UBFC Université Bourgogne Franche-Comté

IUT Belfort-Montbéliard

Université Lumière Lyon 2

Editeurs : Jean-Claude Régnier, Régis Gras, Antoine Bodin, Raphaël Couturier, Gérard Vergnaud †

LABEX ASLAN Université de Lyon

icar Interactions, Corpus, Apprentissages, Représentations UMR 5191 CNRS • ENS de Lyon • Université Lyon 2

Лаборатория социокогнитивной лингвистики и обучения иноязычному дискурсу
Laboratory of Sociocongnitive Linguistics and Foreign Language Discourse Teaching

National Research Tomsk State University

Analyse statistique implicative

Analyses quali-quantitatives des liens orientés entre variables et/ou groupes de variables

Éditeurs :

Jean-Claude Régnier
Régis Gras
Antoine Bodin
Raphaël Couturier
Gérard Vergnaud (1933- †6 juin 2021)

Actes du 11ème colloque d'Analyse Statistique Implicative

Edition :
Université Bourgogne Franche-Comté - Besançon

Version numérique ISBN : 978-2-9562045-5-8

Version imprimée ISBN : 978-2-9562045-4-1

Année 2021

Version en date du : 5/11/2021

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 :

- IUT de Belfort - Montbéliard Université Bourgogne – Franche Comté (Belfort France)
- Université Lumière Lyon2 – Université de Lyon (Lyon France)
- Laboratoire de recherche : Unité Mixte de Recherche UMR 5191 – ICAR (Lyon France)
- Labex ASLAN Université de Lyon
- UMR 6004 -LS2N Laboratoire des Sciences du Numérique de Nantes

ou leur parrainage scientifique *:*

- Université d'Etat de Tomsk (Tomsk - Sibérie - Russie)
- Société française de Statistique (SFdS France),

- ARSA – Association de Recherche en Statistique Appliquée (Tunisie)
- Association Extraction et Gestion des Connaissances (EGC) (France),
- Association pour la Recherche en Didactique des Mathématiques (ARDM France),
- Centro de Estudios e Investigación Enzo Faletto (USACH – Chili)
- UMR 6004 -LS2N Laboratoire des Sciences du Numérique de Nantes (France)
- G.R.I.M. (Gruppo di Ricerca sull'Insegnamento delle Matematiche), Dipartimento di Matematica, Università di Palermo (Italie)
- Groupe de Formation et Recherche sur l'Analyse Statistique Implicative – Programme Science sans frontières (UFRPE – Recife – Brésil)
- PUC - Pontifícia Universidade Católica de São Paulo (Brésil)
- REVISTA AKADEMEIA (Santiago de Chili)
- REVISTA EMP - Educação Matemática Pesquisa (São Paulo- Brésil)
- REVISTA VIDYA – Ensino das Ciências e de Matemática (Santa Maria – Brésil)
- Société Francophone de Classification (SFC France)
- Laboratoire Société Éducation Travail - Université de Tizi-Ouzou

Président du comité scientifique : Jean-Claude Régnier

Vice-Président du comité scientifique : Antoine Bodin
Fondateur : Régis Gras
Président d'honneur : Gérard Vergnaud († 6 juin 2021)
Président du comité d'organisation : Raphaël Couturier

INDEX DES AUTEURS

Ailin Gamoneda.....311
Aitzol Lasa.....149
Ana Maria Dos Anjos Carneiro-Leão 169
André Totohasina 79, 97, 199, 293
Andrianasy Angelo Djistera 67, 293
Antoine Bodin 191
Aurino Lima Ferreira.....227
Bakys Bruno Ralahady 67, 293
Bruce Masonova Solozafy Bemena....97
Christian Pellois.....327
Christian Razanatsoavina 67, 293
Cristiana Marinho Da Costa 255
Daniel Rajaonasy Feno.....79, 97
Eleni Deliyianni.....125
Erika Pedersen 311
Federico Vasches.....311
Flávia Pereira Righi.....275
Germán Nieto 109
Gilbert Saporta.....11
Greice Scremin 275
Haritz Iribas 149
Iliada Elia.....125
Iván Benzaquén 311
Jaione Abaurrea 149
Jean-Claude Laberche.....67, 293
Jean-Claude Régnier5, 169, 191, 255, 361, 399
Jeanne Guiet-Silvain..... 377
Jeanne Ravaosolo..... 67
Josoa Michel Tovohery..... 79
Larisa Zamora Matamoros..... 49
Leonardo Dalla Porta..... 275
María Cardieri..... 311
Miguel Wilhelmi..... 149
Nadja Maria Acioly-Régnier ... 169, 227
Nelsa María Sagaró Del Campo 49
Océane Touron..... 399
Olga Belletich 149
Pablo Carranza..... 311
Pablo Gregori..... 109
Parfait Bemarisika 199
Paula Virgínia Chaves Cabral Andrade 227
Quércia Carvalho Eloi 255
Rafael Santos De Aquino..... 169
Raphaël Couturier..... 39, 199
Régis Gras..... 5, 13, 39, 191
Saddo Ag Almouloud 39
Sofia Anastasiadou 13
Tolga Tekin..... 361
Vladimir Lira Véras Xavier De Andrade 169, 227, 255, 437

