferenciado" que incorporam diversas formas. As ações de combate ao racismo, citadas anteriormente, bem como as ações de reservas de vagas nas universidades se constitui em Políticas Afirmativas, ou seja, em recursos de reparação aos efeitos das práticas discriminativas (Silva, et.al. 2009).

Assim, as ações afirmativas no ensino superior são iniciativas que objetivam promover o acesso a universidade de estudantes pertencentes a grupos socialmente excluídos. Essas ações podem ser implementadas por meio de diferentes recursos, podendo ou não se estruturada por meio de reserva de vagas. Um exemplo são as experiências de cursinhos preparatórios para o vestibular direcionados a jovens negros. Porém, vale destacar que a Política de Reserva de Vagas tem sido o principal tipo de ação afirmativa implementada nas universidades brasileira (Silva, et.al. 2006).

Vale lembrar, que a Política de Cotas, atualmente obrigatória para as universidades e institutos federais, já foi implementada em mais 60 universidades (estaduais e federais) brasileiras. Destas instituições, as pioneiras foram a Universidade Estadual do Rio Janeiro – UERJ e a Universidade Estadual do Norte Fluminense – UENF (SILVA, et.al. 2009). Outra instituição, também precursora das Políticas de Reserva de Vagas, foi a Universidade Federal da Bahia - UFBA, que a partir do ano de 2005, vêm reservando um número significativo de suas vagas para não brancos egressos de escolas públicas.

Na UFBA, a referida política foi conjeturada a partir de 1998, em um documento construído pelo Centro de Estudos Afro-Orientais – CEAO, mas foi apenas em 2001 que a proposta adquiriu espaço na Universidade, quando um grupo de estudantes não brancos, vinculados ao Diretório Central dos Estudantes – DCE, fez uma manifestação em favor do Sistema de Reserva de Vagas. Como resultado, em 2003, juntamente com o CEAO, o Comitê Pró-Cotas da UFBA enviou à Reitoria uma proposta de criação de uma equipe específica para trabalhar a temática. Aprovada a formação da equipe, em 2004, o Conselho de Ensino, Pesquisa e Extensão – CONSEPE, juntamente com o

Conselho Universitário, ambos deferiram uma proposta do então Reitor Prof. Naomar de Almeida, permitindo que a Universidade Federal da Bahia implantasse o Sistema de Reserva de Vagas no vestibular do ano de 2005 (Anunciação, 2005).

Conforme documentos disponíveis na página da UFBA, o Sistema de Reserva de Vagas da UFBA considera quatro categorias para a inscrição dos candidatos ao vestibular, cada uma com as respectivas porcentagens de vagas reservadas, nas duas fases:

Categoria A	Categoria B	Categoria C	Categoria D
candidatos de escola pública que se declararam não brancos ou pardos	candidatos de escola pública de qualquer etnia ou cor	candidatos de escola pública que se declararam índiodescendentes	reservados a todos os candidatos de qualquer procedência escolar, etnia ou cor
36,55%	6,45%	2%	55%):

Tabela 3. categorias para a inscrição dos candidatos ao vestibular.

Vale ressaltar a importância das ações de reserva de vagas nas universidades para a população negra, uma vez que a referida política tem permitido que esses jovens possam ingressar na universidade. Sob esta perspectiva, Sell (2002, p. 79) defende que as Políticas Afirmativas no ensino superior podem ser utilizadas com fins "eticamente aceitáveis" para a edificação de uma sociedade menos "assimétrica" no "acesso ao poder" e nas oportunidades surgidas. Ainda, destaca:

Não se trata de privilégios, mas sim da busca de realização da igualdade material, a partir de uma reinterpretação do conceito de igualdade formal. Atualmente, não-brancos e brancos são formalmente iguais e materialmente diferentes. A Ação Afirmativa os tornaria formalmente diferentes no presente (oferecendo certas preferências aos não brancos) para torná-los, no futuro, materialmente mais próximos dos brancos (Sell, 2002, p. 80).

Certamente, a políticas de reserva de vaga tem significantemente aproximado negros e brancos, uma vez que os primeiros têm tido a oportunidade de frequentar espaços universitários que até então eram reservados a elite branca brasileira. Mas, quais são os outros resultados da referida política? Antes de tentar responder tal questão, apresenta-se o tópico a seguir que aborda os procedimentos metodológicos da pesquisa, assim como alguns resultados da pesquisa de doutorado que fundamenta este artigo.

3 Aportes metodológicos

Como já destacado**,** a proposta de tese que fundamenta este artigo objetiva analisar as expectativas, alcances e estratégias de inserção no mercado de trabalho dos estudantes, cotistas e não cotistas, dos cursos de maior e menor prestígio social, da Universidade Federal da Bahia (UFBA) – Brasil, em fase de transição universidade-mercado de trabalho, no período de 2014-2015. Entende-se que a transição universidade-mercado de trabalho é o momento em que o estudante começa sua preparação para sair da universidade (último semestre do curso de graduação) e ingressar no mercado de trabalho (até seis meses após a conclusão do curso de nível superior).

Como a Medicina, o Direito e as Engenharias pertencem ao grupo de cursos historicamente valorizados e conhecidos como de maior prestigio social; os estudantes de tais cursos são identificados como pertencentes às áreas de "Alto Prestigio Social" – APS, e demais alunos como vinculados às áreas de "Baixo Prestígio Social" – BPS. Apesar de não ser objetivar neste trabalho discutir prestígio social dos cursos, acredita-se que os alunos oriundos das áreas APS, em comparação com os estudantes de áreas de BPS, possuem maior expectativa de inserção no mercado de trabalho. Dessa forma, duas importantes variáveis serão observadas, á variável modalidade de ingresso na universidade (Cotas e não cotas), assim como a variável prestígio do curso (APS, BPS). A seleção das referidas variáveis se justifica em função do recorte proposto neste artigo, bem como em função de ainda não ter sido finalizado a segunda etapa de coleta de dados.

Para o alcance do citado objetivo, desenhou-se um estudo em duas dimensões: uma primeira, de caráter quantitativo; a segunda, qualitativa. Sobre a abordagem qualitativa, é importante destacar a sua relevância para esta pesquisa, pois assegurar "que aspectos da vida social dos sujeitos sejam apreendidos para se trabalhar as percepções e motivações em suas relações e interações socioculturais". Os ambientes e as "temporalidades" singulares específicas de cada circunstância podem ser estudados e sistematizados por meio de passos científicos ou abordagens que buscam compreender o fenômeno investigado (Castro; Abramovay, 2006, p. 42). Já a abordagem quantitativa justifica-se, tendo em vista o emprego da quantificação, tanto na coleta de dados (Richardson, 1989), quanto na forma em que serão examinados os dados, como frequências e percentagens (Richardson, 1989).

3.1 Seleção da Amostra

Tendo em vista a numerosa quantidade de alunos formandos em 2014 (ano de realização da pesquisa), na Universidade Federal da Bahia – UFBA, selecionou-se 16 (dezesseis) cursos, a saber: Engenharia Elétrica, Ciência da Computação - bacharelado, Geologia, Engenharia de Minas (**Área I**: Matemática, Ciências Físicas e Tecnologias); Medicina, Oceanografia, Medicina Veterinária, Fonodialógia (**Área II**: Ciências Biológicas e Profissões da Saúde); Direito, Jornalismo, Secretariado e Arquivologia (**Área III:** Filosofia e Ciências Humanas); Design, Artes Cênicas-bacharelado, Artes Cênicas-licenciatura, Artes Plásticas - bacharelado (**Área IV**: Letras; **Área V)**. A seleção dos cursos ocorreu a partir das notas de maior e menor corte no vestibular em 2005, quando foi implementado a Política de Cotas na UFBA.

Conforme levantamento, foi obtido aproximadamente 517 formandos, cotistas e não cotistas em 2014. Desse total, 295 estudantes foram localizados, assim como apresentaram interesse em participar. Todavia, tendo em vista o foco deste artigo que é apresentar as expectativas de inserção no mercado de trabalho dos estudantes negros cotistas e brancos não cotistas[13], a amostra foi reduzida a 121 estudantes[14].

[13] Assim, foi eliminado os estudantes cotistas da Categoria A, B e C, bem como aqueles que se declaram de outras etnias e que ingressaram pelo sistema universal.

[14] Outra observação importante, é que os estudantes serão identificados como pertencente às áreas de maior ou menor prestigio social, sem identificar o curso.

3.2 Instrumento de coleta de dados

Em função do número expressivo de estudantes formandos, o que poderia inviabilizar o contato com todos os sujeitos, optou-se pela aplicação de questionário. Para isso, levantou-se os emails de todos os formandos e, posteriormente, foi enviado um link de acesso ao questionário. A ferramenta utilizada foi o Google docs que permite criar, editar e partilhar diversos documentos, a exemplo de questionários.

O referido questionário está sendo aplicado em momentos diferentes. No primeiro, quando o estudante encontra-se em fase de saída da universidade; e o segundo quando completa seis meses de conclusão do curso. Em relação ao primeiro momento, o questionário busca coletar informações que estejam relacionados, sobretudo, às expectativas profissionais. Já o segundo, visa verificar o alcance das expectativas (posição de ocupação dos egressos; natureza das atividades desenvolvidas pelo egresso na ocupação exercida; renda mensal, entre outros); e, as estratégias adotadas para ingresso no mercado de trabalho.

Neste momento, já foi concluído a primeira faze da pesquisa; e está sendo aguardado que os alunos completem seis meses de encerramento do curso, previsto para acontecer em outubro de 2015. É importante lembrar que, a partir revisão de literatura, o questionário foi construído com base em algumas variáveis e categorias que permitem a identificação das expectativas dos estudantes. Dessa forma o questionário abordou, entre outras questões, os seguintes assuntos: **1**. A posse de um plano profissional; **2**. atividade a ser realizada nos primeiros seis meses após a conclusão do curso; **3**. Nível da expectativa de inserção profissional em Salvador; **4.** Expectativa de tempo para inserção profissional; e, **5**. Expectativa salarial. Tais questões são apresentadas no quadro a seguir:

Questão 1
Você considera que já tem um plano profissional pós-formatura? **A** ☐ Sim. **B** ☐ Em parte. **C** ☐ Não.
Questão 2
O que você pretende fazer durante os PRIMEIROS SEIS MESES após a conclusão do curso de graduação? (**assinale a mais significativa**) **A** ☐ Apenas trabalhar **B** ☐ Apenas continuar os estudos **C** ☐ Continuar os estudos e trabalhar **D** ☐ Tirar um período de férias **E** ☐ Está em dúvida **F** ☐ Não sabe responder
Questão 3
Considerando a situação atual do mercado de trabalho da sua profissão e as sua possibilidades pessoais, quais são suas expectativas de inserção no mercado de trabalho nos PRIMEIROS SEIS MESES APÓS CONCLUSÃO DO CURSO, em Salvador: A ☐ Muito grande B ☐ Grande **C** ☐ Razoável **D** ☐ Pequena **E** ☐ Nenhuma
Questão 4
Dentro de quanto tempo você acredita que iniciará sua carreira profissional nessa profissão:

A ☐ Já trabalha na área profissional B ☐ Imediato **C** ☐ Até 6 meses **D** ☐ De 7 a 12 meses **E** ☐ De 13 a 18 meses **F** ☐ De 19 a 24 meses **G** ☐ Mais de 24 meses **H** ☐ Não sabe responder
Questão 5
Tendo em vista o contexto atual do campo de trabalho da sua área de formação, qual a sua perspectiva de remuneração salarial para os PRIMEIROS 6 (SEIS) MESES no exercício da sua profissão? **A** ☐ Nenhuma **B** ☐ Até 1,5 salários mínimos (até R$ 1.086,00). **C** ☐ De 1,5 a 3 salários mínimos (R$ 1.086,01 a R$ 2.172,00). **D** ☐ De 3 a 4,5 salários mínimos (R$ 2.172,01 a R$ 3.258,00). **E** ☐ De 4,5 a 6 salários mínimos (R$ 3.258,01 a R$ 4.344,00). **F** ☐ De 6 a 10 salários mínimos (R$ 4.344,01 a R$ 7.240,00). **G** ☐ De 10 a 30 salários mínimos (R$ 7.240,01 a R$ 21.720,00). **H** ☐ Acima de 30 salários mínimos (mais de R$ 21.720,01). **I** ☐ Não sabe responder

Quadro 1 - questões abordadas no questionário **Fonte**: Elaborado pelos autores.

3.3 Análises dos Dados

Como já destacado, os dados coletados neste estudo estão sendo tratados a partir do Software CHIC (Classificação Hierárquica Implicativa e Coesitiva) (Borges, 2009). O referido software interpreta resultados de estudos de análise estatística implicativa (A.S.I). Conforme Gras et.al (2013) a A.S.I menciona um campo teórico que foca o conceito de implicação estatística, mais precisamente o conceito de quase implicação, para que se possa diferenciá-lo da implicação lógica dos estudos da matemática. O estudo do conceito de quase implicação, nas áreas da probabilidade e da estatística, contribuiu para a elaboração de ferramentas teóricas que "instrumentam um método de análise de dados" (7º Colóquio Internacional sobre a Análise Estatística Implicativa).

A análise estatística implicativa:

> [...] destina-se a descobrir e a estruturar em forma de regras, um conjunto de dados cruzando sujeitos (ou objetos) e variáveis por meio de uma modelagem estatística de quase-implicação: se a variável ou uma combinação de variáveis a é observada na população, então, em geral, a variável b é também observada. As variáveis envolvidas podem ser de vários tipos: binário, modal, numérico, intervalar. Os conjuntos de regras obtidos podem ser estruturados de acordo com diferentes abordagens complementares (gráfico implicativo, hierarquia orientada) (7º Colóquio Internacional sobre a Análise Estatística Implicativa).

De acordo com Borges (2009, p.73) o CHIC foi construído por uma equipe de pesquisadores, sob a coordenação do professor Régis Gras, do Núcleo de Pesquisa e Didática da Matemática da Universidade de Rennes, na França, em 1992. Sua elaboração teve como objetivo sintetizar e organizar as respostas de professores para a obtenção de uma "tipologia de comportamentos", através de um "tratamento multidimensional de dados estatísticos". Este software permite extrair um conjunto de dados, com o cruzamento de sujeitos, variáveis, atributos, regras de associação entre

variáveis, dentre outros; faz análise de similaridades, de forma que seja possível visualizar semelhanças e classes de variáveis "em nível de uma árvore hierárquica" (Borges, 2009, p. 73). Diante disso, observa-se que o Software CHIC é uma importante ferramenta de análise de dados qualitativos, que será adotada neste estudo pela riqueza na análise dos dados que esse software permitirá alcançar. Após a apresentação dos passos metodológicos, apresenta-se o subtópico a seguir que discorre sobre os resultados alcançados.

4 Discussão e resultados

A tabela explorada pelo softawe CHIC foi constituída por 121 linhas (que retratam o número de sujeitos participantes do estudo) e 30 colunas (que tratam das variáveis). As variáveis trabalhadas são caracterizadas como: binárias (questões 1, 2 e 4); modais (questão 3); e, intervalar (questão 5). Estabeleceu-se como variáveis principais as que contemplam as expectativas profissionais dos estudantes e como variáveis secundárias as tocantes a modalidade de ingresso na universidade, assim como o curso em que o estudante está inserido.

A seguir, apresenta-se o quadro 2 que com as variáveis analisadas.