Sommaire

REMERCIEMENTS 1
Président du comité scientifique : Jean-Claude Régnier 1

INDEX DES AUTEURS II
INTRODUCTION DE L'OUVRAGE : ANALYSE STATISTIQUE IMPLICATIVE. ANALYSES QUALI-QUANTITATIVES DES LIENS ORIENTÉS ENTRE VARIABLES ET/OU GROUPES DE VARIABLES . 5
Jean-Claude Régnier, Régis Gras 5

CONFÉRENCE INVITÉE : INTERPRÉTABILITÉ DES MODÈLES PRÉDICTIFS 11
Gilbert Saporta 11

LA HIERARCHIE COHESITIVE : AVATAR FRACTAL DES THEORIES DU DEVELOPPEMENT 13
Régis Gras, Sofia Anastasiadou 13

FRACTALITÉ DU RYTHME DU TEMPS SUR L'ARBRE DE VIE 39
Régis Gras, Raphaël Couturier, Saddo Ag Almouloud 39

METODOLOGÍA DE APLICACIÓN DEL ANÁLISIS ESTADÍSTICO IMPLICATIVO EN LAS INVESTIGACIONES MÉDICAS DE CAUSALIDAD 49
Nelsa María Sagaró Del Campo, Larisa Zamora Matamoros 49

LES STRATÉGIES DE RIPOSTES, DONT LES PLANTES TRADITIONNELLES, À LA PANDÉMIE DE COVID-19 CHEZ DES ÉTUDIANTS UNIVERSITAIRES MALGACHES :CAS DE L'UNIVERSITÉ D'ANTSIRANANA 67
Christian Razanatsoavina, Bakys Bruno Ralahady, Andrianasy Angelo Djistera, Jeanne Ravaosolo, Jean-Claude Laberche 67

UNE TECHNIQUE D'EXTRACTION DES RÈGLES D'ASSOCIATION QUANTITATIVES BASÉE SUR LE COEFFICIENT DE VARIATION 79
Josoa Michel Tovohery, André Totohasina, Daniel Rajaonasy Feno 79

UNE ÉTUDE COMPARATIVE DES PROPRIÉTÉS DES MESURES MGK ET ODD-RATIO 97
Bruce Masonova Solozafy Bemena, André Totohasina, Daniel Rajaonasy Feno 97

COMPARISON OF TWO CLUSTER ANALYSIS OF VARIABLES: STATISTICAL IMPLICATIVE ANALYSIS VS CLUSTOFVAR 109
Germán Nieto, Pablo Gregori 109

ANALYZING MULTIPLE-REPRESENTATION FLEXIBILITY AND PROBLEM-SOLVING ABILITY THROUGH LONGITUDINAL EMPIRICAL DATA: INSIGHTS FROM IMPLICATIVE ANALYSIS ... 125
Eleni Deliyianni, Iliada Elia 125

TRIANGULATION EN RECHERCHE QUALITATIVE À L'AIDE DE L'ANALYSE STATISTIQUE IMPLICATIVE 149
Miguel Wilhelmi, Olga Belletich, Haritz Iribas, Jaione Abaurrea, Aitzol Lasa . 149

ANÁLISE ESTATÍSTICA IMPLICATIVA E ANÁLISE DE VARIÂNCIA: ESTUDO ESTATÍSTICO COMPARATIVO SOBRE O DESEMPENHO ESCOLAR EM SALA DE AULA MULTICULTURAL ... 169
Rafael Santos De Aquino, Ana Maria Dos Anjos Carneiro-Leão Nadja Maria Acioly-Régnier Vladimir Lira Véras Xavier De Andrade Jean-Claude Régnier 169

ÉTUDE DE LA DISTANCE ENTRE GRAPHES IMPLICATIFS POUR LA PRISE EN COMPTE DE LA TEMPORALITÉ EN ANALYSE STATISTIQUE IMPLICATIVE 191
Régis Gras, Antoine Bodin, Jean-Claude Régnier, ... 191

UNE CONSTRUCTION CONDENSÉE ET INTERACTIVE DU GRAPHE IMPLICATIF D'UNE BASE DE DONNÉES 199
Parfait Bemarisika, Raphaël Couturier,André Totohasina..... 199

TENDÊNCIAS DE PESQUISAS SOBRE RESILIÊNCIA EDUCACIONAL EM TESES E DISSERTAÇÕES NO BRASIL: ANÁLISE À LUZ DA ASI. 227
Paula Virgínia Chaves Cabral Andrade Aurino Lima Ferreira, Nadja Maria Acioly-Régnier, Vladimir Lira Véras Xavier De Andrade 227

COMPETÊNCIAS SOCIOEMOCIONAIS (CSE) NA EDUCAÇÃO: TENDÊNCIAS DA PRODUÇÃO CIENTÍFICA NAS EDIÇÕES DO CONEDU 2014-2020 À LUZ DA A.S.I..... 255
Cristiana Marinho Da Costa, Quércia Carvalho Eloi, Vladimir Lira Véras Xavier De Andrade, Jean-Claude Régnier. 255

CONCEPÇÕES SOBRE O ENSINO DE ÁLGEBRA: UMA INVESTIGAÇÃO A PARTIR DA ANÁLISE ESTATÍSTICA IMPLICATIVA..... 275
Flávia Pereira Righi, Leonardo Dalla Porta, Greice Scremin 275

CHANGEMENT RADICAL DES MÉTHODES D'ENSEIGNEMENT UNIVERSITAIRES VERS LE NUMÉRIQUE FACE À LA PANDÉMIE DE COVID-19 À MADAGASCAR : CAS DE L'UNIVERSITÉ DE ANTSIRANANA..... 293
Bakys Bruno Ralahady , Christian Razanatsoavina, Andrianasy Angelo Djistera, Jean-Claude Laberche, André Totohasina 293

ESTUDIO SOBRE VIOLENCIA EN RELACIONES SEXO-AFECTIVAS DE JÓVENES DE LA PROVINCIA DE NEUQUÉN- ARGENTINA MEDIANTE A.S.I. 311
Pablo Carranza, Federico Vasches, Ailin Gamoneda, María Cardieri, Iván Benzaquén, Erika Pedersen 311

MODÉLISATION DU DÉVELOPPEMENT COGNITIF MISE À L'ÉPREUVE PAR L'ASI : APPROCHE GÉNÉRALE, APPROCHE DIFFÉRENTIELLE ET QUESTIONS MÉTHODOLOGIQUES 327
Christian Pellois 327

AFFORDANCE DE TWITTER POUR LES USAGES SOCIO-PROFESSIONNELS PAR LES ENSEIGNANTS 361
Tolga Tekin, Jean-Claude Régnier..... 361

L'INCERTITUDE DES SCHÈMES POURTANT BIEN ORGANISÉS 377
Jeanne Guiet-Silvain 377

INTERACTIONS ENTRE ÉLÈVES DANS LA CLASSE NON MAITRISÉES PAR L'ENSEIGNANT. ÉTUDE EXPLORATOIRE PAR UNE ENQUÊTE PAR QUESTIONNAIRE AUPRÈS D'ENSEIGNANTS DE L'ÉCOLE PRIMAIRE..... 399
Océane Touron, Jean-Claude Régnier,..... 399

LE LOGO ASI 11 COMMENTÉ… 437
Vladimir Lira Véras Xavier De Andrade 437

INDEX DES CONTACTS DES AUTEURS..... 442

INTRODUCTION DE L'OUVRAGE :

ANALYSE STATISTIQUE IMPLICATIVE. ANALYSES QUALI-QUANTITATIVES DES LIENS ORIENTÉS ENTRE VARIABLES ET/OU GROUPES DE VARIABLES

Jean-Claude Régnier[12], Régis Gras[3]

En 2021, avec la tenue du 11ème colloque ASI à Belfort, nous continuons à mesurer le parcours de 43 années de développement de ce qui fut considéré comme une méthode puis un cadre théorique, et dénommé Analyse statistique implicative, depuis la soutenance de thèse de doctorat[4] de Régis Gras, point de départ repérable de sa fondation en 1979. 21 années depuis le premier colloque ASI en 2000 se sont déjà écoulées.