Variável	Código	Tipo	Descrição
V1	IngresST	Binária	Modalidade de ingresso na Universidade – estudantes que se declararam brancos e que ingressaram pelo "Sistema tradicional" de vestibular.
V2	IngresSC	Binária	Modalidade de ingresso na Universidade – estudantes de escola pública que se declaram pretos ou pardos e que ingressaram pelo "Sistema de cotas" de vestibular.
V3	CursAPS	Binária	Estudantes de cursos de alto prestígio social
V4	CursBPS	Binária	Estudantes de cursos de baixo prestígio social
V5	PlanP01	Binária	Estudantes que possuem um plano profissional
V6	PlanP02	Binária	Estudantes que possuem em parte um plano profissional
V7	PlanP03	Binária	Estudantes que não possuem um plano profissional
V8	ExpectA1	Binária	Expectativas para os primeiros seis meses após obtenção do diploma – interesse em apenas trabalhar
V9	ExpectA2	Binária	Expectativas para os primeiros seis meses após obtenção do diploma – interesse em apenas continuar os estudos
V10	ExpectA3	Binária	Expectativas para os primeiros seis meses após obtenção do diploma –interesse em continuar os estudos e trabalhar
V11	ExpectA4	Binária	Expectativas para os primeiros seis meses após obtenção do diploma – interesse em tirar um período de férias
V12	ExpectA5	Binária	Expectativas para os primeiros seis meses depois de saída da universidade – Está em dúvida
V13	ExpectA6	Binária	Expectativas para os primeiros seis meses após obtenção do diploma – Não sabe responder
V14	ExpectIP	Modal	Expectativas de inserção no mercado de trabalho de Salvador – Muito Grande, Grande, Razoável, Pequena, Nenhuma.
V15	ExpectICP1	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - Já trabalha na área profissional
V16	ExpectICP2	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional – Imediato
V17	ExpectICP3	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - Até 6 meses

V18	ExpectICP4	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - De 7 a 12 meses
V19	ExpectICP5	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - De 13 a 18 meses
V20	ExpectICP6	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - De 19 a 24 meses
V21	ExpectICP7	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - Mais de 24 meses - Não sabe responder
V22	ExpectICP8	Binária	Expectativas quanto ao tempo para iniciar a carreira profissional - Não sabe responder
V23	ExpectS1	Intervalar	Expectativa salarial – Nenhuma
V24	ExpectS2	Intervalar	Expectativa salarial - Até 1,5 salários mínimos (até R$ 1.086,00).
V25	ExpectS3	Intervalar	Expectativa salarial - De 1,5 a 3 salários mínimos (R$ 1.086,01 a R$ 2.172,00).
V26	ExpectS4	Intervalar	Expectativa salarial - De 3 a 4,5 salários mínimos (R$ 2.172,01 a R$ 3.258,00).
V27	ExpectS5	Intervalar	Expectativa salarial - De 4,5 a 6 salários mínimos (R$ 3.258,01 a R$ 4.344,00).
V28	ExpectS6	Intervalar	Expectativa salarial - De 6 a 10 salários mínimos (R$ 4.344,01 a R$ 7.240,00).
V29	ExpectS7	Intervalar	Expectativa salarial - De 10 a 30 salários mínimos (R$ 7.240,01 a R$ 21.720,00).
V30	ExpectS8	Intervalar	Expectativa salarial - Acima de 30 salários mínimos (mais de R$ 21.720,01).
V31	ExpectS9	Intervalar	Expectativa salarial - Não sabe responder

Quadro 2 - Variáveis analisadas **Fonte:** elaboração dos autores.

Vale destacar, que o tratamento dos dados foi realizado com tipo de implicação segundo a "teoria clássica" e a "lei Binominal". A apresentação dos resultados será organizada em dois momentos; no primeiro, as análises realizadas com base no gráfico de similaridade; e, no segundo, fundamentadas no gráfico implicativo.

Antes de expor sobre os resultados alcançados, apresenta-se o perfil dos sujeitos que compuseram a amostra deste estudo; no que refere-se a modalidade de ingresso na universidade, bem como o tipo de curso que o estudante está inserido.

Conforme já destacado neste texto, a amostra foi reduzida a 121 estudantes, sendo 61% brancos ingressantes pelo sistema tradicional de vestibular e 39% de pardos ou negros do sistema de cotas. Tal resultado já era previsto, uma vez que 55% das vagas na UFBA são reservadas para aqueles que desejam entrar pelo sistema universal.

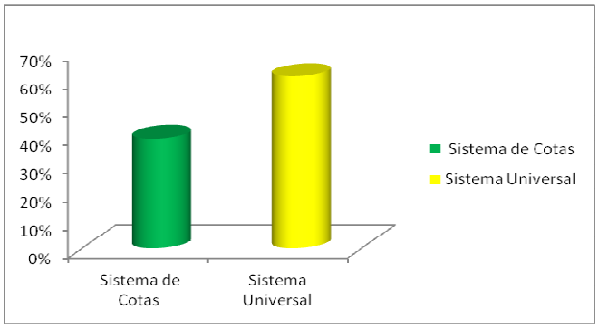


Gráfico 1 - Modalidade de ingresso na UFBA **Fonte:** pesquisa de campo

No que se refere ao nível de prestígio dos cursos, observou-se que 46% dos participantes são dos cursos de alto prestígio social e 54% de baixo prestígio social. Em relação ao primeiro, a participação dos pardos ou negros cotistas foi de 32%; já os estudantes brancos do sistema universal, o valor foi de 68%. Em relação aos cursos de baixo prestígio social, nota-se que a participação dos negros cotistas é de 45%, enquanto que os brancos do sistema universal a porcentagem é de 55%.

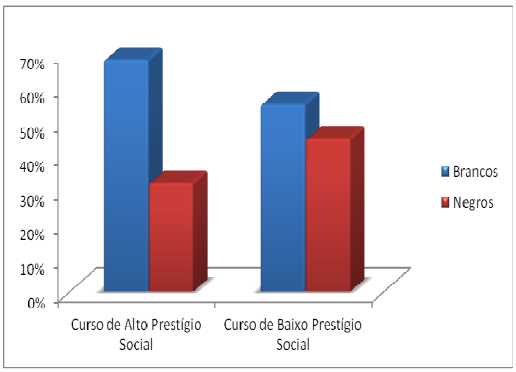


Gráfico 2 - Modalidade de ingresso na UFBA **Fonte:** pesquisa de campo

Nesse contexto, nota-se que os estudantes brancos têm maior porcentual nos dois tipos de cursos. Outra análise importante é a diferença de porcentagem entre brancos não cotistas e negros cotistas dos cursos de alto prestígio social, haja vista que os últimos apresentaram um valor inferior a 35%. Tal diferença pode ser justificada, pois historicamente a população negra teve maior participação em ocupações desvalorizadas ou mal remuneradas, conforme Fonseca (2009).

4.1 Análises do gráfico de similaridade

Conforme pode ser observado a seguir, a árvore de similaridade é formada por 7 (sete) classes com similaridade entre 22 variáveis. A primeira classe apresenta 3 (três) variáveis; a segunda 5 (cinco); a terceira 3 (três); a quarta 4 (quatro); a quinta 2 (duas); a sexta classe é composta por 2 (duas); e a sétima 3 (três). A primeira classe, apresenta um nó significativo[15], sendo formada por variáveis que indicam que os indivíduos que já trabalham na sua área de formação (ExpectICP1), têm, sobretudo, expectativa salarial de 3 a 4,5 salários mínimos para os primeiros seis meses após conclusão do curso universitário (ExpectS4) ou revelam não apresentar expectativa salarial (ExpectS9).

[15] Similaridade bastante significativa entre variáveis, conforme é mostrado em destaque vermelho na árvore.

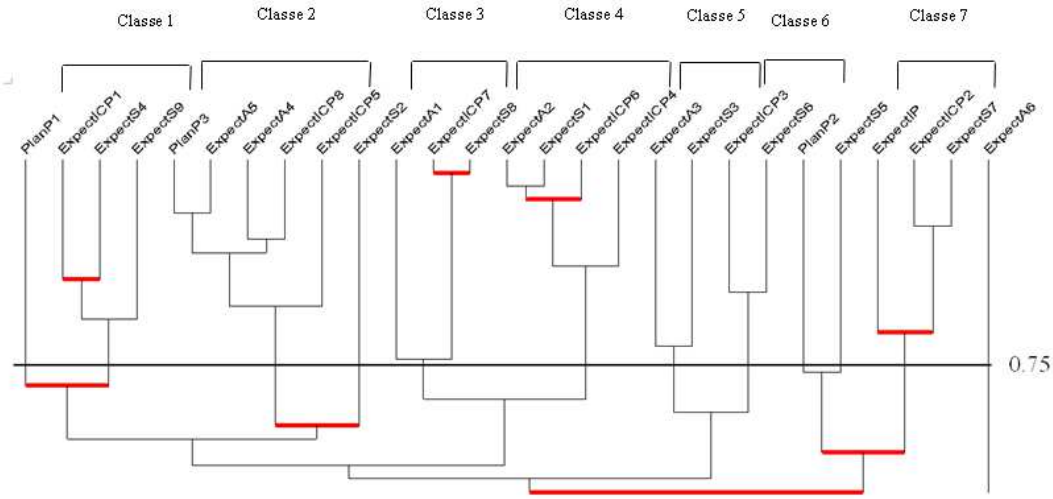


Figura 1: Árvore de similaridades

Como o maior nível de similaridade encontra-se entre as duas primeiras variáveis, dar-se-á maior atenção a essa subclasse. Vale destacar que a variável suplementar CursBPS (estudantes dos cursos de menor prestígio social) foi a que mais contribuiu para a formação dessa classe, com risco de 0.0185. A segunda, foi a variável IngresST (estudantes brancos que ingressaram pelo Sistema Tradicional), com risco de 0.354.

Diante disso, pode-se afirmar que o prestígio social do curso interfere na expectativa salarial, uma vez que, na classe 1, os estudantes dos cursos de BPS não esperam um bom retorno salarial. Além disso, destaca-se ainda a importância da experiência profissional. Trata-se de estudantes já inseridos no mercado de trabalho e, por isso, já podem identificar os aspectos positivos e negativos que envolvem o campo de trabalho da formação realizada.

Contribuição à classe 1: ExpectICP1,ExpectS4,ExpectS9
A variável IngresST contribui a esta classe com um risco de: 0.354
A variável IngresSC contribui a esta classe com um risco de: 0.551
A variável CursAPS contribui a esta classe com um risco de: 0.988
A variável CursBPS contribui a esta classe com um risco de: 0.0185
A variável que contribui mais a esta classe é CursBPS com um risco de : 0.0185

Quadro 3 - contribuições das variáveis suplementares para a formação da classe 1 de similaridade

A segunda classe e composta por estudantes que ainda não elaboraram seu projeto profissional (PlanP3), ou seja, ainda não sabem quais são os objetivos a serem alcançados após a conclusão do curso universitário (ExpectA5); estão em dúvida se vão apenas continuar os estudos, se preferem exclusivamente trabalhar ou ainda fazer as duas coisas (ExpectA4); estudantes que pensam em tirar um período de férias e que não têm previsão de tempo para inserção profissional (ExpectICP8) ou apresentam uma estimativa de no mínimo 1 (um) ano e 6 (seis) meses para início do exercício profissional (ExpectICP5); e, revelam uma expectativa salarial bastante baixa, até 1,5 salários mínimos (R$1.086,00). A variável que mais contribuiu para a formação dessa

classe também é Curso de Baixo Prestígio Social BPS, com risco de 0.153; a segunda variável com maior cooperação foi IngresST (ingressantes pelo sistema tradicional), com risco de 0.413.

Nessa direção, a falta de um plano profissional e a baixa expectativa salarial também pode ser justificada pelo prestígio social do curso dos estudantes. Tendo em vista a visão negativa que se constrói em relação a certas áreas profissionais, determinados estudantes se sentem desestimulados em relação a sua inserção no mercado de trabalho.

Contribuição à classe2: PlanP3,ExpectA5,ExpectA4,ExpectICP8,ExpectICP5,ExpectS2

A variável IngresST contribui a esta classe com um risco de : 0.413
A variável IngresSC contribui a esta classe com um risco de : 0.467
A variável CursAPS contribui a esta classe com um risco de : 0.865
A variável CursBPS contribui a esta classe com um risco de : 0.153

A variável que contribui mais a esta classe é CursBPS com um risco de : 0.153

Quadro 4 – contribuições das variáveis suplementares para a formação da classe 2 de similaridade

A classe três, também com um nó significativo, mostra que os estudantes que pretendem apenas trabalhar durantes os primeiros 6 (seis) meses após obtenção do diploma (ExpectA1), tendem a ter uma estimativa de 30 salários mínimos (ExpectS8). Porém, acreditam que vão demorar em torno de 24 meses para iniciar a carreira profissional (ExpectICP7). A variável que contribuiu para a formação da classe é CursAPS (curso de alto prestígio social), com risco de 0.13; em seguida, a variável IngresST, com risco de 0,239.

Apesar dos estudantes da classe 3 não esperarem um ingresso no mercado de trabalho imediato, é interessante observar a preocupação em apenas trabalhar após a conclusão do curso. Por se tratar de estudantes oriundos de cursos de alto prestigio social, estes possivelmente acreditam que não haverá necessidade de um investimento imediato em formação continuada para a inserção no mercado de trabalho.

Contribuição à classe 3 : ExpectA1,ExpectICP7,ExpectS8

A variável IngresST contribui a esta classe com um risco de : 0.239
A variável IngresSC contribui a esta classe com um risco de : 0.543
A variável CursAPS contribui a esta classe com um risco de : 0.13
A variável CursBPS contribui a esta classe com um risco de : 0.852

A variável que contribui mais a esta classe é CursAPS com um risco de : 0.13

Quadro 5 – contribuições das variáveis suplementares para a formação da classe 3 de similaridade

Em relação à classe 4, também com nó significativo, o gráfico aponta que os estudantes que não apresentam nenhuma expectativa salarial (ExpectS1), acreditam que podem demorar de 7 a 12 meses (ExpectICP4) ou de 19 a 24 meses (ExpectICP6) para alcançar a inserção profissional e pretendem apenas investir na formação continuada, em um primeiro momento (ExpectA2).

O interesse em apenas continuar os estudos por parte dos estudantes pode explicar a estimativa de acesso ao mercado de trabalho de forma mais tardia. Assim, a

preocupação é, primeiramente, investir em formação e, posteriormente, tratar do exercício profissional.

A variável típica da classe 3 é CursAPS (curso de alto prestígio social), com risco de 0.13; e ainda IngresST (ingressantes pelo sistema tradicional), com 0.239. Trata-se de um resultado diferente do que foi observado na formação da classe 3, pois apesar do alto prestígio do curso, os estudantes da classe 4 tem interesse em dar continuidade aos estudos. Assim, entre os estudantes dos cursos de alto prestígio social, podem-se identificar dois grupos: indivíduos que pensa em apenas trabalhar; indivíduos que pretendem, em primeiro momento, investir na formação continuada.

Contribuição à classe 4 : ExpectA2,ExpectS1,ExpectICP6,ExpectICP4

A variável IngresST contribui a esta classe com um risco de : 0.239
A variável IngresSC contribui a esta classe com um risco de : 0.543
A variável CursAPS contribui a esta classe com um risco de : 0.13
A variável CursBPS contribui a esta classe com um risco de : 0.852

A variável que contribui mais a esta classe é CursAPS com um risco de : 0.13

Quadro 6 – contribuições das variáveis suplementares para a formação da classe 4 de similaridade

Na classe 5, observa-se que os estudantes que pretendem tanto estudar quanto trabalhar (ExpectaA3) têm uma estima salarial de 1,5 a 3 salários mínimos (ExpectS3). A variável que mais contribuiu para a construção da classe é CursBPS (curso de baixo prestigio social), com risco de 0.0072; e ainda a váriavel IngresSC (estudantes negros ingressantes pelo sistema de cotas), com risco de 0.0411.

Contribuição à classe 5: ExpectA3,ExpectS3

A variável IngresST contribui a esta classe com um risco de : 0.901
A variável IngresSC contribui a esta classe com um risco de : 0.0411
A variável CursAPS contribui a esta classe com um risco de : 0.996
A variável CursBPS contribui a esta classe com um risco de : 0.0072

A variável que contribui mais a esta classe é CursBPS com um risco de : 0.0072

Quadro 7 – contribuições das variáveis suplementares para a formação da classe 5 de similaridade

A expectativa salarial relativamente baixa justifica-se tendo em vista que o estudante que opta em estudar e trabalhar, não poderá disponibilizar uma parte significativa do seu tempo para o exercício profissional; o que consequentemente impedirá o alcance de maior retorno salarial.

Dado interessante é que enquanto entre os estudantes dos cursos de APS observa-se a formação de dois grupos (aqueles que só pretendem estudar e aqueles que planejam apenas trabalhar), entre os estudantes dos cursos BPS formou-se apenas um grupo: indivíduos que pretendem investir tanto na inserção profissional imediata, mas também na formação continuada. Tendo em vista que, historicamente, no Brasil os cursos menos valorizados, como as licenciaturas, são muitas vezes frequentados pela população de menor poder aquisitivo ou ainda por jovens cujas famílias não podem apoiá-los

financeiramente; é possível compreender a preocupação destes estudantes em conseguir trabalho, enquanto meio sobrevivência.

É importante notar ainda que ao contrário do que ocorreu na formação das classes anteriores, na classe 5 observa-se que são os estudantes cotistas negros que contribuíram para a formação da classe. Todavia, tal contribuição é secundária, uma vez que a variável prestígio do curso é que mais tem colaborado na formação das classes vistas.