En cette période de novembre 2021, le procès qui s'est ouvert à Paris pour juger les auteurs des horreurs du 13 novembre 2015, nous remémore que le 8ème colloque ASI8 s'était tenu à Radès en novembre 2015 dans un contexte générateur d'inquiétudes qui avaient traversé notre communauté scientifique ASI-iste en raison des attentats commis sur le territoire même de la Tunisie, au point de soulever des questions sur le maintien de l'évènement. Mais les actions criminelles menées ont ceci de terribles qu'elles semblent frapper à l'aveugle pour un regard extérieur. Or ce fut avec une détermination cruelle et meurtrière que ces auteurs ont commis leur crime le 13 novembre 2015 sur le territoire même de la France. Nous le répétons aujourd'hui pour ne pas oublier les souffrances des femmes et des hommes blessés directement ou indirectement dans leur chair et dans leur cœur. Nos pensées leur sont de nouveau adressées avec une espérance exprimée par un cri déjà centenaire : *plus jamais ça* !

Ce retour vers le 8ème colloque ASI de 2015, nous évoque aussi le souvenir de notre collègue et ami François Pluvinage qui avait été notre conférencier invité. Il nous a quittés en mars 2020. Il fut parmi ceux et celles qui ont soutenu le développement de l'Analyse statistique implicative par son regard critique et éclairé, en particulier, comme membre du comité scientifique des 9ème et 10ème colloques ASI en 2017 et 2019.

Les années 2020 et 2021 furent marquées par un évènement mondial, celui de la pandémie Covid 19, fortement relayé par les médias, qui ne fut pourtant pas unique dans l'histoire puisque déjà en 1923 dans la pièce de théâtre *Knock ou le triomphe de la médecine*[5] de Jules Romains, le personnage du docteur Parpalaid évoquait celle de 1918, la tristement célèbre grippe espagnole de 1918 qui aurait pu s'appeler grippe américaine[6], mais aussi celle de 1890. Notons que les impacts de la pandémie Covid-19 ont fait l'objet de deux communications par nos collègues de Madagascar pour ce colloque ASI11.

En 2021, nous avons toutefois décidé de la mise en œuvre le projet de l'organisation du 11ème colloque ASI qui avait été formulé à l'issue du 10ème colloque. Pour des raisons

---

1 Président du comité scientifique de ASI11 - UMR 5191 ICAR, Université de Lyon - Lyon 2 (France)

2 Laboratory of sociocognitive linguistics and foreign language discourse teaching - National Research Tomsk State University

3 Fondateur - Polytech, LS2N, Université de Nantes (France)

4 Gras, R. (1979), Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques, Thèse d'état, Université de Rennes 1.

5 Romains, J. (1924) *Knock ou le triomphe de la médecine* Paris : éditions Gallimard p.26-27 – Edition de poche de 1963

6 « La grippe espagnole, la grande tueuse » France2 – 29/09/2021 à 22h50 dans « Infrarouge »

pratiques, nous avons maintenu le même lieu que pour les deux précédents colloques, c'est-à-dire à Belfort en France. Nous avons aussi prévu que ce colloque se déroule selon une modalité hybride en présence et à distance.

La préparation de notre 11ème colloque a, de nouveau, été endeuillée par la disparition de notre ami et collègue Gérard Vergnaud qui nous a quittés le 6 juin 2021. Il fut un soutien important tout au long des 43 années de l'histoire de l'Analyse statistique implicative. Il fut membre du comité scientifique depuis le 4ème colloque ASI en 2007. Il fut notre conférencier invité lors du 9ème colloque ASI en 2017. Il était notre président d'honneur de l'actuel colloque ASI 11.

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

Il ressort que l'Analyse Statistique Implicative peut être définie 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*. »[7] (Gras, Régnier, 2009 p.12). Le présent ouvrage est constitué des articles issus d'un appel à contributions lancé dans le cadre du onzième colloque sur l'Analyse Statistique Implicative – ASI11 - organisé à Belfort en novembre 2021. Ces articles ont été soumis à la lecture critique[8] 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.

ASI et le Comité scientifique :

- Nadja **Acioly-Régnier**, EA4571 Éducation, Cultures, Politiques Université de Lyon - ESPé- Université Lyon 1 (France)
- 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)
- Antoine **Bodin**, Consultant en matière de curriculum et d'évaluation dans le domaine des mathématiques (France)
- Martine **Cadot**, LORIA, Nancy, (France)
- Dante **Castillo Guajardo** Centro de Estudios e Investigación Enzo Faletto - Universidad de Santiago de Chile (Chili)
- Cileda **de Queiroz e Silva Coutinho** PPG Educação Matematica - Pontifícia universidade Católica de São Paulo (Brésil)
- Raphaël **Couturier**, FEMTO-ST département DISC Université de Franche-Comté (France)
- Leonardo **Dalla Porta** Universidade Franciscana de Santa Maria (RS - Brasil)
- Eleni **Deliyianni** Ministry of Education, Culture and Sport Nicosia of Cyprus (Chypre)
- Ahmed **Dhouibi** , Institut supérieur des études technologiques de Radès (Tunisie)
- Benedetto **Di Paola**, Département de mathématiques - Université di Palermo (Italie)

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 Editions

8 Chaque contribution a été anonymement soumise à trois relecteurs dont les comptes rendus d'expertise ont été retournés anonymement aux auteurs.