Analisando a classe 6, percebe-se que os estudantes que acreditam que vão precisarem de até seis meses para entrarem no mercado de trabalho (ExpectICP3), apresentam uma boa expectativa de remuneração, em torno de 6 a 10 salários (ExpectS6). Contribuiu na formação da classe a variável cursAPS (curso de alto prestigio social), risco de 0.0856; e depois a variável IngresST (brancos ingressantes pelo Sistema Tradicional), com 0.43. Novamente, destaca-se que o alto prestígio do curso pode influenciar na alta expectativa salarial.

Contribuição à classe 6: ExpectICP3,ExpectS6 A variável IngresST contribui a esta classe com um risco de : 0.43 A variável IngresSC contribui a esta classe com um risco de : 0.414 A variável CursAPS contribui a esta classe com um risco de : 0.0856 A variável CursBPS contribui a esta classe com um risco de : 0.898 A variável que contribui mais a esta classe é CursAPS com um risco de : 0.0856

Quadro 8 – contribuições das variáveis suplementares para a formação da classe 6 de similaridade

E, por fim, a classe 7 (sete) que é caracterizada por estudantes com expectativa muito alta de inserção no mercado de trabalho em Salvador (ExpectIP); que acreditam que terão uma inserção profissional imediata (ExpectICP2); assim como um bom rendimento salarial de 6 a 10 salários mínimos (ExpectS6). A variável típica da classe é CursAPS (curso de alto prestigio social), com risco de 0.00777; em seguida, IngresST (estudantes brancos ingressantes pelo sistema tradicional).

Contribuição à classe 7: ExpectIP,ExpectICP2,ExpectS7 A variável IngresST contribui a esta classe com um risco de : 0.413 A variável IngresSC contribui a esta classe com um risco de : 0.467 A variável CursAPS contribui a esta classe com um risco de : 0.00777 A variável CursBPS contribui a esta classe com um risco de : 0.989 A variável que contribui mais a esta classe é CursAPS com um risco de : 0.00777

Quadro 9 – contribuições das variáveis suplementares para a formação da classe 7 de similaridade

Tendo em vista os resultados apresentados com base no gráfico de similaridade, é possível perceber que a variável que mais contribuí para a formação das classes é o prestigio/valorização do curso em que o estudante está inserido. Como analisado, os estudantes dos cursos de APS normalmente já possuem um plano de inserção profissional, têm boas expectativas salariais e acreditam que, se quiserem, vão entrar no mercado de trabalho em curto período de tempo. Situação contrária, é normalmente observada com os estudantes dos cursos BPS.

A outra variável suplementar modalidade de ingresso na universidade não apresentou influência considerável sobre as classes. Diante disso, a partir dos dados trabalhados não é possível afirmar que as diferenças de expectativas entre os estudantes podem sem explicadas com base na modalidade de ingresso na universidade (cotas ou sistema universal), mas, sobretudo no prestígio do curso.

4.2 Análises do gráfico implicativo

Por meio do software CHIC, foi obtido o gráfico a seguir com o valor mínimo de 0.80; nível de confiança considerado adequado para obtenção de uma avaliação estatística satisfatória.

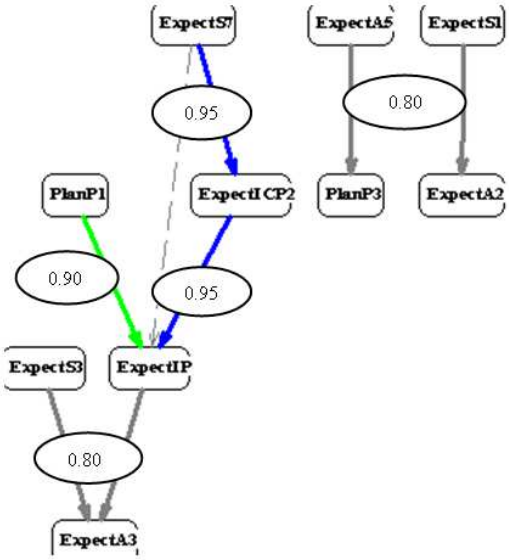


Figura 2: Gráfico Implicativo

Conforme pode ser observado, foram obtidas duas implicações simples:

ExpectS1 ⇨ ExpectA2: com nível de confiança em 0,80, tal caminho mostra que a ausência de expectativa salarial está condicionada a preocupação do estudantes em investir na continuidade dos estudos depois da obtenção do diploma. Está implicação pode ser compreendida, uma vez que se o estudante pretende apenas estudar por um determinado período, este, possivelmente, não terá nenhuma remuneração.

ExpectA5 ⇨ PlanP3: também com nível de confiança em 0,80, observa-se que a ausência de um plano profissional com objetivos bem definidos e estratégias específicas para alcançá-las implica na dúvida do que fazer após a conclusão do curso de ensino superior. Tal relação é também facilmente compreendida, havia vista que a construção de um plano profissional envolve a definição e reflexão a respeito dos objetivos a serem alcançados a médio e longo prazo.

No gráfico é possível ainda visualizar um caminho mais longo de quase implicação entre as variáveis **ExpectS7 ⇨ ExpectICP2 ⇨ ExpectIP ⇨ ExpectA3**. Esse caminho, com nível de confiança 0,90, significa que alta expectativa salarial (de 10 a 30 salários) dos estudantes é consequência de uma expectativa bastante positiva quanto ao tempo para o início do exercício profissional (inserção imediata nos primeiros seis

meses); e, por fim, no planejamento profissional que contemple formação continuada. Dessa forma, é possível inferir que, para os estudantes, uma boa expectativa de inserção profissional (com alto salário; entrada rápida no mercado de trabalho) relaciona-se com a construção de um plano profissional que inclua formação continuada para os primeiros seis meses após a conclusão do curso de graduação.

Outro aspecto interessante a ser observado é que a alta expectativa de inserção profissional em Salvador (ExpectIP) está relacionada com a elaboração de um plano profissional (PlanP1), ou seja, com a identificação das oportunidades de carreiras e ações para alcança-las. Merece ainda destaque a implicação entre ExpectS3 e ExpectA3, uma vez que ela fortalece a ideia de que quando o estudante espera por um bom retorno salarial, esse também demonstra preocupação em investir nos estudos.

Vale destacar que a variável que mais contribuiu para a formação do caminho analisado é CursAPS, com risco de 0.0541; em segundo a variável IngresST (estudantes brancos que ingressaram pelo sistema tradicional de vestibular), com risco de 0.359.

Assim, tal resultado reafirma as análises feitas anteriormente, com base na árvore de similaridade, uma vez que foi observado que a alta expectativa de inserção profissional está normalmente condicionada, sobretudo, ao prestígio ou a valorização do curso.

5 Considerações finais

A implementação de Ações Afirmativas de ingresso no Ensino Superior, de maneira específica para a população negra, veicula um escopo explícito de transformação. Ao efetivá-la, o Estado, as universidades, assim como os responsáveis pela elaboração dessa política passam a trabalhar em torno da busca pela correção dos efeitos da discriminação cometida no passado, tendo como intuito maior a consolidação de um projeto de igualdade e de edificação de uma sociedade mais democrática (Munanga e Gomes, 2006, p.186).

Tendo em vista o papel importante das Ações Afirmativas e compreendendo que essa não deve apenas ampliar o número de estudantes negros nas universidades, mas principalmente modificar o sistema de estratificação social brasileiro; a avaliação dos seus resultados se coloca como necessária. Investigar se existem diferenças entre as expectativas e alcances profissionais de inserção no mercado de trabalho, entre estudantes negros cotistas e brancos não cotistas, permitem reflexões sobre as contribuições da referida política e, consequentemente, auxilia no debate em torno da reformulação ou elaboração de novas políticas de apoio à população negra.

Apesar de o presente artigo ter abordado apenas as expectativas profissionais, não explorando aspectos relacionados a alcances de inserção profissional, foi possível realizar análises importantes. Conforme revelado nos gráficos de similaridade e de implicação, as diferenças nas expectativas profissionais não estão relacionadas a modalidade de ingresso na universidade (pelas cotas ou pelo sistema tradicional de vestibular), mas pelo tipo de curso que o estudante está inserido (alto ou baixo prestígio).

Normalmente, os estudantes dos cursos de alto prestígio social (Medicina, Direito, Engenharias), apresentam expectativa mais altas de inserção profissional em Salvador;

têm boas expectativas salariais e acreditam que poderão entrar no mercado de trabalho logo nos primeiros seis meses após obtenção do diploma. O mesmo resultado não foi observado com estudantes de cursos de baixo prestígio social. É interessante observar ainda que, enquanto os estudantes dos cursos mais valorizados pretendem investir exclusivamente na continuidade dos estudos após a conclusão da graduação ou apenas no exercício profissional, tendo em vista as grandes possibilidades de inserção imediata; os estudantes dos cursos de baixo prestígio social acreditam que vão precisar intercalar trabalho e investimento em formação continuada.

Dessa forma, tendo em vista a discussão que fundamentou este artigo "as desigualdades raciais", os resultados são positivos, uma vez que não obtivemos dados que aponta iniquidades entre negros e brancos, ou seja, a questão do pertencimento racial não explica as diferenças de expectativas entre estudantes. Todavia, para análise mais aprofundada, faz-se necessário verificar se houve diferenças no processo de inserção profissional; tratamento que se encontra em desenvolvimento.

Referências

[1] Anunciação, M. P. P (2005). *Análise da demanda social, segundo a cor e procedência*. UFBA. 2001 – 2005. Disponível em: <http://www.proplan.ufba.br/demanda_social01-05.pdf>. Acesso em: 12 set 2011.

[2] Borges, M.A.F (2009). *Apropriação das tecnologias de informação e comunicação pelos gestores educacionais* **(**Tese de Doutorado – Faculdade de Educação da Universidade Católica de São Paulo – PUC/SP). São Paulo.

[3] Cardoso, C. B (2008). Efeitos da política de cotas na Universidade de Brasília: uma analise do rendimento e da evasão. (Dissertação de mestrado, Universidade Federal de Brasília. Brasília - DF. Disponível em: *<http://repositorio.bce.unb.br/handle/10482/1891>*. Acesso em 30 de setembro de 2011.

[4] Castro, M. G.; Abramovay, M (2006). Relações raciais na escola: reproduções de desigualdades em nome da Igualdade. Brasília: UNESCO, INEP.

[5] Diesse – Departamento Intersindical de Estatística e Estudos Socioeconômicos. *Os negros no trabalho (2013).* Disponível em: <http://www.dieese.org.br/>. Acesso em 14 de janeiro de 2014.

[6] Fernandes, F (1978). A integração dos negros na sociedade de classes. São Paulo: Dominus.

[7] Fonseca, N. S. F (2006). Visibilidade e ocultação da diferença: imagens de negro na cultura brasileira. In: FONSECA, N. S. F.(org.) *Brasil afro-brasileiro***.** 2 ed. Belo Horizonte: Autêntica, p. 87-115.

[8] Gonineau, A (2006). Essai sur l´inégalité des races humaines. Paris, Gallimard-Pleiade (1ª ed. 1853). Apud. Fonseca, N. S. F. Visibilidade e ocultação da diferença: imagens de negro na cultura brasileira. In: Fonseca, N. S. F.(org.) *Brasil afro-brasileiro.* 2 ed. Belo Horizonte: Autêntica, p. 87-115.

[9] Guimarães, A. S. A (2002). *Como trabalhar com 'raça' em sociologia*. Educação e Pesquisa, v. 29, n. 1, p. 93-107, jan./jun. p. 93 – 107.

[10] Gras, R., Régnier, J.-C., Marinica, C., Guillet, F. (2013), *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire à la recherche de causalités.*, Éditions Cépaduès, Toulouse

[11] Hasenbalg, C. A (2005). *Discriminação e desigualdades raciais no Brasil*. 2.ed. Belo Horizonte: Editora UFMG. Acesso em: 13 de janeiro de 2014.
[12] IBGE – Instituto Brasileiro de Geografia e Estatística. *Síntese de Indicadores Sociais: uma análise das condições de vida da população brasileira*. 2012 Disponível em:
[13] ftp://ftp.ibge.gov.br/Indicadores_Sociais/Sintese_de_Indicadores_Sociais_2012/SIS_2012.pdf>
[14] Jaccoud, L (2008). O combate ao racismo e a desigualdade: o desafio das políticas públicas de promoção da igualdade racial. In: Theodoro, M. (org.). *As políticas públicas e a desigualdade racial no Brasil: 20 anos após a abolição*. Brasília: IPEA, p.131-166.
[15] Munanga, K.; Gomes, N. L (2006). *O negro no Brasil de hoje*. São Paulo: Global.
[16] Osorio, G. R (2009). Desigualdade racial e mobilidade social no Brasil: um balanço das teorias. In: Osório, R. G. *A desigualdade de renda no Brasil: 1976 – 2006*. (Tese de doutorado – Universidade Federal da Bahia, Programa de Pós-Graduação em Sociologia). Brasília-DF.
[17] Richardson, R. J. et.al (1989). *Pesquisa Social: métodos e técnicas*. São Paulo: Atlas.
[18] Rousseau, J. J (2006). *Discurso sobre a origem e os fundamentos da desigualdade entre os homens*. Trad. Lourdes Santos Machado. São Paulo: Abril Cultural, 1973. *Apud* SOUZA, F.;. As ações afirmativas como instrumento de concretização da igualdade material (dissertação de mestrado). Universidade. Universidade Federal do Paraná.
[19] Santos, D. B. R (2009). *Para além das cotas: a permanência de estudantes negros no ensino superior como política de ação afirmativa*. (Tese de Doutorado – Universidade Federal da Bahia, Programa de Pós-Graduação em Educação.
[20] Sell, S. C (2002). Ação afirmativa e democracia racial: uma introdução ao debate no Brasil. Florianópolis: Fundação Boiteux,
[21] Silva, A. et.al (2009a). A promoção da igualdade racial no primeiro semestre de 2007 e os programas de ação afirmativa nas universidades públicas. In: Jaccoud, L. (org.) *A construção de uma política de promoção da igualdade racial: uma analise dos últimos 20 anos*. Brasília: IPEA, p. 171-204.
[22] Silva, A. et.al (2009b). Entre o racismo e a desigualdade: da constituição a promoção de uma política de igualdade racial (1988-2008). In: Jaccoud, L. (org.) *A construção de uma política de promoção da igualdade racial: uma analise dos últimos 20 anos*. Brasília: IPEA, p. 147-170.
[23] Souza, F.(2006) *As ações afirmativas como instrumento de concretização da igualdade material* (Dissertação de mestrado – Universidade Federal do Paraná, Programa de Pós-Graduação em Direito) Curitiba – PR.
[24] Theodoro, M. (org.) (2008) As políticas públicas e a desigualdade racial no Brasil: 120 anos após abolição. Brasília: IPEA, p.69-99.
[25] Theodoro, M. (2008). À guisa de conclusão: o difícil debate da questão Racial e das políticas públicas de combate à desigualdade e à Discriminação racial no Brasil. In: THEODORO, M. (org.) *As políticas públicas e a desigualdade racial no Brasil: 120 anos após abolição*. Brasília: IPEA, p. 167-176.
[26] UNIVERSIDADE FEDERAL DA BAHIA (2005). Análise da demanda social segundo cor e procedência. Pró-Reitoria de Planejamento e Administração – setor de informação e documentação. Salvador: UFBA.

[27] UNIVERSIDADE FEDERAL DA BAHIA (2004). Resolução 01/04. Salvador: UFBA.

[28] UNIVERSIDADE FEDERAL DA BAHIA *(2005)***.** Sistema de cotas no vestibular 2005: análise dos resultados. Disponível em: <http://www.vestibular.ufba.br/docs/vest2005cotas.pdf>. Acesso em: 15 de dez, 2013.

[29] UNIVERSIDADE FEDERAL DA BAHIA *(2016)***:** Plano de Desenvolvimento Institucional 2012 – 2016. Disponível em: <https://www.ufba.br/sites/devportal.ufba.br/files/pdiufba_2012-16.pdf>. Acesso em 15 de dez, 2013.

[30] Varela, S. F (2009). *Discriminação racial indireta e ação afirmativa no emprego sob a perspectiva dos direitos coletivo* (Tese de doutorado – Departamento de Sociologia, Universidade de Brasília), Brasília-DF.

[31] 7º *Colóquio Internacional sobre Análise Estatística Implicativa* (2013). Disponível em: <http://sites.univ-lyon2.fr/asi7/lang/br/Appel_ASI7_BR.pdf_>. Acesso em: 12 jan. 2014.