- Jean-Jacques **Droesbeke**, Université Libre de Bruxelles (Belgique)
- 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**, LS2N UMR 6004, Equipe DUKe - École Polytechnique de l'Université de Nantes (France)
- Michel **Henry**, IREM de Franche-Comté (France)
- Tatiana **Kabanova,** National Research Tomsk State University - Tomsk (Sibérie, Russie)
- Pascale **Kuntz**, LS2N UMR 6004, Equipe DUKe - École Polytechnique de l'Université de Nantes (France)
- Vladimir **Lira Veras Xavier de Andrade**, PPG Ensino das Ciências e da Matemática - UFRPE Recife (Brésil)
- Samuel Edmundo **Lopez Bello**, Universidade Federal do Rio Grande do Sul - UFRGS (Brésil)
- Maria del Pilar **Orus Baguena**, IMAC - Universitat Jaume I de Castellón de la Plana (Espagne)
- Jean-Louis **Piednoir**, Inspecteur Général Honoraire de mathématiques MEN (France)
- André **Totohasina** , Université d'Antsiranana, (Madagascar)
- 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 non symétrique 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, 7 et 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 : https://sites.univ-lyon2.fr/asi/5/ Revue QRDM[9] - 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, une 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 : https://sites.univ-lyon2.fr/asi/6/ Revue QRDM5 - 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 : https://sites.univ-lyon2.fr/asi/7/

Le huitième colloque ASI8 s'était tenu à Radès (Tunisie) les 11-14 novembre 2015. Il est possible d'accéder aux publications : https://sites.univ-lyon2.fr/asi/8/

Le neuvième colloque ASI9 s'était tenu à Belfort (France) les 4-7 octobre 2017. Il est possible d'accéder aux publications : https://sites.univ-lyon2.fr/asi/9/

Le dixième colloque ASI10 s'était tenu à Belfort (France) les 2-5 octobre 2019. Il est possible d'accéder aux publications : https://sites.univ-lyon2.fr/asi/10/

A.S.I. – Analyse statistique implicative : publications majeures.

Rappelons qu'en 2009, 2013, 2017 et 2018, quatre ouvrages majeurs francophones ont été publiés chez Cépaduès à Toulouse :

- *Analyse Statistique Implicative. Une méthode d'analyse de données pour la recherche de causalités*[10].
- *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire pour la recherche de causalités*[11]
- *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales*[12]
- *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*[13].

Ils permettent de faire un point global sur les actualisations significatives de l'ouvrage de 1996, *L'implication statistique*, édité par La Pensée Sauvage Editeur (Grenoble France). Une

9 Quaderni Di Ricerca In Didattica

10 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

11 Gras, R., Régnier, JC., Marinica, C., Guillet, F. (Eds) (2013) *Analyse Statistique Implicative. Méthode exploratoire et confirmatoire pour la recherche de causalités*. Toulouse : Cépadues

12 Gras, R., Régnier, JC., Lahanier-Reuter, D. Marinica, C., Guillet, F. (Eds) (2017) *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociale*s. Toulouse : Cépadues

13 Gras, R., (2018) *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*. Toulouse : Cépadues

autre publication majeure anglophone a été réalisée avec l'ouvrage[14] de 2008, *Statistical Implicative Analysis* diffusé par Springer-Verlag, (Berlin-Heidelberg, Allemagne). Ajoutons une publication en espagnol qui a aussi été produite avec l'ouvrage[15] de 2009 *Teoria y Aplicaciones del Analisis Estadistico Implicativo*, diffusé par Universitat Jaume-1, (Castellon Espagne). Enfin en 2020 l'ouvrage en langue portugaise du Brésil produit par Régnier, J.C., Lira Veras Xavier de Andrade, V. (Eds) (2020). *Análise Estatística Implicativa e Análise de Similaridade no Quadro Teórico e Metodológico das Pesquisas em Ensino de Ciências e Matemática com a utilização do software CHIC.* Brasileira Editora Universitaria da UFRPE. Version numérique: ISBN 978-65-86466-11-9

Dans ce onzième colloque A.S.I.11 de 2021 à Belfort, les dimensions internationales et interculturelles, initiées dès ASI5 par l'introduction de cinq idiomes de travail : anglais, espagnol, français, italien et portugais, ont été maintenues, certes non sans difficultés et obstacles à dépasser, mais parce que cela nous semble fondamental pour enrichir nos champs et nos lieux de recherche. Les idiomes arabe, chinois, russe et turc ont été introduits afin de présenter plus largement les fondements du colloque. Ces idiomes ne sont cependant pas admis comme langue de communication au colloque.