A QUESTÃO DA PERMANÊNCIA E DESISTÊNCIA DOS ESTUDANTES DE LICENCIATURA EM CIÊNCIAS E MATEMÁTICA NO BRASIL. ESTUDO[1] EXPLORATORIO NO CFP/UFCG ABORDADO PELO QUADRO DA ANÁLISE ESTATÍSTICA IMPLICATIVA

Valéria Maria de LIMA BORBA[2], Anna Paula de AVELAR BRITO LIMA[3], Jean-Claude RÉGNIER[4]

LA QUESTION DU MAINTIEN ET DE L'ABANDON DES ÉTUDIANTS DANS LE PARCOURS DE LICENCE SCIENCES ET MATHÉMATIQUES AU BRESIL. ÉTUDE EXPLORATOIRE AU SEIN DU CFP/UFCG DANS LE CADRE DE L'ANALYSE STATISTIQUE IMPLICATIVE

THE QUESTION OF RETENTION AND DROPOUT OF STUDENTS IN UNDERGRADUATE TEACHER-TRAINING COURSES IN SCIENCE AND MATHEMATICS OF THE BRAZIL. A STUDY IN THE FRAMEWORK OF STATISTICAL IMPLICATIVE ANALYSIS.

RESUMO

Este estudo se propõe a identificar e compreender os comportamentos com relação às permanências e desistências dos estudantes ingressos nos cursos de Licenciatura em Ciências com Habilitação em Matemática e de Licenciatura em Matemática, ambos do CFP/UFCG, campus Cajazeiras. Participaram desse estudo 22 estudantes dos curso supra citados. O instrumento de coleta de dados foi um questionário. O objetivo do questionário foi obter subsídios referentes ao desempenho destes estudantes: auto avaliação dos estudantes acerca de seus desempenhos antes e durante os primeiros períodos dos cursos de Licenciatura. Os dados foram analisados a partir da abordagem da Análise Estatística Implicativa – A.S.I. utilizando-se o software CHIC na versão 6.0. As analises feitas dos dados modelados com o CHIC possibilitaram uma compreensão mais aprofundada do fenômenos em estudo, indicando algumas relações importantes como a influência da universidade na percepção dos estudantes do fracasso no curso.

Palavras-chave : *desistências e permanência, comportamentos, ensino e aprendizagem, cálculo I, estudantes de licenciatura.*

RÉSUMÉ

Cette étude vise à identifier et de comprendre les comportements de maintien ou d'abandon des étudiants dans le parcours de formation de licence science et mathématique et de

[1] Esta pesquisa contou com o apoio da CAPES através da bolsa PVE para o programa PPGEC-UFRPE.

[2] Doutoranda do Programa de Pós-Graduação em Ensino das Ciências e Matemática da Universidade Federal Rural de Pernambuco – UFRPE; e-mail : valbo66@yahoo.com.br

[3] Programa de Pós-Graduação em Ensino das Ciências e Matemática da Universidade Federal Rural de Pernambuco (UFRPE), Recife, Brasil, e-mail: apbrito@gmail.com

[4] Pesquisador Visitante Especial PVE/CAPES no PPGEC/UFRPE (Programa de Pós-Graduação em Ensino das Ciências da Universidade Federal Rural de Pernambuco) – UMR5191 – ICAR Université Lumiere Lyon 2 (FRA) jean-claude.regnier@univ-lyon2.fr

licence mathématique, les deux de CFP/UFCG, campus Cajazeiras. Ici, nous avons centré notre étude sur un échantillon de 22 étudiants qui ont participé à cette étude. L'instrument de construction des données est un questionnaire. Le but du questionnaire était d'obtenir des données relatives à la performance de ces étudiants : auto-évaluation des étudiants au sujet de leurs performances avant et pendant les premières années de cours de licence. Les données ont été analysées à partir de l'approche de l'Analyse Statistique Implicative avec l'aide du logiciel CHIC (Classification Hiérarchique Implicative et Cohésitive). L'analyse faite des données modélisées avec CHIC a permis une meilleure compréhension des phénomènes étudiés. Indiquant certaines relations importantes telle que l'influence de l'université dans la perception de l'échec au cours des élèves.

Mots-clés : *maintien, abandon, comportement, enseignement et apprentissage, calcul, étudiants de licence.*

ABSTRACT

This study sets out to identify and understand behaviours with respect to the drop-out and continuation of students attending Licentiate courses in Sciences with a Qualification in Mathematics and in Mathematics, both at CFP/UFCG, Cajazeiras campus. 22 students of the courses mentioned above took part in this study. The data collection instrument was a questionnaire. The purpose of the questionnaire was to obtain underlying reasons for the performance of these students: the students' self-evaluation of their performances before and during the first periods of the Licentiate courses. Data were analysed based on Statistical Implicative Analysis (SIA) approach using CHIC (Cohesive Hierarchical Implicative Classification) software. The analysis made of the data modelled with CHIC enabled a deeper understanding of the phenomena under study to be obtained. Some important relationships were revealed such as the influence of the university on the students' perception of failure on the course.

Keywords: *dropouts and continuers, behaviors, teaching and learning, calculus I, licentiate students.*

1 Introdução

O presente estudo se propõe a analisar as razões das permanências e das desistências de estudantes ingressos nos cursos de Licenciatura em Ciências com Habilitação em Matemática e de Licenciatura em Matemática ambos do Centro de Formação de Professores (CFP) da Universidade Federal de Campina Grande (UFCG), campus Cajazeiras, entre 2011 e 2013 (período da pesquisa), observando os fatores que implícita e explicitamente contribuem para uma possível ruptura existente no desenvolvimento da aprendizagem matemática dos estudantes. Tal ruptura se revela no comportamento matemático dos ingressos no curso supracitado, causando uma queda no desempenho dos estudantes que iniciam nesses cursos. Para isto caminharemos através de vários temas.

É importante salientar que os problemas relativos à evasão e a baixa terminalidade dos cursos de Licenciatura em Matemática não é um fenômeno restrito à realidade local do CFP da UFCG, uma vez que várias universidades públicas e privadas tanto no Brasil quanto no exterior (pelo menos no mundo ocidental), já perceberam o problema e buscam formas de enfrentá-lo. Entretanto, o que se percebe é que alguns programas que tratam do problema, quando conseguem algum progresso, os mesmos não são facilmente transferíveis para outras instituições.

Para melhor entender as razões da permanência e da desistência e os fatores que levam os dos estudantes a tais atos, nos cursos de Licenciatura em Ciências com Habilitação em Matemática e de Licenciatura em Matemática, propõe-se aqui retomar aspectos importantes da própria epistemologia da matemática que poderão ajudar a compreender tal fenômeno.

2 A Matemática como ferramenta sócio histórica 2

A importância da matemática no processo de humanização e da construção da cidadania é inquestionável, uma vez que se utiliza a matemática desde as mais corriqueiras às mais complexas atividades, em contextos escolar e extraescolar. Portanto, torna-se imprescindível pensar de que forma esse saber está sendo tratado na escola e nas universidades, espaços privilegiados para a inserção dos estudantes no saber matemático. Mais que uma disciplina escolar, a matemática, assim como todo o conhecimento construído no decorrer da história, é herança de toda a humanidade, logo a existência de pessoas impedidas de ter acesso a tal ferramenta representa grave injustiça sociocultural. Entretanto, é mais comum encontrarmos pessoas com sintomas de "Matofobia[5]" que pessoas que transitam pelo mundo da matemática sem problemas. Assim, é importante mapearmos a origem dessa situação que vem, historicamente, impedindo o sucesso dos estudantes em tal disciplina.

2.1 Desarticulação entre a matemática escolar e a matemática do cotidiano extraescolar

A escola brasileira tem como parâmetro avaliativo de inteligência um enfoque apoiado em uma visão de homem do início do século XX, em função da qual se costuma avaliar a inteligência a partir de duas amplas categorias de competências: a verbal, revelada pela capacidade de expressão e uso conveniente das palavras, e a matemática, definida pela capacidade de solucionar problemas, desenvolvidos por meio de uma percepção espacial de objetos concretos. Contudo, atualmente podemos nos calcar em olhares outros que salientam outros aspectos referentes ao desenvolvimento humano.

Cabe então salientar que historicamente a matemática tem sido considerada como uma disciplina formal, universal e descontextualizada, contendo "verdades eternas", obtidas pelo poder da lógica e que lida com um conjunto articulado de conceitos abstratos, caracterizando-se como uma área de conhecimento institucionalizado. Por outro lado, tal disciplina tem sido responsável em muitos países por uma significativa parcela de fracasso escolar. (Huillet e Mutemba, 2000; Taylor, 1999; Robert, 1998)

A presença e importância da matemática na escola é consequência de sua presença na sociedade e, portanto, as necessidades matemáticas que surgem na escola deveriam estar subordinadas às necessidades matemáticas da vida em sociedade como um todo, tornando os usuários dessa área de conhecimento habilitados para tratar com a mesma em todos os momentos em que tais conhecimentos fossem solicitados.

[5] Medo da matemática, cf. Nimier, 1988.

Torna-se imperativo esclarecer que não é objetivo desse estudo fazer apologia à matemática aplicada imediatamente ao "mundo extraescolar", uma vez que se compreende a importância da utilização desse conhecimento em contexto abstrato. A escola tem uma responsabilidade social e não deveria permitir que seus estudantes saíssem despreparados para atuar numa sociedade que exige cada vez mais dos seus integrantes o domínio de saberes construído historicamente pela humanidade.

Isto significa que, dentre outras competências, a matemática precisa habilitar os estudantes a resolver problemas cotidianos, formulando-os e agindo matematicamente. Isso implica em capacita-los a lidar com uma complexidade cada vez maior à medida que progridem de um nível para outro dentro do sistema educacional, em qualquer área de formação. Para isso, no caso da matemática, se faz necessário que os estudantes compreendam as estruturas, as ideias e os métodos matemáticos e não simplesmente apliquem fórmulas ou algoritmos.

Os Parâmetros Curriculares Nacionais (BRASIL, 1998, pg. 09) referentes à Matemática do Ensino Médio comenta que: "(...) No Ensino Médio, quando nas ciências torna-se essencial uma construção abstrata mais elaborada, os instrumentos matemáticos são especialmente importantes (itálicos nossos)." Esta afirmação afiança a importância e a necessidade da aprendizagem matemática, não só como 'ferramenta' para resolução de problemas cotidianos, mas também, como meio pelo qual o pensamento abstrato se estrutura dando condições aos estudantes de projetarem-se para situações em que não haja ancoragem nos objetos concretos.

Assim, continua o mesmo documento:

> (...) A Matemática ciência, com seus processos de construção e validação de conceitos e argumentações e os procedimentos de generalizar, relacionar e concluir que lhe são característicos, permite estabelecer relações e interpretar fenômenos e informações. As formas de pensar dessa ciência possibilitam ir além da descrição da realidade e da elaboração de modelos. (BRASIL, 1998, pg. 09).

Tal documento mostra claramente que a matemática, mais que o suporte científico ao qual está intimamente vinculada, deverá também capacitar os estudantes para que eles possam analisar, argumentar, relacionar e interpretar fatos, ideias, fenômenos sociais e científicos etc. Em outras palavras, ajudar na formação de sujeitos críticos com ferramental técnico-científico capaz de conduzi-los a pensar e intervir na realidade dentro e fora da escola.

2.2 Como está a situação de abandono no ensino superior? Estudos.

É perceptível o grande número de estudantes que fracassam ao ingressarem no ensino superior, em cursos que exigem certo domínio das chamadas disciplinas científico-tecnológicas (Matemática, Física, Química e Biologia). Tal situação tem sido descrita em vários trabalhos de pesquisa que buscam explicações para o fenômeno em foco.

Várias são as justificativas daqueles que ministram e organizam os programas dos cursos de Matemática nas universidades; dentre as mais frequentes estão aquelas que colocam a culpa do fracasso nos níveis de escolarização anteriores (ensinos médio, fundamental e infantil), o que evidencia a postura de culpar o aluno pelo fracasso ocorrido. Frases dos professores responsáveis pelos ciclos gerais de formação, segundo

segundo Malta (2004, p 41) " os alunos estão ingressando nas universidades cada vez mais despreparados" e alguns professores participantes dessa pesquisa: P1:"*o aluno não chega à universidade com uma 'bagagem' suficiente para lidar com a estrutura da Matemática do nível superior*", ou P3 :"*a Matemática da universidade é muito mais abstrata que a do ensino médio, logo falta base para que haja compreensão nesse novo nível de abstração*", ou ainda P1: "*faltam os pré-requisitos necessários para que o aluno se saia bem nessa Matemática superior*".

Tal fato é apontado por Arruda & Ueno (2003) quando analisam o ingresso, desistência e permanência no curso de Física da Universidade Estadual de Londrina (estado do Paraná). Nesse texto, os autores mostram que muitos dos estudantes que entram no curso de Física da universidade supracitada e acabam desistindo por enfrentarem diversos tipos de problemas, inclusive com relação à complexidade dos problemas apresentados e a real dificuldade em dar conta da quantidade de itens de conteúdo. Os autores comentam ainda que o conteúdo apresentado caracteriza-se por um nível de complexidade inédito para os estudantes, exigindo enorme esforço e disponibilidade e exercendo influência negativa na permanência dos mesmos no curso em questão.

Outro trabalho, nesse mesmo sentido, é o de Santos (1994), que procurou investigar as possíveis causas de desistência e reprovação, na disciplina de Cálculo Diferencial e Integral I, ministrada para dezessete cursos de graduação da Universidade Federal do Ceará. O estudo foi realizado com 196 estudantes da supracitada universidade, tendo encontrado como principal resultado que, apesar dos estudantes terem se considerado com bom desempenho em matemática no ensino médio, consideraram o fraco desempenho, constatado pelo autor na disciplina de Cálculo, como resultado de uma deficiente formação básica em Matemática. Comentando tal conclusão, o autor supracitado acrescenta ainda a pobreza e o conservadorismo pedagógico na forma como as aulas seriam ministradas, o que agravaria ainda mais o quadro.

Esses estudos fazem supor a existência de certo desnível ou não-compatibilização entre o ensino da matemática no nível médio e no ensino superior. Tal desnível parece decorrer de dois tipos de fatores: deficiência, pura e simples, de conteúdo (ou seja, não se está ensinando no ensino médio aquele mínimo de conhecimentos que deveria ser ensinado), e mudança abrupta de contrato didático (o ambiente psicossocial de funcionamento da sala de aula de matemática do ensino superior seria muito diferente, em termos "contratuais", do ambiente de ensino do nível médio).

O primeiro grupo de fatores (déficit de ensino) é encampado em várias análises, como se pode depreender da leitura da Lei de Diretrizes e Bases da Educação Nacional (LDBEN/96), no artigo 35, item I, que salienta que o Ensino Médio tem como finalidade: "*a consolidação e o aprofundamento dos conhecimentos adquiridos no ensino fundamental, possibilitando o prosseguimento de estudos* (itálicos acrescentados)".

2.3 Ensinar matemática: aspectos envolvidos no processo

Dentre as várias pesquisas sobre o tópico ensinar e aprender matemática podemos citar a de Da Rocha Falcão & Meira (1994), que explica que ensinar matemática tem, historicamente, significado transferir princípios, algoritmos e conceitos gerais, cuja

articulação caberia ao aprendiz. Dito de outra forma, ensinar pressupõe a transmissão de conhecimento acumulado historicamente e preparado institucionalmente para o ensino (ou seja, "curricularizado" e normatizado para a transmissão escolar), cabendo ao professor a tarefa de gestor desse processo. Por outro lado, espera-se tradicionalmente que o estudante seja cognitivamente "capaz" de dar conta de seu papel, articulando o conhecimento que lhe é ensinado de forma adequada. Essa forma de ensinar, que tradicionalmente permeia nossas escolas, com relação à matemática, segundo Silva (2002), não tem sido suficiente para proporcionar aos estudantes os meios de significação da informação disponível na sociedade, objetivo maior da transposição didática acima aludida.

Apesar disso, observa-se que na história da humanidade, nunca tivemos tantos cientistas e engenheiros como nos tempos atuais; o avanço científico e tecnológico, produzido por estes cientistas e profissionais de áreas mais distintas, incluem setores que mesmo não sendo diretamente vinculados a uma formação matemática específica utilizam-se da matemática de certo nível de sofisticação (como é o caso da utilização, por psicólogos, de algoritmos estatísticos de descrição multidimensional). Isto denota que em todos os setores da sociedade os indivíduos, mesmo aqueles com "passado" escolar de reprovação em matemática (e mesmo "matofobia") estão sendo forçados cotidianamente a ressignificar a matemática e torná-la uma ferramenta útil, interessante e importante no seu dia-a-dia (o que não os impede de continuarem desprezando e detestando a matemática acadêmica, aquela que foi, e é "ensinada" na escola, e da qual têm péssima recordação).

Resumindo, apesar de a escola ser considerada como o lugar por excelência onde se realiza a educação e, sendo proposição dessa educação habilitar os integrantes da sociedade a tomarem posse de um corpo de um patrimônio de conhecimento historicamente acumulado e a ser compartilhado, percebe-se que há mais rupturas que continuidades no processo de ensino.

Vários pesquisadores da área de educação matemática (Vergnaud, 1983, Robert, 1998; Chevallard, 1985; Brousseau,1998; Huillet e Mutemba, 2000; Taylor,1999, entre outros) já chegaram a essas conclusões e atualmente buscam formas de ressignificar conteúdos com o propósito de tornar mais claro para os estudantes à importância de cada um para que realmente exista aprendizagem. Outros pesquisadores buscam nas organizações de sala de aula aspectos relevantes para o desenvolvimento do fenômeno educacional situado nesses espaços.

2.4 Expectativas acerca do estudante de matemática

Na nossa sociedade, as pessoas que escolhem profissões que têm suas bases fundamentadas na Matemática são percebidas de forma diferente e especial. Isto se dá, historicamente falando, pela importância atribuída a esta disciplina desde os tempos de Pitágoras. Mas, não são sem certa cobrança que as pessoas que se dedicam à matemática ganham seus status de "mais inteligentes" que os "seres humanos normais" que escolhem outras profissões fundamentadas em outras disciplinas.

Entretanto, quando se trata de estudantes, as expectativas dos professores em relação a estes são definidores, segundo Robert (1998), da postura didático-pedagógica apresentada pelos professores quando do trato com o conhecimento matemático. Eles se

espelham, na maioria das vezes, em atitudes desenvolvidas pelos matemáticos profissionais em seu fazer cotidiano.

Robert (1998) parte assim para mapear certas características habituais das práticas dos matemáticos profissionais, que segundo a autora serve de modelo aos professores. Ela esclarece que as práticas são diversas e dependem notadamente da subjetividade de cada profissional, do momento histórico, etc. Dessa forma, ela elenca algumas características que diferenciam o trabalho do matemático profissional em detrimento ao dos estudantes que ainda estão no ensino médio ou iniciando à universidade:

A primeira característica é a que os matemáticos têm a seu dispor um grande número de conhecimento e tendo estes já organizados e em pleno funcionamento cognitivo. Que segundo a autora, isto possibilita pela condição de poder procurar sistematicamente respostas para um dado problema, articulando todas as experiências com resolução de problemas as quais o matemático já se deparou, ou seja, ampliando sua possibilidade de resolução. Contudo, isto só é possível pelo acúmulo e organização dos conhecimentos disponíveis a estes profissionais.

Esta é uma das competências que os professores frequentemente sentem falta em seus estudantes, pois, eles parecem não conseguir fazer as relações entre os conteúdos que está sendo ensinado com outros supostamente já “aprendidos”.

A segunda é que os experts problematizam os objetos de conhecimento com maior eficiência, utilizando uma ampla quantidade de questionamentos, explorando assim os conhecimentos sistematizados com maior aproveitamento, mesmo que haja aspectos implícitos nos mesmos.

Assim, os experts estão sempre se questionando acerca de algum conceito, elaborando hipóteses, buscando refletir sobre a estrutura destes, questionando-se sobre a homogeneidade, a coerência, o caráter local ou global pertinente ao conceito, se ele é finito ou infinito, sobre a possibilidade de existência, de unidade, da “exaustividade”, etc. Em outras palavras, eles estão sempre articulando suas reflexões acerca de um determinado conceito às suas práticas; organizando e desorganizando todo conhecimento já elaborado, articulando saberes, reorganizando-os, quase que automaticamente, em uma procura incessante acerca da estruturação de soluções possíveis a tais questionamentos. Sendo assim que eles constroem seus “estoques” de conhecimento tornando-os disponíveis para serem “acessados” quando necessário.

Esse estoque de conhecimento e experiências caracteriza-se em particular pela existência de situações de referência, suficientemente familiares que lhes permite contar com os conhecimentos já dominados, em termos de experiência, e sobre os quais eles formam sua base conceitual e se apoiam quando surgem novos desafios ou questionamentos. Observa-se então que a falta desta competência impede uma maior desenvoltura dos estudantes quando expostos a situação formal de resolução de problemas.

O trabalho implementado pelo expert, nas fases mais avançadas de seu desenvolvimento, faz com que ele busque desenvolver relações, generalizações, mas também explorar as particularizações originais de cada novo conceito ou conhecimento que lhe chegue às mãos. Em sua busca pela formalização, eles conseguem revirar o problema e “tudo” o que existe acerca deste; eles fazem variar parâmetros, hipóteses, domínios de aplicação, o quadro em que o problema esta situado, os registros do

mesmo, os pontos de vista adotados, etc. Ele transforma, muda de estratégia e notadamente retorna ao início do problema na busca da compreensão do seu objeto de estudo e trabalho. Mas, também o expert se utiliza do cálculo, gastando um longo tempo nessa atividade, mesmo quando ele tem ferramentas, como o computador, para ajudá-lo.

Enfim, os experts utilizam a linguagem escrita, sendo esta de importância fundamental em seus trabalhos como fonte organizadora do conhecimento que está sendo apropriado. Dessa forma, a passagem do cálculo à forma escrita de explicação, com todo o formalismo necessário, engendra, às vezes, uma dinâmica de questionamento mais precisa, tornando possível uma maior exigência e um rigor de explicitação sobre o que está sendo escrito, notadamente, na fase final da redação.

Portanto, ousa-se dizer que os profissionais matemáticos podem ser considerados como exploradores intrépidos, corajosos, mas, muito bem ancorados em ferramentas matemáticas, e, portanto, com uma base muito bem organizada. Ao contrário do que se pode dizer do estudante do ensino médio e os estudantes iniciantes em cursos superior.

Porém, tais lacunas apresentadas pelos estudantes não ficam restritas aos aspectos conceituais e de compreensão e organização de conhecimento. Como salienta Robert (1998), essas dificuldades se ampliam, pois existem expectativas por parte dos professores, em termos de prática matemática, atingindo os níveis dos conteúdos e das habilidades complexas de manipulação de conhecimento.

Assim, e possível ver que as novas exigências em termos de demonstração e formalização estão geralmente relacionadas às novas e complexas demandas na atividade matemática, uma vez que os estudantes galgam mais um patamar na sua história acadêmica. Podendo, esses novos dados, em matéria de prática pessoal, ser visto tanto de forma empírica como a partir de estudos mais específicos e que buscam descobrir ocorrências complexa na atividade matemática.

Dessa forma, Robert (1998) relaciona algumas dessas novas demandas às práticas esperadas dos estudantes e as dificuldades encontradas na elaboração de tais práticas. Uma dessas novas demandas surge através da necessidade de demonstração, já no nível superior de resolução de problemas, que se espera que os estudantes consigam conduzir suas atividades de forma satisfatória, porém o que se observa é que esta não é uma habilidade já desenvolvida, mas, ainda em desenvolvimento e, dessa forma, tornam-se causa de dificuldades que se apresentam em situação de resolução de problemas.

As demonstrações se inscrevem cada vez mais no quadro da lógica elementar. Na universidade esta atividade passa a ser condição prévia implícita (estabelecida em contrato didático implícito, uma vez que é condição si ne qua non de sucesso), que tudo, em termos matemáticos, deverá ser demonstrado.

2.5 Mudanças operacionais ou um novo olhar sobre a formação e prática dos professores

Para que haja coerência entre a prática didático-pedagógica do professor e suas concepções é necessário que este reflita sobre os processos que perpassam as relações que se estabelecem dentro da sala de aula, ou seja, é preciso que o professor reflita sobre o que é conhecimento e como o adquirimos e, através de que processos, construímos o conhecimento. Também é preciso entender o que é aprendizagem, como o ser humano aprende e como o professor pode participar de forma a realmente ativa desse momento

do estudante. É a partir dessas reflexões que o sujeito, inicialmente apenas um formando, torna-se um professor, pois tais reflexões irão fundamentar sua prática pedagógica separando-o da reprodução de experiências enquanto estudantes e, no seu próprio processo de aprendizagem, realize para síntese de sua ação docente particular.

A base conceitual na qual o professor apoia suas reflexões serve como fundamento para o planejamento dos ambientes de aprendizagem. Este paradigma deverá ser suficientemente articulado para que possa responder as tanto em termos metodológicos como da pluralidade de estilos, notadamente, únicos em aspectos tanto psicológicos quanto experiencial.

2.5.1 O processo de formação do professor de ensino superior

Não são poucos os pesquisadores que buscam através de seus discursos incentivarem um processo de revisão e atualização dos cursos de formação de professores e mesmo das condições de trabalho destes, em todos os níveis de escolarização, mostrando que, mais que ampliação quantitativa de instituições de nível superior, existe a necessidade de se investir na qualidade, pois, o sujeito que procura se qualificar através dos cursos de formação, muitas vezes, não consegue nem adquirir a competência do saber da área de atuação, muito menos o saber pedagógico inerente e necessário à profissão.

Antes de ser uma "bênção" da democratização do ensino, os cursos de formação de professores tem uma grande parcela de responsabilidade na baixa qualidade da educação nacional, tornando-se, em muitos casos, motivo de gracejo da população, pois normalmente abraçam a profissão docente aquelas pessoas que não se dariam bem em profissões que exigissem saberes mais específicos, como Direito, Engenharia, Medicina etc[6].

Contudo, é importante salientar que ser professor demanda a articulação de saberes específicos das áreas de conhecimento e conhecimentos didáticos-pedagógicos, como salienta Pimenta e Anastasiou (2002), acerca das transformações das práticas docentes. Elas afirmam que tais práticas só se realizarão se o professor ampliar sua consciência sobre a própria prática, a de sala de aula e a da escola como um todo, o que pressupõe conhecimentos teóricos e críticos sobre a realidade. Enfatizando o que foi sugerido: a transformação da prática do professor decorrerá da ampliação de sua consciência crítica sobre essa mesma prática.

Entende-se que a democratização do ensino passa pelo desejo de mudança dos professores, por sua formação, por sua valorização profissional e por suas condições de trabalho. Vários são os pesquisadores que têm defendido a importância do investimento no desenvolvimento profissional da docência como um todo e da docência superior em particular.

A formação do professor deve ser antes de tudo epistemológica, ou seja, devendo-se reconhecer a docência como um campo de conhecimentos específicos necessários para a atuação dos agentes com coerência e clareza do fazer pedagógico cotidiano. Assim, a docência se constitui num campo específico de intervenção profissional na prática social.

---

[6] Nesse sentido, diz-se, popularmente, que "*quem sabe faz, quem não sabe ensina*"...

Pimenta e Anastasiou (2002) salientam que tal reflexão e prática não são utilizadas em muitos cursos superiores, onde pesquisadores dos vários campos de conhecimento como historiadores, químicos, filósofos, biólogos, cientistas políticos, físicos, matemáticos, artistas, etc. e profissionais de outras áreas de conhecimento como médicos, dentistas, engenheiros, advogados, economistas, etc; adentram o campo da docência como decorrência natural dessas atividades e por razões e interesses vários. Eles trazem os conhecimentos profissionais, porém não têm em sua formação os aspectos pedagógico-didáticos, logo, na maioria das vezes, não se questionam sobre os saberes próprios do ser professor. Semelhantemente, as instituições que os acolhem o fazem por serem o que são, desobrigando-se de torná-los professores. Esta situação gera traumas tanto no profissional/pesquisador-professor quanto no processo de ensino e aos seus resultados.

A ideia, aqui não é de culpabilização com relação aos profissionais e instituições, mas, de valorizar a formação e aperfeiçoamento dos docentes do ensino superior, buscando uma prática reflexiva e menos traumática para os envolvidos, ou seja, estudantes e professores.

Ser professor requer saberes e conhecimentos científicos, pedagógicos, educacionais, sensibilidade, indagação teórica, criatividade para encarar as situações ambíguas, incertas, conflituosas no contexto escolar. É da natureza da atividade docente proceder à mediação reflexiva e crítica entre as transformações sociais concretas e a formação humana dos estudantes, questionando os modos de pensar, sentir, agir e de produzir e distribuir conhecimentos.

Nos cursos de formação de professores é muito comum enfatizarem-se os saberes em detrimento dos deveres que os professores devem assumir junto aos seus estudantes, ou seja, o dever de estar sempre atualizados com as pesquisa de sua área de conhecimento e com aquelas que os capacitam na docência. O dever de assegurar aos seus estudantes a condição de cidadania que perpassa o ato próprio do aprender; o dever de assegurar uma aprendizagem significativa, fazendo-os perceber a importância do conhecimento científico-formal em suas atividades cotidianas.

Ao problematizar e analisar as situações da prática social de ensinar, o professor utiliza o conhecimento elaborado das ciências, das artes, da filosofia, da pedagogia e das ciências da educação como ferramenta para a compreensão e a proposição do real, tornando o trabalho docente impregnado de intencionalidade, pois este visa à formação humana por meio de conteúdos e habilidades, de pensamento e ação, o que implica escolhas, valores, compromissos ético. Isso significa introduzir objetivos de natureza conceitual, procedimental e valorativa, em relação aos conteúdos da matéria que ensina; transformar o saber científico ou tecnológico em conteúdos formativos; selecionar e organizar conteúdos de acordo com critérios lógicos e psicológicos, em função das características dos estudantes e das finalidades do ensino; utilizar métodos e procedimentos de ensino específicos, inserindo-os em uma estrutura organizacional em que participe de decisões e ações coletivas. Por isso, para ensinar, o professor precisa de conhecimentos e práticas que ultrapassem o campo de sua especialidade.

2.5.2 Algumas reflexões acerca da formação dos professores de Matemática do Ensino Superior

Apesar das sucessivas críticas ao ensino tradicional e de todos os avanços alcançados em pesquisas na área da Educação Matemática voltadas para a melhoria da prática pedagógica do professor e, consequentemente, favorecendo a aprendizagem matemática dos alunos, percebe-se que pouco mudou, ou seja, o ensino bancário ainda é amplamente utilizado nas escolas e nas universidades, pois há certa dificuldade em transpor os resultados obtidos com esses estudos para aplicação concreta em sala de aula (Camara Dos Santos, 2009). Tal fato favorece o distanciamento progressivo entre as pesquisas desenvolvidas e a realidade da sala de aula.

Desta forma, há um problema na maneira como a Matemática é ensinada nas escolas brasileiras. Tal fato se justifica pelos conteúdos que são trabalhados em sala de sala, que parecem ser desinteressantes, inúteis, e obscuros para os estudantes. Sobre esse assunto, Cardoso (2008, p.5) discute que:

> [...] a grande maioria dos conteúdos escolares não é, de fato, utilizada ou aplicada pelo aluno no seu cotidiano. Perde-se, assim, a oportunidade de reconhecer que o aluno traz consigo, para a sala de aula, uma bagagem cultural rica em conhecimentos matemáticos que podem ser o ponto de partida para a formação e ampliação desses conhecimentos.

É importante destacar que esta problemática ocorre tanto na Educação Básica como no Ensino Superior. Desta forma, quando falamos em pesquisas educacionais ou em formação continuada, não nos remetemos apenas aos professores dos ensinos fundamental e médio, mas, também, estamos incluímos o ensino superior, pois este apresenta, em grande escala, várias lacunas referentes ao processo de ensino e aprendizagem da Matemática, que podem influenciar, em muita das vezes, na permanência e desistência dos graduandos nos cursos superiores.

Para tanto, o distanciamento progressivo pode ser percebido com muita ênfase também nas universidades brasileiras. Tal distanciamento consiste na dificuldade de articulação entre os resultados obtidos nas pesquisas educacionais e a realidade em sala de aula, ocasionando um empilhamento de reflexões advindas de pesquisas sem que com isso haja uma reflexão critica direcionada à sala de aula.

Há alguns anos atrás, se questionavam que a maioria dos professores não tinha acesso a essas informações, isto é, aos resultados obtidos com as pesquisas educacionais. Desta forma, o conhecimento construído com tais pesquisas ficava restrito apenas a universidade e aos professores idealizadores e outros pesquisadores envolvidos nos estudos. Tal fato representava um grande obstáculo para a aplicação das experiências estabelecidas nestes trabalhos com o contexto da sala de aula.