ASI et les thèmes privilégiés

- Concepts fondamentaux en ASI : modèles statistiques, types de variables, variables principales et supplémentaires;
- Avancées nouvelles en cours, diverses métaphores de l'ASI, 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, l'ASI, logique paracohérente
- 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
- Spécificités de la formation à l'ASI : usage du logiciel CHIC, interprétation des représentations graphiques (graphe implicatif ; arbre de la hiérarchie cohésitive)
- 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.;
- 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.
- Intérêt de l'ASI pour l'évaluation d'épreuves internationales ;
- Problématiques de la didactique de l'ASI.

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

15 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

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. 9

ASI et les Défis 2021

- Défi 1: Cône implicatif : comment qualifier et quantifier les qualités globales des variables père, d'une part, et des variables fils par rapport au sommet du cône implicatif. Identifier les liaisons les plus cohérentes via le sommet du cône.
- Défi 2: On dispose d'un réseau d'arcs d'un graphe implicatif d'origine A. Ce graphe présente un caractère dynamique dont les arcs sont pondérés par des instances satisfaisant les règles. Il serait peut-être possible de construire une métaphore mécanique illustrant un tel graphe.
- Défi 3: Enrichir l'extension aux variables continues par des exemples authentiques ; les traiter et les analyser
- Défi 4: Faire une double analyse d'un fichier de données binaires ; d'une part avec l'analyse implicative et, d'autre part, en utilisant une méthode bayésienne.
- Défi 5: Recherche et traitement du degré d'homogénéité/hétérogénéité interne d'une population d'ensemble présentant un ordonnancement général des données compatible avec une structure implicative/cohésitive particulière
- Défi 6: Une variable a étant donnée et des conséquences b, c, d…. étant connues par a=>b, a=>c, a=>d, …est-il possible de définir une implication de a sur la conjonction de b, c, d…c'est-à-dire : a => (b et c et d et..) ?
- Défi 7: Établir en quoi la logique sous-jacente à l'ASI, la logique statistique implicative LSI, où l'on gère les contradictions à l'abri d'une certaine dialectique, est une logique paracohérente (ou paraconsistantes)
- Défi 8: La hiérarchie cohésitive semble être une métaphore du développement cognitif de l'homme. Ne serait-ce pas également celle de l'évolution au sens darwinien ?
- Défi 9: Définir pour une analyse donnée la notion de densité de l'ensemble des relations implicatives (règles). L'étudier en fonction du seuil retenu (ex. 0.95, 0.8., …) et qualifier la compacité d'un graphe implicatif par un rapport entre le nombre de règles représentées et le seuil. Cette étude peut faire penser à la notion de dimension fractale d'une courbe.

Délibérément comme pour les précédents colloques, nous avons tenté dans la mesure du possible de maintenir un programme de A.S.I. 11 articulé autour d'une alternance équilibrée entre des communications portant sur des approches théoriques ou sur des mises en application, et des travaux pratiques sur les logiciels CHIC et R-CHIC

A.S.I. – Analyse statistique implicative :

Le présent ouvrage donne à voir une partie de l'activité de production scientifique suscitée par l'appel à communication du 11ème colloque A.S.I.11. Les contributions constituent les chapitres successifs de ce livre. Rappelons que, cette fois, les auteur-es sont originaires d'Argentine, Brésil, Chypre, Cuba, Espagne, France, Madagascar et Turquie

CONFÉRENCE INVITÉE :

INTERPRÉTABILITÉ DES MODÈLES PRÉDICTIFS

Gilbert Saporta[1]

Les succès du Machine Learning dans les problèmes de régression ou de classification supervisée ont conduit de nombreux auteurs à théoriser l'opposition entre comprendre et prédire, tant sur le plan conceptuel que pratique. Leo Breiman (2001) opposait ainsi ces deux cultures. Vladimir Vapnik (2006) renchérissait en montrant que « Better models are sometimes obtained by deliberately avoiding to reproduce the true mechanisms » . Ce fut le triomphe des boîtes noires : peu importait que le modèle, ou plutôt l'algorithme, soit compréhensible du moment qu'il prédisait bien. La disponibilité de gigantesques bases de données, (Big Data), conduisit un peu vite Chris Anderson (2008) à prédire la fin de la démarche scientifique. L'application à des décisions concernant la vie des individus finit par susciter des polémiques (Cathy O'Neil, 2016). Des codes d'éthique et des règlements, tel le RGPD européen (règlement général sur la protection des données, 2016) consacrèrent le droit à l'explication. Dans cet exposé, on reviendra tout d'abord sur le dilemme entre prédire et comprendre. On distinguera ensuite les modèles explicables des modèles interprétables. Ces derniers incluent les modèles causaux, ce qui permettra de faire le lien avec l'ASI. L'explicabilité des algorithmes (X-AI explainable artificial intelligence) est un domaine de recherche actif avec des approches globales ou locales, agnostiques ou spécifiques, pour mesurer l'importance des variables, utilisant souvent des modèles de substitution (Christoph Molnar, 2021). Au-delà de la transparence nécessaire, le traitement équitable (fairness) ou non-discriminant par des algorithmes prédictifs est devenu une exigence éthique qui donne lieu à une abondante littérature non exempte de paradoxes sur les mesures d'équité (Mitchell et al., 2021). Ce que l'on nomme biais des algorithmes n'est la plupart du temps que la conséquence des biais des données exploitées. On se demandera en conclusion s'il faut ou non renoncer aux boîtes noires.

Références

[1] Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7), 16-07

[2] Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199-231

[3] Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, 8 (141-163)

[1] Professeur émérite de statistique appliquée au Conservatoire National des Arts et Métiers (CNAM). Membre permanent de l'équipe Méthodes Statistiques de Data Mining et Apprentissage du Centre d'études et de recherche en informatique et communications. Professeur invité depuis 2017 à Beihang University à Pékin en Chine. Président d'honneur de la Société Française de Statistique [SFDS]) depuis 2017. Président de la Fondation d'utilité publique "La Science Statistique" depuis 2008. Depuis 2003, Membre d'honneur de la Societatea de Probabilităţi şi Statistică din România. Il a été président de l'International Association for Statistical Computing [IASC] et Vice-président de l'International Statistical Institute [ISI-IIS] entre 2005 et 2007. Ses travaux portent sur l'analyse des données multidimensionnelles. gilbert.saporta@cnam.fr

[4] Molnar, C. (2021). Interpretable machine learning , *A Guide for Making Black Box Models Explainable*, https://christophm.github.io/interpretable-ml-book.

[5] O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishers.

[6] Vapnik, V. (2006) *Estimation of Dependences Based on Empirical Data*, 2nd edition, New York: Springer

LA HIERARCHIE COHESITIVE : AVATAR FRACTAL DES THEORIES DU DEVELOPPEMENT

Régis Gras[1], Sofia Anastasiadou[2]

RÉSUMÉ

Nous avons rencontré plusieurs fois dans nos recherches autour de l'Analyse Statistique Implicative, des situations conduisant à des ensembles de données croisant sujets et variables, dont le traitement se traduisait par un certain ordre au sein des règles comportementales des sujets. L'outil qui nous paraissait le mieux adapté à leur traduction sagittale était le graphe implicatif car il était orienté. Mais celui-ci ne permettant pas de décrire la structure ordinale de l'ensemble des règles, nous lui avons associé, avec succès, semble-t-il, un autre graphe, la hiérarchie cohésitive, également orientée. On rappellera ici qu'elle apporte, spécifiquement et de façon dynamique, un ordonnancement des familles de règles, déjà sous-jacent au graphe implicatif. Cette procédure a été appliquée à plusieurs concepts et situations relevant des sciences humaines, de la psychologie, la sociologie, mais aussi relevant de la biologie, en particulier l'ontogenèse et la phylogenèse. Il est souvent apparu, à l'analyse de certaines hiérarchies cohésitives, une propriété inhérente à l'ensemble des sujets, à savoir, caché ou non, un certain développement anthropologique. Des exemples l'illustreront dans ce chapitre. Un contre-exemple montrera en revanche qu'une hiérarchie cohésitive peut ne pas traduire un développement. Cependant, dans les cas favorables, relativement à la structure du développement, soulignée par les images graphiques, on montrera qu'il s'agit d'un objet de la géométrie fractale et non pas d'un objet euclidien. Fractalité et développement représenteront les axes coordonnés de cet article et pour cela seront discutés à travers plusieurs exemples.