Contudo, hoje, há diversas opções pelas quais os professores podem ter acesso aos dados das pesquisas, principalmente em eventos científicos (congressos, encontros, simpósios, etc.), a exemplo dos ENEM's (Encontros Nacionais de Educação Matemática), EREM's (Encontros Regionais de Educação Matemática) etc., nos quais tais profissionais tem a possibilidade de participarem de diversas atividades (minicursos, palestras, mesas-redondas, comunicações orais, entre outros) realizadas por pesquisadores de todas as partes do país, tornando desta forma, um fórum de discussões e socialização de experiências educacionais.

A partir dessa temática Ascoli e Brancher enfatizam a importância e a relevância que o acesso as informações de variada ordens deverá fazer parte do processo de formação do professor que ensina nas diversas modalidades de ensino.

> Atualmente, a maior parte da população tem acesso às informações com muita facilidade. As crianças e adolescentes recebem estímulos diferenciados de antigamente, pois estão constantemente em contato com as novas tecnologias, o que resulta em reações e aprendizagens diferenciadas. Tais realidades exigem das escolas e dos educadores práticas pedagógicas mais motivadoras, como músicas, danças, brincadeiras, jogos, teatros, enfim atividades lúdicas que estimulem os educandos a construírem uma aprendizagem mais significativa. (Ascoli; Brancher, 2006, p.1)

Hoje, com o grande avanço tecnológico, temos a disposição um grande leque de informações, das mais variadas, obtidas apenas com um simples clique do mouse do computador, assim, tanto professores como alunos podem ter acessos a esses subsídios. O que nota-se é que muitos professores não conseguem articular os resultados das pesquisas educacionais com o contexto de sua sala de aula, por não refletirem adequadamente sobre sua atuação como docente.

Muito embora, a situação acima descrita possa ser analisada com os professores de maneira geral, no ensino superior isto é bastante evidente, pois no caso dos cursos de licenciatura em Matemática, parte considerável do corpo docente são bacharéis em matemática (ou possuem mestrado e doutorado em Matemática Pura), sendo que sua própria formação universitária não o possibilitou de realizar tal reflexão. Desta forma, abre-se aqui uma nova vertente a ser trabalhada na formação continuada do professor de Matemática, buscando impulsionar reflexões que ultrapassem o modo do fazer tradicional tão comum nas salas de aula no ensino superior.

Sobre a metodologia e didática docente praticada no ensino da Matemática, é fato que entre a maioria dos professores do ensino superior (e também os da educação básica) o ensino tradicional é bastante evidente, no qual há uma ênfase na memorização dos conteúdos escolares e na repetição mecânica do que se “estudou” em sala de aula, sem a preocupação do real entendimento (compreensão/aprendizagem) pelo aluno, o que favorece o fracasso na Matemática e, posteriormente, abandono do curso.

E, geralmente, quando é discutido sobre a formação de professores a indagação que se faz é...

> Quando tratamos do assunto “formação de professores”, pensamos imediatamente na formação para a docência nos ensinos fundamental e médio, raramente remetemos tal pensamento à formação de professores universitários, causa uma impressão de que para este nível não é necessário uma formação. Porém, uma das críticas mais comuns dirigidas aos cursos superiores diz respeito à didática dos professores universitários, ou seria melhor dizer, à falta dela. (Utrera ,2011, p.5)

Desta forma, o professor de nível superior deve fazer uma reflexão sobre sua prática pedagógica, em especial sobre seus processos didáticos e metodológicos, promovendo aprendizagem aos estudantes e não sentimentos de medo e desistência.

Como citado anteriormente, a maioria dos professores de Matemática dos cursos de Licenciatura em Matemática é constituída por bacharéis, mestres e doutores em Matemática. Contudo, tais profissionais lecionam as disciplinas como se não estivem

formando professores e sim bacharéis, desta forma, tais profissionais são excelentes matemáticos, porém como professores de Matemática apresentam várias lacunas.

Desta forma, evidencia-se que muitos dos problemas existentes na educação atual são creditados aos professores e suas práticas reprodutivista, a-crítica, não reflexiva, antiquada, pouco transformadora etc (Malta, 2004; Pimenta e Anastasiou, 2002; Santos, 1994, entre outros). Entretanto, é preciso reconhecer que muito do que acontece em sala de aula deve-se ao padrão de formação prévia desses professores, não se podendo, portanto, focar em responsabilidades do professor, esquecendo-se do quadro institucional de formação e exercício profissional do mesmo.

Feitas as presentes considerações de caráter introdutório, passa-se na seção seguinte do presente estudo à proposta de operacionalização metodológica.

3 Metodologia

O presente estudo se propõe a refletir sobre as razões das desistências e da permanência de estudantes ingressos nos cursos de Licenciatura em Ciências com Habilitação em Matemática e de Licenciatura em Matemática ambos do Centro de Formação de Professores (CFP) da Universidade Federal de Campina Grande (UFCG), campus Cajazeiras, entre 2011 e 2013 (período da pesquisa), observando os fatores que implícita e explicitamente contribuem para uma possível ruptura existente no desenvolvimento da aprendizagem matemática dos estudantes. Tal ruptura se revela nos comportamento matemático dos ingressos no curso supracitado, causando uma queda no desempenho dos estudantes que iniciam nesses cursos.

Tal objetivo geral pode ser explicitado em termos dos seguintes objetivos específicos:

- Identificar os comportamentos dos estudantes ingressos no curso de Matemática acerca da disponibilidade e compreensão sobre o curso que iniciam;
- Compreender como os alunos se percebem em seu processo de aprendizagem da matemática no ensino superior;
- Compreender como os alunos percebem o processo de ensino dos professores que ensinam matemática nos ensino superior.
- Discutir os fatores que contribuem para o sucesso e o fracasso nos cursos supracitados.

Tendo em vista o atendimento destes objetivos iremos utilizar como ferramenta de análise dos dados a Análise Estatística Implicativa Dessa forma, apresentar-se-á, nesta seção, os critérios de seleção para a construção da amostra e o procedimento de análise dos dados.

3.1 Construção da amostra

Este estudo contou com 22 alunos dos cursos de Licenciatura em Ciências[7] com Habilitação em Matemática[8] (entre o 6º e 10º períodos) e de Licenciatura em

[7]No curso de Ciências, a disciplina de Cálculo I é oferecida no 5º semestre

[8]No curso de Matemática, a disciplina de Cálculo I é ofertada no 2º semestre. Sendo que a turma "mais avançada" encontra-se no 4º período.

Matemática (3º e 4º períodos), ambos do CFP/UFCG, campus Cajazeiras, entre 2011 e 2013 (período da pesquisa). O questionário foi aplicado com alunos que cursavam o segundo semestre de 2011 e que se disponibilizaram a responder o questionário. Para responder ao questionário era necessário ter cursado a disciplina de Cálculo 1 (aprovado ou reprovado). Procuramos também observar a influência do gênero como variável suplementar para observar se existe alguma influência nos resultados. Tivemos uma participação 12 indivíduos do sexo Masculino e 10 do feminino.

O instrumento de coleta de dados foi um questionário de sondagem e auto avaliação para os estudantes dos cursos supracitados e que continha treze questões das quais dez eram quantitativas/qualitativas e três qualitativas. O objetivo do questionário foi obter subsídios referentes ao desempenho destes estudantes em termos de desempenho no curso; auto avaliação dos estudantes acerca de seus desempenhos antes e durante os primeiros períodos dos cursos de Licenciatura em Ciências com Habilitação em Matemática e Licenciatura em Matemática; autoconceito dos estudantes e investimento semanal nas disciplinas de matemática.

3.2 Procedimento de análise dos dados

Neste trabalho os dados foram analisados, buscando relacionar, os 13 itens (aqui transformados em variáveis) que constavam no questionário respondido pelos 22 sujeitos participantes da pesquisa com o intuito de se verificar as relações existente entre sujeitos e variáveis. Tais relações foram tratadas a partir da abordagem da Análise Estatística Implicativa – A.S.I.; utilizando-se o software CHIC[9] na versão 6.0; O qual tem como função essencial "extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecer um índice de qualidade de associações e de representar uma estrutura das variáveis obtida por meio de regras" (Couturier et al, 2004, p.1 apud Andrade, Régnier, Lima, 2014, p. 1057).

As variáveis da análise correspondem as perguntas do questionários. Na tabela 1, apresentamos as codificações das variáveis. A variável 4, trata das respostas a uma pergunta, ela transformada em 5 variáveis (V4a, V4b....) em função da resposta apresentada.

[9] O termo CHIC é uma abreviatura de: Classificação Hierárquica Implicativa e Coesiva (Gras et al, 2013).

Código	Descrição
V01AvMed	Como você se avaliava enquanto aluno Matemática no Ensino Médio? (Muito bom =1; Bom =2/3; Ruim = 1/3; Muito Ruim=0)
V02AvOut	Como as pessoas avaliam você enquanto aluno de Matemática? (Muito bom =1; Bom =2/3; Ruim = 1/3; Muito Ruim=0)
V03ProfAv	Como seus professores do Ensino Médio avaliavam você? (Muito bom =1; Bom =2/3; Ruim = 1/3; Muito Ruim=0)
V04reprov	Você já foi reprovado em alguma disciplina na universidade? Sim=1 ; Não =0
V04aCal1	Foi reprovado em Cálculo 1? Sim=1; Não =0
V04bCal2	Foi reprovado em Cálculo 2? Sim=1; Não =0
V04cMat1	Foi reprovado em Matemática I ? Sim=1; Não =0
V04dOdis	Foi reprovado em outras disciplinas ? Sim=1; Não =0
V05AutAv	Como você se avalia agora depois de ter entrado na universidade e ter cursado Cálculo I? (Muito bom =1; Bom =2/3; Ruim = 1/3; Muito Ruim=0)
V06AvProf	Como você avalia seus professores de Cálculo? (Muito bom =1; Bom =2/3; Ruim = 1/3; Muito Ruim=0)
V07selet1	Você entrou na universidade após o seu primeiro processo seletivo? Sim=1; Não =0
V08Crep	Você sabia que o seu curso tem um índice de reprovação alto? Sim=1; Não =0
V08aCdif V08bOres	Por que você acha que seu curso tem um índice de reprovação alto? Curso difícil: Sim=1; Não =0; Outra resposta: Sim=1; Não =0
V9aDed V9bOres	Quais os fatores que são determinantes para o sucesso no seu curso? Dedicação (persistência): Sim=1; Não =0; Outra resposta: Sim=1; Não =0.
V9cFded V9dOres	Quais os fatores que são determinantes para o fracasso no seu curso? Falta dedicação (persistência): Sim=1; Não =0; Outra resposta: Sim=1; Não =0.
V10a+2h V10b-2h	Quantas horas você dedica, por dia, para estudar as disciplinas da universidade? Mais 2h: Sim=1; Não =0; Menos 2h: Sim=1; Não =0;
V11aAfi V11bOmt	Por que você veio fazer o curso de Matemática? Por afinidade: Sim=1; Não =0; Outro motivo: Sim=1; Não =0.
V12aCom V12bOmt	Qual a maior diferença que você sentiu, em se tratando de conteúdos, ao entrar na universidade? Complexidade [dos conteúdos do ensino superior]: Sim=1; Não =0; Outro motivo: Sim=1; Não =0.

Tabela 1 – Codificação das variáveis desta pesquisa

4 Resultados e análises

Para análise dos dados com o uso do CHIC selecionamos as opções: nós significativos; cálculo longo; implicação segundo a teoria clássica; lei binominal. Apresentamos a seguir os resultados com as análises.

4.1 Análise de similaridade

Para nossa análise, vamos considerar os índices acima de 0,71 que é estatisticamente mais significativa.

Na figura 1, temos como variáveis com maior similaridade as variáveis V04cMat1 e V09cFded com nível de similaridade de (0,97) e que apresenta o nó mais significativo.

Estas variáveis correspondem a “Foi reprovado em matemática 1” (V04c) e “Quais os fatores que são determinantes para o fracasso no seu curso? Falta de dedicação e persistência”(V09c) isto demonstra que o fato de ter sido reprovado em Matemática I tem uma forte ligação com a falta de dedicação o que contribui para o fracasso nos cursos objetos dessa pesquisa.

O que se pode dizer acerca da relação acima citada é que para que os estudantes, dos cursos que tem a Matemática como disciplina central, cheguem a obter sucesso em seus cursos precisam, o mais cedo possível, desenvolver estratégias, comumente utilizadas pelos matemáticos profissionais, segundo Robert (1998), transformando conceitos aparentemente soltos em uma grande “teia” de saberes e assim fazerem as relações entre os conteúdos que está sendo ensinado com outros supostamente já “aprendidos”.

O segundo nível de similaridade é o formado pelas variáveis V04aCal1 e (V04cMat1 e V09cFded) com nível (0.93) de similaridade. Estas variáveis correspondem a “Foi reprovado em cálculo 1” (V04aCal1) e as variáveis “Foi reprovado em Matemática I” (V04cMat1) e o que determina o fracasso no seu curso é a “a falta de dedicação (persistência)” (V09cFded) o que reforça o que já foi dito na análise anterior de que para os estudantes que responderam ao questionário a reprovação em disciplinas está diretamente ligado a falta de dedicação aos estudos, em particular, nos cursos elencados, fornecem subsídios importantes para o fracasso.

No terceiro nível, formado pelas variáveis V04dOdis e V11bOmt com nível, com um índice de 0.89 de similaridade. Estas variáveis correspondem a “Foi reprovado em outras disciplinas” (V04d_Odis) e “Por que você veio fazer o curso de Matemática? Outro motivo” (V11b_Omt). Nesse nível, apesar de haver uma relação alta entre as variáveis, não foi identificada uma razão, com base no estudo, para essa relação.

No quarto nível, formado pelas variáveis V07selet1 e V08bOres, que apresentam um nível de similaridade (0.83). Estas variáveis correspondem a “Você entrou na universidade após o seu primeiro processo seletivo?” (V07selet1) e “Por que você acha que seu curso tem um índice de reprovação alto? Outra_resposta” (V08b_Ores). Nesse nível há uma clara relação entre entrar na universidade no primeiro processo seletivo, ou seja, se haver nenhum trauma de reprovação na entrada da universidade e uma resposta diferente da ideia de que o curso tem alto índice de reprovação por ser difícil. Ou seja, não ter sido traumatizado com reprovações no início da carreira acadêmica deu uma ideia diferente a respeito do curso para os estudantes em questão.

No quinto nível de similaridade formado pelas variáveis V04reprov e V04b_Cal2 (0.75) de similaridade, estas variáveis correspondem a “Você já foi reprovado em alguma disciplina na universidade? Sim=1” (V04reprov) e “Foi reprovado em Cálculo 2? Sim=1” (V04bCal2), aqui cabe a mesma análise realizada no primeiro nível de similaridade.

No sexto nível de similaridade formado pelas variáveis V08aCdif e V12aCom com (0.71) de similaridade, estas variáveis correspondem a “Por que você acha que seu curso tem um índice de reprovação alto? Curso_difícil” (V08a_Cdif)” e “Qual a maior diferença que você sentiu, em se tratando de conteúdos, ao entrar na universidade?”, que teve a “Complexidade dos conteúdos do ensino superior” (V12a_Com) como resposta válida. Nesse nível pode-se ver que há uma forte coerência ao relacionar a dificuldade trazida pelo curso, que tem um alto índice de reprovação com a complexidade dos conteúdos tratados, pois com foi visto nas falas dos professores do ensino superior,

segundo Malta (2004, p 41) " os alunos estão ingressando nas universidades cada vez mais despreparados" e alguns professores participantes dessa pesquisa: P1:"*o aluno não chega à universidade com uma 'bagagem' suficiente para lidar com a estrutura da Matemática do nível superior*", ou P3 :"*a Matemática da universidade é muito mais abstrata que a do ensino médio, logo falta base para que haja compreensão nesse novo nível de abstração*", ou ainda P1: "*faltam os pré-requisitos necessários para que o aluno se saia bem nessa Matemática superior*".

No sétimo nível de similaridade formado pelas variáveis V09a_Ded e V10a-2h com (0.71) de similaridade, estas variáveis correspondem a "Quais os fatores que são determinantes para o sucesso no seu curso? Dedicação (persistência)" (V09a_Ded) e "Quantas horas você dedica, por dia, para estudar as disciplinas da universidade? Menos de 2h" (V10a-2h). Observamos uma similaridade entre quem considera a dedicação como determinante para o sucesso no curso e a quantidade de horas investidas em estudo por dia. Quem estuda menos de duas horas considera que, apesar disso, para ter sucesso no curso é necessário se dedicar mais horas no curso.

No oitavo nível de similaridade formado pelas variáveis V09bOres e V10a+2h com (0.71) de similaridade, estas variáveis correspondem a "Quais os fatores que são determinantes para o sucesso no seu curso? Outra resposta" com "Quantas horas você dedica, por dia, para estudar as disciplinas da universidade? Mais de duas horas". Esse resultado indica que quem respondeu que já estuda mais de duas horas por dia considera outros elementos que a dedicação para o sucesso no curso, uma vez que já se dedica mais de 2 horas por dia em estudos.

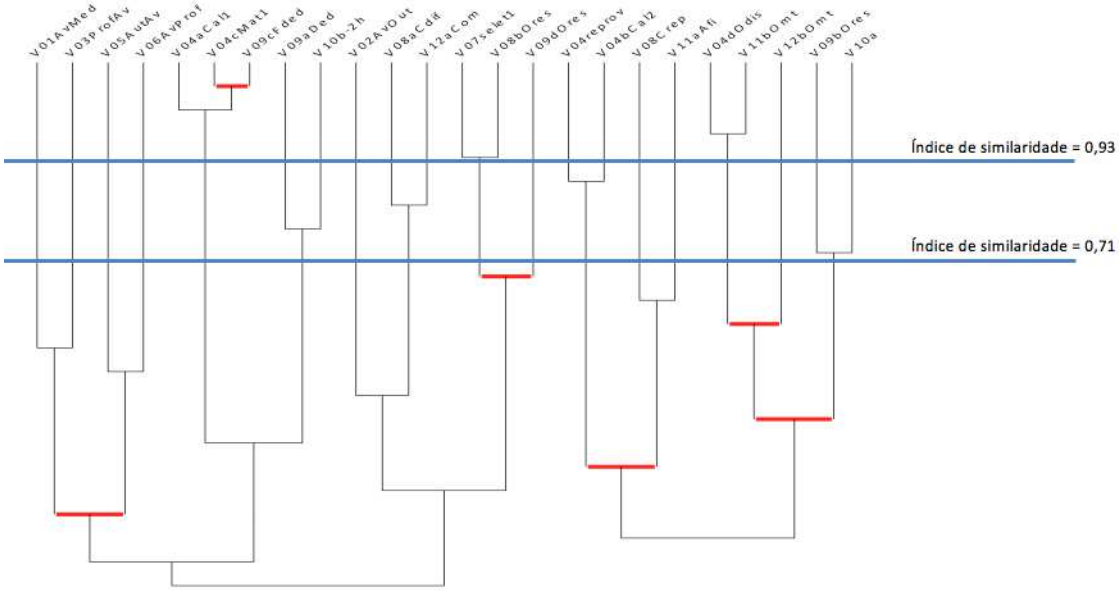


Figura 1 – Árvore de similaridade com os resultados apresentados pelo CHIC.

4.2 Análise do gráfico implicativo

Na figura 2, apresentamos o resultado do gráfico implicativo calculado segundo a teoria clássica e a lei binominal. Foram considerados o índice mínimo de 0,90 com a cor azul e 0,80 com a cor verde de modo a indicar as implicações mais fortes.

Acrescentamos ao gráfico implicativo o valor usando com as cores usadas nos caminhos.

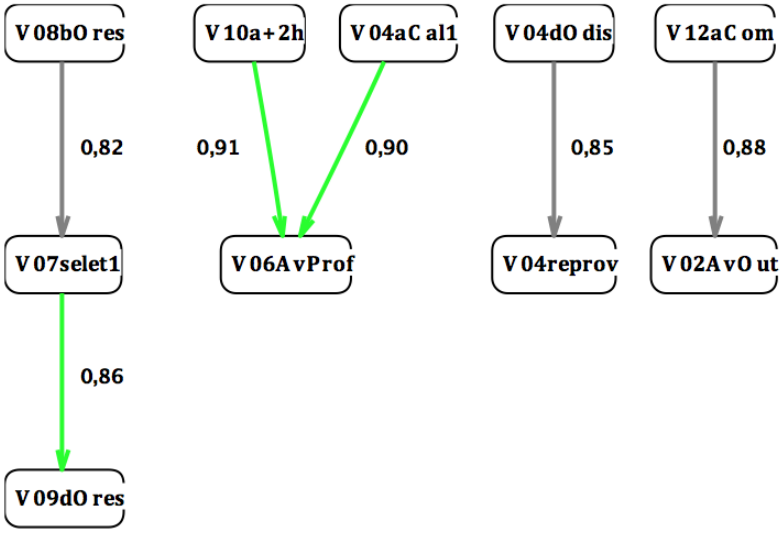


Figura 2 – Gráfico Implicativo.

Apresentamos os valores das relações indicadas no gráfico da figura 2 na figura 3 (copiado da tela do CHIC). Os valores marcados com um retângulo correspondem as implicações indicadas no gráfico da figura 2. As demais, são relações do ponto de vista estatístico menos significativas.

4.2.1 Interpretação dos resultados do gráfico implicativo

O caminho V08b_Ores⇒V07selet1⇒V09Ores mostra o maior índice de implicação respectivamente de (0,82 e 0,86) indica que há uma implicação entre se conceber que o alto índice de reprovação do curso advenha de outras possibilidades diferentes da dificuldade dos cursos, o fato de ter sido aprovado no primeiro processo seletivo implica em outra possibilidade diferente da falta de dedicação para se chegar a fracassar nos cursos citados. Pode-se dizer, então, que quem foi aprovado no seu primeiro processo seletivo percebe que os altos índices de reprovação dos cursos tem outras causas que não a dificuldade já observada na grande maioria das pessoas ao lidar com a matemática. Talvez, por não ter tantas dificuldades na área de matemática, pois conseguiu ingressar sem maiores dificuldades na universidade percebe que o que está relacionado ao fracasso no curso são outros aspectos não detectados nesse estudo, já que para esses estudantes a matemática tem desenvolvido outras competências que os ajudou a resolver problemas cotidianos, formulando-os e agindo matematicamente como está posto em Brasil (1998).

A rede V10a_+2h ⇒ V06AvProf⇐V04aCal1 mostra o segundo maior índice de implicação respectivamente de (0,91 e 0,90). Temos uma ligação entre ser reprovado em Cálculo 1 (V04aCal1) e uma avaliação positiva do professor de Cálculo (V06AvProf) e estudar mais de 2horas por dia. Esta implicação indica que o fato de ser reprovado em Cálculo 1 não modifica a avaliação positiva do professor de cálculo e que é necessário para obter sucesso nos curso um investimento de mais de duas horas diárias

de estudo(V10a_=+2h). Assim, o fato de ser reprovado não desvaloriza o esforço do professor que leciona a disciplina, mostrando que quem passa mais de 2 horas estudando por dia também avalia de forma positiva o professor de Cálculo.

O caminho V04d_Odis⟹V04_reprov sinaliza o terceiro maior índice de implicação (0,85). Está implicação, pode indicar que o fato de ter sido reprovado em outras disciplinas, ligadas ou não a matemática, é um forte indicador para o fracasso na universidade.

O caminho V12a_Com ⟹V02avOut apresenta o quarto maior índice de implicação (0,88), indicando que apesar da complexidade dos conteúdos apresentados no ensino superior, se o estudante é bem avaliado em matemática pelas pessoas em seu entorno isto o levará ao sucesso nesse nível de ensino. Esta variável (V12a_Com) Qual a maior diferença que você sentiu, em se tratando de conteúdos, ao entrar na universidade? Complexidade [dos conteúdos do ensino superior] implica em que para o estudante ter sucesso no ensino superior ele precisará considerar que no caso da matemática, se faz necessário que compreendam as estruturas, as ideias e os métodos matemáticos e não simplesmente apliquem fórmulas ou algoritmos, mais do que isto que é possível ver que há novas exigências em termos de demonstração e formalização e estas estão geralmente relacionadas às novas e complexas demandas na atividade matemática, uma vez que os estudantes galgam mais um patamar na sua história acadêmica. Portanto, ter uma boa avaliação externa nas competências supracitadas propiciará um investimento maior para se obter o sucesso desejado.

4.3 Análise da árvore coesitiva

Na figura 4, temos a árvore coesitiva com os resultados apresentados indicando os nós significativos. Para elaborar o gráfico eliminamos algumas variáveis que indicavam relações não desejáveis (implicação entre duas possíveis respostas da mesma variável) Com base neste gráfico podemos observar algumas classes. Analisaremos apenas aquelas que estejam com uma coesão acima de 0,80. Apresentamos uma análise das classes considerando a sua ordem de qualidade decrescente.

V07selet1 ⟹ V09dOres: classificação ao nível 1, com uma coesão de 0,956. Indica que quem entrou na universidade no seu primeiro processo seletivo (V07selet1) tende a responder que os fatores determinantes para o fracasso no seu curso (V09dOres) são outros que a falta de dedicação e persistência. Isto pode ser explicado pela competência já adquirida principalmente nas disciplinas de Matemática vinculadas aos cursos, levando os estudantes a colocar em outro pólo a razão para o fracasso.

V04dOdis ⟹ V04reprov: classificação ao nível 2, com uma coesão de 0,916. Indica que quem foi reprovado tende a ter sido reprovado em disciplinas que não estão diretamente ligadas as disciplinas matemática mostrando que muitas vezes o que leva ao fracasso nos cursos perpassam por outras disciplinas, ligadas ou não a matemática, como física, disciplinas do eixo de formação, demonstrando que para se obter o sucesso é importante se dedicar de forma integral, pois cada disciplina tem seu peso e importância para a formação de futuros professores.

V10a+2h ⟹ V06AvProf: classificação ao nível 3, com uma coesão de 0.899. Indica que quem estuda mais de 2horas por dia tende a ter uma avaliação positiva do professor de Cálculo (V06AvProf). Isto pode ser explicado por uma das expectativas

que se tem do aluno no ensino superior, ou seja, que ele seja suficientemente autônomo ao ponto de que o professor seja realmente um mediador de conhecimento e que o estudante busque no seu dia a dia suprir suas lacunas.

V09cFded⟹ V04aCal1: classificação ao nível 4, com uma coesão de 0.858. Indica que quem fracassa no curso por falta de dedicação tende a ser reprovado em Cálculo 1. Isto pode ser explicado por ser está a disciplina matemática de recepção dos estudantes advindos do ensino médio. Ou seja, a primeira disciplina da Matemática que será trabalhada no ensino superior, logo, é a partir de Cálculo 1 que os estudantes realmente passaram a conviver com uma maior complexidade.

Além disso, estudos como os de Santos (1994), de Arruda & Ueno (2003) e Malta (2004) fazem supor a existência de certo desnível ou não-compatibilização entre o ensino da matemática no nível médio e no ensino superior. Tal desnível parece decorrer de dois tipos de fatores: deficiência, pura e simples, de conteúdo (ou seja, não se está ensinando no ensino médio aquele mínimo de conhecimentos que deveria ser ensinado), e mudança abrupta de contrato didático (o ambiente psicossocial de funcionamento da sala de aula de matemática do ensino superior seria muito diferente, em termos "contratuais", do ambiente de ensino do nível médio).

V12aCom⟹ V02AvOut: classificação ao nível 5, com uma coesão de 0.848. Indica que os estudantes que entende que a maior diferença entre o ensino médio e o superior, em se tratando de conteúdos, ao entrar na universidade é a complexidade [dos conteúdos do ensino superior] são bem avaliados pela sociedade enquanto alunos de matemática. Pode-se dizer que apesar da complexidade dos conteúdos apresentados no ensino superior, se o estudante já tem um bom relacionamento com a matemática advindo dos anos em que cursou o ensino fundamental e médio, ele tenderá a se sair bem, pois compreende que é necessário dedicação e estudo para continuar a obter sucesso nesse nível de ensino.

V08bOres⟹ (V07selet1⟹ V09dOres): classificação ao nível 6, com uma coesão de 0.814. Indica que quem entrou na universidade no seu primeiro processo seletivo (V07selet1) tende a responder que os fatores determinantes para o fracasso no seu curso (V09dOres) são outros que a falta de dedicação e persistência e que entende que o alto índice de reprovação do curso advenha de outras possibilidades diferentes da dificuldade dos cursos, o fato de ter sido aprovado no primeiro processo seletivo e outra possibilidade diferente da falta de dedicação para se chegar a fracassar nos cursos citados.

V11bOmt⟹ V12bOmt: classificação ao nível 7, com uma coesão de 0.808. Indica que quem veio fazer os cursos por outros motivos diferentes da afinidade com os mesmos não percebe a complexidade dos conteúdos que são exigidos para se obter sucesso. Isto muitas vezes acontece com os estudantes que apenas obtém notas suficiente para ingressar na universidade e aceitam a situação, muitas vezes sem refletirem se realmente gostariam de se aprofundar naquela área de conhecimento. Infelizmente, quando isto acontece a possibilidade de fracasso é grande.

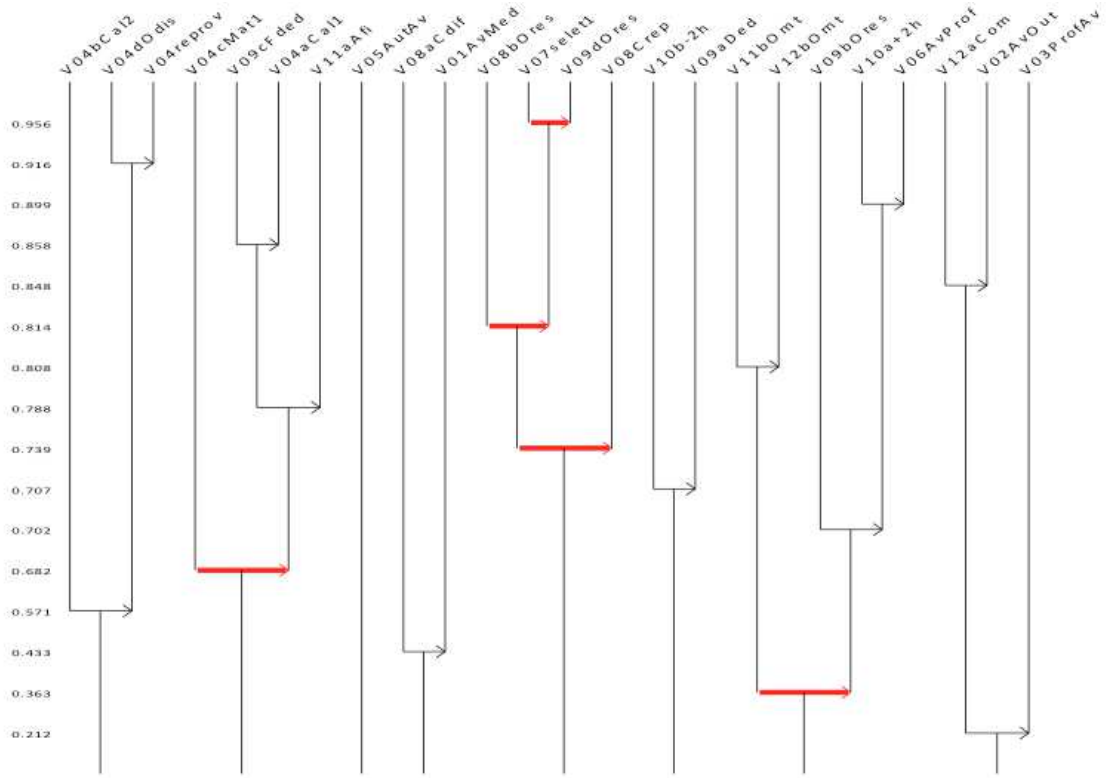


Figura 3. Árvore coesitiva

5 Considerações finais

No estudo, ora relatado, verificou-se que os estudantes compreendem a dificuldades e também a necessidade pessoal de investimento. Assim, observou-se que os estudantes afirmam que para obter sucesso nos seus cursos é necessário ter dedicação e perseverança, ficando evidente que apenas o próprio esforço é o diferencial para um bom desempenho ao longo do curso na universidade. Tal fato pode ser explicado a partir das relações de similaridade, implicação e coesão entre as variáveis apresentadas.

Viu-se, então, que as variáveis que possuíram o maior nível similaridade foram as variáveis V04cMat1 e V09cFded (0,97). Estas variáveis correspondem a "Foi reprovado em matemática 1" (V04c) e "Quais os fatores que são determinantes para o fracasso no seu curso? Falta de dedicação e persistência"(V09c) isto demonstra que o fato de ter sido reprovado em Matemática I tem uma forte ligação com a falta de dedicação o que contribui para o fracasso nos cursos objetos dessa pesquisa. Isto indica também que os participantes da pesquisa percebem que a persistência e dedicação são fatores importante para o sucesso nos cursos elencados

Pode-se dizer, dessa forma, que para os estudantes cheguem a obter sucesso em cursos que tem a Matemática como disciplina central precisam, o mais cedo possível, desenvolver estratégias comumente utilizadas pelos matemáticos profissionais, segundo Robert (1998), transformando conceitos aparentemente soltos em uma grande "teia" de saberes e assim fazer as relações entre os conteúdos que está sendo ensinado com outros supostamente já "aprendidos". Além de compreender que precisam mudar sua

atitude em relação a matemática vivida até o ensino médio, pois no ensino superior os conteúdos envolvem mais complexidade.

No gráfico implicativo O caminho V08b_Ores ⟹ V07selet1 ⟹ V09Ores mostrou o maior índice de implicação respectivamente de (0,82 e 0,86) indica que há uma implicação entre se conceber que o alto índice de reprovação do curso advenha de outras possibilidades diferentes da dificuldade dos cursos, o fato de ter sido aprovado no primeiro processo seletivo e outra possibilidade diferente da falta de dedicação para se chegar a fracassar nos cursos citados. Pode-se dizer, então, que quem foi aprovado no seu primeiro processo seletivo percebe que os altos índices de reprovação dos cursos tem outras causa que não a dificuldade inerente. Tal situação talvez se explique, por não ter tantas dificuldades na área de matemática, pois conseguiu ingressar no curso para o qual se preparou e o qual tem o desejo de investir, o que está relacionado com outras possibilidades para o fracasso no curso, já que para esses estudantes a matemática tem desenvolvido outras competências que os ajudou a resolver problemas cotidianos, formulando-os e agindo matematicamente como sugere Brasil, (1998).

A árvore coesitiva mostrou que o caminho V07selet1 ⟹ V09dOres com classificação ao nível 1, e uma coesão de 0,956, Indicou que quem entrou na universidade no seu primeiro processo seletivo (V07selet1) tende a responder que os fatores determinantes para o fracasso no seu curso (V09dOres) são outros que a falta de dedicação e persistência. Isto pode ser explicado pela competência já adquirida principalmente nas disciplinas de Matemática vinculadas aos cursos. Esse nível se compatibilizou com o primeiro nível de implicação, demonstrando uma coerência entre os dois instrumento.

Com a intensão de fechamento desse estudo pode-se dizer que muitos dos fatores que conduzem ao sucesso do estudantes nos cursos objeto desse estudo é a persistência e dedicação aos cursos, procurando refletir sobre as escolhas feitas no momento de se escolher a profissão que irá abraçar, além de ter consciência de suas lacunas conceituais e procurar saná-las, desenvolvendo, assim, uma atitude autônoma bastante exigida dos estudantes do ensino superior.

Referências

[1] Andrade, V. L. V. X. de ; Régnier, J-C ; Brito Lima, A. P. de A. (2014), Apresentação dos dados em atividades que envolvem as medidas de tendência central e de dispersão nos livros didáticos do ensino médio no Brasil e na França. Análise das formas e das suas organizações no quadro da A.S.I. Educação Matemática Pesquisa (Online), v. 16, p. 1047-1067.

[2] Arruda, S. M. ; Ueno, M H. (2003), Sobre o ingresso, desistência e permanência no curso de Física da Universidade Estadual de Londrina: algumas reflexões. Pg. 159 – 175. Revista Ciência & Educação. São Paulo: Escrituras; ISSN nº 1516-7313; vol. 9; nº 2.

[3] Ascoli, C. C. B; Brancher, V. R.. (2006), Jogos matemáticos: algumas considerações sobre os processos de ensino e aprendizagem. In: 12ª Jornada Nacional de Educação, 2006, Santa Maria. Anais da 12ª Jornada Nacional de Educação,

Educação e sociedade: perspectivas educacionais do século XXI. Santa Maria: UNIFRA.

[4] Brasil. (2001), Conselho Nacional de Educação. Diretrizes Curriculares Nacionais Educação Básica. Brasília,.

[5] Brasil.. (1996), Lei de Diretrizes e Bases da Educação Nacional. Lei nº 9394/96.

[6] Brasil..(1998), Ministério da Educação. Parâmetros Curriculares Nacionais - Ensino Médio/Matemática. Brasília: MEC/SEF,.

[7] Brasil.. (2004), Ministério da Educação. Catálogo do Programa Nacional do Livro para o Ensino Médio. Brasília: MEC/SEF.

[8] Cabral, T. C. B.; Baldino, R. R.(2004), O ensino de matemática em cursos de engenharia de sistemas digitais. Em: H. N. Cury (Org.). Disciplinas matemáticas do ensino superior: reflexões, relatos, propostas. (pp. 139 –186). Porto Alegre: EDIPUCRS.

[9] Camara Dos Santos, M. (2009), O Cabri-Géomètre e o desenvolvimento do pensamento geométrico: o caso dos quadriláteros. In: BORBA, R.; GUIMARÃES, G. (Org.). A Pesquisa em Educação Matemática: Repercussões na sala de aula. São Paulo: Cortez.

[10] Cardoso, E. R.(2008), Jogos matemáticos no contexto escolar. Trabalho de Conclusão de Curso (Aperfeiçoamento em Programa de Desenvolvimento Educacional) – Secretaria de Estado de Educação do Paraná. SEED: Maringá.

[11] Da Rocha Falcão, J. T., Meira, L. A.(1994), A experiência matemática na escola de 1º grau. Em: A educação matemática em revista. SBEM–Ano 1 nº 2.

[12] Freire, P. (1996) Pedagogia da Autonomia: saberes necessários à prática educativa. São Paulo: Paz e Terra.

[13] Huillet, D; Mutemba, B.(2000), The relation of mozambican secondary school teacher to a mathematical concept: the case of limits of functions. Research report completo submetido ao 24th International Meeting of Psychology of Mathematics Education (PME), Hiroshima (Japan).

[14] Malta, I.(2004), Linguagem, leitura e matemática. Em: H. N. CURY (Org.). Disciplinas matemáticas do ensino superior: reflexões, relatos, propostas. (pp. 41 – 62). Porto Alegre: EDIPUCRS.

[15] Maia, L. de S. L.(2009), Vale a pena ensinar Matemática. In: R. BORBA, G. Guimarães (Org.). A Pesquisa em Educação Matemática**:** Repercu ssões na sala de aula. São Paulo: Cortez.

[16] Nascimento, J. C.(2003), O conceito de limites em Cálculo: obstáculos e dificuldades de aprendizagem no contexto do ensino superior de matemática. Tese de doutorado não publicada; Pós-Graduação em Psicologia – UFPE, Recife.

[17] Nimier, J.(1988), Les modes de rélations aux mathématiques. Paris: Méridiens-Klincksieck.

[18] Pimenta, S. G.; Anastasiou, L. G. C. (2002), Docência no ensino superior. São Paulo: Cortez.

[19] Robert, A.(1998), Outils d`analyse des contenus mathématiques à enseigner au Lycée et L` Université. Recherches em didactique des mathématiques, vol. 18, nº 2, pp. 139-190. França: editions La Pensée Sauvage.

[20] Santos, R. M.(1994), Avaliação do desempenho no processo de ensino-aprendizagem de Cálculo Diferencial e Integral I: o caso da Universidade Federal do Ceará. Dissertação de Mestrado.

[21] Silva, J. A. M.(2002), Educação matemática e exclusão social: tratamento diferenciado para realidades desiguais. Brasília: Plano Editora.

[22] Silva, M. C. R., Vendramini, C. M. M.(2005), Autoconceito e desempenho de universitários na disciplina estatística. Psicol. Esc. Educ. (Impr.), Campinas, v. 9, n. 2, Dec. . <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-85572005000200008&lng=en&nrm=iso>. access on 18 Mar. 2013. http://dx.doi.org/10.1590/S1413-85572005000200008

[23] Taylor, J. A. (1999), Undergraduate Mathematics and the role of Mathematics learning support. site: www.pci.usq.edu.au/staff/spundi/delta99/papers/taylor.pdf

[24] Utrera, R. G.(2011), A preocupação com a desistência dos alunos no ensino superior em decorrência da falta de pré-requisitos. Monografia (Especialização em Docência e Pesquisa para o Ensino Superior) – Programa de Pós-Graduação da Universidade Metropolitana de Santos. UNIMES: Santos.

Sommaire

Introduction – L'analyse statistique implicative : des sciences dures aux sciences humaines et sociales 4 – 10
Jean-Claude REGNIER, Régis GRAS

Conférences

Un survol paradigmatique de l'analyse statistique implicative 11 – 23
Régis GRAS

Big Data et Datamining: entre mythes et réalités 24
Yahia SLIMANI

Regards sur les rapports entre statistique et enseignement mathématique 25 – 28
François PLUVINAGE

A – Développement du cadre théorique A.S.I.

Chapitre 1 – Notion de champ implicatif en analyse statistique implicative 29 – 46
Régis GRAS, Pascale KUNTZ, Nicolas GREFFARD

Chapitre 2 – Un mariage arrangé entre l'implication et la confiance ? 47 – 68
Régis GRAS, Raphaël COUTURIER, Pablo GREGORI

Chapitre 3 – Analyse d'un questionnaire « enseignants de mathématiques » par différentes méthodes 69 – 80
Régis GRAS et Antoine BODIN

Chapitre 4 – Variable nodale et cône implicatif 81 – 98
Dominique LAHANIER-REUTER, Régis GRAS, Marc BAILLEUL

Chapitre 5 – Hiérarchie de règles en a.s.i. et conceptualisation 99 – 116
Régis GRAS, Nadja Maria ACIOLY-RÉGNIER

Chapitre 6 – Ajout de la confiance au graphe implicatif 117 – 129
Souhila GHANEM et Raphaël COUTURIER

Chapitre 7 – Extension de l'analyse statistique implicative au cas des variables continues quelconques 130 – 142
Régis GRAS, Jean-Claude RÉGNIER

Chapitre 8 – Classifying objective interestingness measures based on the tendency of value variation 143 - 172
Nghia QUOC PHAN, Hiep XUAN HUYNH, Fabrice GUILLET, Régis GRAS

Chapitre 9 – Medidas alternativas de los grados de adhesión de individuos a la implicación y similaridad para variables modales 173 – 190
Larisa ZAMORA y Pablo GREGORI

Chapitre 10 – Variability of GRAS classical implication index and other rule quality measures under small size sampling from bivariate binary processes 191 – 209
Pablo GREGORI, Raphaël COUTURIER

Chapitre 11 – Pourquoi et comment transformer des variables quantitatives en qualitatives ? Application à la mélodie de la langue française 210 – 242
Martine CADOT, Anne BONNEAU

Chapitre 12 – Apport de la combinaison de la méthode d'analyse statistique implicative (ASI) avec la théorie de réponses aux items (IRT) 243 – 262
Hayette KHALED, Raphaël COUTURIER

B – Application aux recherches en didactiques disciplinaire et professionnelle

Chapitre 13 – La coordinación de los procesos de aproximación en la comprensión del límite de una función en un punto. Una aproximación a través del análisis implicativo 263 – 279
Joan PONS, Julia VALLS, Salvador LLINARES

Chapitre 14 – Les supports d'enseignement dans la représentation du métier chez des professeurs des écoles débutants. 280 – 294
Marc BAILLEUL, Laurence LEROYER

Chapitre 15 – Uma análise das diferentes praxeologias no ensino das equações de segundo grau. um olhar a partir da quadro teorico da A.S.I 295 – 325
Marcus Bessa DE MENEZES, Marcelo Câmara DOS SANTOS

Chapitre 16 – Professores que ensinam matemática nos anos iniciais do ensino fundamental: Análise das Tendências de Pesquisa no Brasil (2006-2014) 326 – 355
Luciana SILVA DOS SANTOS, Marcelo CÂMARA DOS SANTOS, Nadja Maria ACIOLY-RÉGNIER, Edênia Maria RIBEIRO DO AMARAL

Chapitre 17 – The influence of the nature of geometrical figures on geometric proofs. 356 - 368
Paraskevi MICHAEL-CHRYSANTHOU, Athanasios GAGATSIS

Chapitre 18 – L'approche socioconstructiviste dans les situations d'enseignement-apprentissage de la biologie en Tunisie : le cas de la reproduction humaine 369 – 385
Fadhila FARHANE HORRIGUE

Chapitre 19 – Étude des significations données à la notion de fraction par des élèves de CM1 et de CM2 de l'École primaire en France 386 – 401
Abdul Aziz ALAHMADATI

Chapitre 20 – Análise das pesquisas didáticas sobre função afim no Ensino Fundamental e Médio no quadro da análise estatística implicativa 402 – 416
Aveilson José DE SANTANA, Vladimir LIRA VERAS XAVIER DE ANDRADE, Jean-Claude RÉGNIER

Chapitre 21 – Représentations et valeurs implicites des enseignants gabonais sur l'éducation à la santé 417 – 434
Laurence NDONG

Chapitre 22 – L'enseignement de la résolution de problèmes mathématiques à l'école primaire au Maroc : représentations des enseignants à l'égard de leurs pratiques 435 – 450
Brahim EL MEKAOUI, Fadhila FARHANE HORRIGUE

Chapitre 23 – Mestrado profissional em ensino de matemática: articulação de competências 451 – 477
Silvia Maria DE AGUIAR ISAIA, Eleni BISOGNIN, Vanilde BISOGNIN, Jean-Claude REGNIER, Nadja ACIOLY-RÉGNIER, Andréia CARDOSO SILVEIRA

C – Application aux recherches du champ de la psychologie

Chapitre 24 – Le traitement statistique de données et la complexité de l'expression et du développement différentiels du sujet psychologique 478 – 501
Christian PELLOIS

Chapitre 25 – Le sujet psychologique : la complexité différentielles de son expression et de son développement au regard du traitement statistique de données 502 – 520
Christian PELLOIS

Chapitre 26 – Bem-estar subjetivo de estudantes brasileiros da área de ciências exatas: abordagem dos dados no quadro da análise estatística implicativa 521 – 535
Fabiana FERREIRA, Céline FAURE, Sofiane BOUZID

D – Application aux recherches du champ des TICE

Chapitre 27 – Les potentiels pédagogiques de l'usage du tableau numérique interactif à l'Ecole primaire en France à la lumière de l'analyse statistique implicative 536 – 560
Hassan ALCHEGHRI

Chapitre 28 – Mapeamento de conhecimentos de professores sobre Tecnologias de Informação e Comunicação e seus usos didático-pedagógicos 561 – 589
Saddo AG ALMOULOUD, Cileda DE QUEIROZ E SILVA COUTINHO
Maria José FERREIRA DA SILVA

E – Application aux recherches en Histoire de l'art

Chapitre 29 – Divers motifs dans les représentations de l'Ascension du Christ en occident entre le IXe et le XIIIe siècle. Étude de liens quasi implicatifs 590 – 602
Magali GUENOT, Jean-Claude REGNIER

F – Meta-analyse des publications contributives à l'A.S.I.

Chapitre 30 – A formação humana dos educadores atraves das comunicações nos colóquios internacionais de Análise Estatística Implicativa (A.S.I.): casos dos colóquios A.S.I.5, A.S.I.6 e A.S.I.7 603 – 623
Djailton PEREIRA DA CUNHA, Nadja Maria ACIOLY-RÉGNIER,
Aurino LIMA FERREIRA

G – Application aux recherches du champ de la sociologie de l'éducation

Chapitre 31 – Expectativas de inserção no mercado de trabalho dos estudantes cotistas e não cotistas da universidade federal da Bahia-Brasil: um estudo no quadro da análise estatística implicativa 624 – 646
Andréia CARDOSO SILVEIRA, Jean-Claude RÉGNIER,
Nadja Maria ACIOLY-RÉGNIER, Robinson MOREIRA TENÓRIO

Chapitre 32 – A questão da permanência e desistência dos estudantes de licenciatura em ciências e matemática no brasil. Estudo exploratório no CFP/UFCG abordado pelo quadro da análise estatística implicativa 647 – 670
Valéria Maria DE LIMA BORBA,
Anna Paula DE AVELAR BRITO LIMA,
Jean-Claude RÉGNIER

Actes du 8ème d'Analyse Statistique Implicative

Edition : ARSA 2015

Bibliothèque Nationale de Tunisie

ISBN: 978-9973-9819-0-5