***Mots-clés** : graphe implicatif, arbre cohésitif, taxonomie, cognition, fractalité,*

ABSTRACT

Several times in our research around Implicative Statistical Analysis, we have encountered situations leading to sets of data crossing subjects and variables, the processing of which resulted in a certain order within the behavioral rules of the subjects. The tool that seemed to us best suited to their sagittal translation was the implicative graph because it was oriented. But since this does not allow us to describe the ordinal structure of the set of rules, we have associated with it, apparently successfully, another graph, the cohesive hierarchy, also oriented. It will be recalled here that it provides, specifically and dynamically, an ordering of families of rules, already underlying the implicative graph. This procedure has been applied to several concepts and situations relating to the human sciences, psychology, sociology, but also relating to biology, in particular ontogeny and phylogeny. When analyzing certain cohesive hierarchies, it has often appeared a property inherent in all subjects, namely, hidden or not, a certain anthropological development. Examples will illustrate this in this chapter. A counterexample, on the other hand, will show that a cohesive hierarchy may not reflect a development. However, in favorable cases, relative to the structure of the development, underlined by the graphic images, it will be shown that it is an object of fractal geometry and not a Euclidean object. Fractality and development will represent the coordinated axes of this article and for this will be discussed through several examples.

***Keywords**: implicative graph, cohesive tree, taxonomy, cognition, fractality.*

1 Introduction

La notion de **développement** est à la fois *plurielle* - présentant plusieurs facettes - et *floue*, contenant l'idée de *progrès* probable mais incertain ou tout au moins l'idée d'une évolution en

1 Ecole Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire LS2N, UMR 6241, Site de la Chantrerie, rue C. Pauc, BP 44306, Nantes cedex 3, e-mail : regis.gras@univ-nantes.fr, regispor@hotmail.fr.

2 University of Western Macedonia, sofan@uom.edu.gr

cours, voire d'une direction orientant des étapes logiquement orientées. Par exemple, en psychologie, une théorie bien connue du développement cognitif de l'enfant a été établie par Piaget et se présente sous la forme dite de 4 stades successifs. De même, G. Bachelard (1972) stratifie la formation de l'esprit scientifique selon trois stades. En biologie, l'**ontogenèse** (ou ontogénie) décrit le développement progressif d'un organisme depuis sa conception jusqu'à sa forme mûre, voire jusqu'à sa mort. Autre exemple relevant du domaine de la paléontologie, celui de la phylogenèse c'est -à-dire de l'évolution d'une espèce ou d'un groupe d'espèces et des relations de parenté entre les organismes vivants. Nous nous permettons d'étendre cette notion de développement à tout changement continu ou discret d'un phénomène dont il est possible de repérer et mesurer les scansions.

Afin d'enrichir l'intuition et de faciliter l'analyse et l'interprétation des phénomènes observés, il est bon de représenter la structure de l'ensemble des différentes étapes décrivant le film du développement. Pour cela, compte tenu de sa morphologie, faite de tortuosités, de césures, de brisures et pour répondre à un tel objectif, la géométrie fractale semble être un outil adapté. Car, grâce à la notion de dimension non entière que l'on peut lui affecter, elle permet de décrire à différents niveaux d'échelle des formes plus complexes que celles de la géométrie euclidienne (ligne droite, cercle...). Nous verrons que le graphe implicatif et la hiérarchie cohésitive présentent les propriétés que lui conférerait la géométrie fractale.

À la suite de lectures diverses et d'incursions de recherche dans le domaine de la paléontologie, il apparaîtrait que le développement de l'arbre de vie, la phylogenèse et l'ontogenèse, en particulier le développement cognitif, respectent la théorie des **équilibres ponctués**, concept que nous empruntons à la paléoanthropologie[3]. Celle-ci puise ses objets, son origine et ses fonctions au sein de la théorie de l'évolution. Mais elle ne serait qu'une facette d'un système plus général englobant l'univers, la vie et les sociétés. Ces trois domaines partageraient des propriétés communes que citent L. Nottale, J. Chaline et P. Grou (2000, p.10)], entre autres :

- seraient sujets à des ruptures,
- seraient structurés en niveaux d'organisation selon la géométrie fractale,
- seraient soumis à des phénomènes d'accélération débouchant par des sauts, sur des ruptures via des crises
- répondraient à une prédictibilité de structures non déterministes et plus collectives qu'individuelles.

Cette théorie des **équilibres ponctués** questionne la théorie darwinienne classique, par ses caractères dynamique, discontinu, chaotique et non linéaire que cette théorie, mais aussi l'ASI - *c'est notre conjecture* - semblent partager. Non seulement, l'ASI paraît révéler l'existence d'un développement, mais en plus paraît l'analyser, en restituer la genèse, orientée et ses étapes. Notons d'ores et déjà que la disjonction des classes de la hiérarchie, l'absence de linéarité de celle-ci, paraissent parfaitement compatibles avec le modèle des équilibres ponctués en faisant la place aux mutations génératrices du passage d'une classe cohésitive à une autre sans liaison mutuelle. Cette propriété ne se rencontre pas, en revanche, avec les hiérarchies de similarité classique.

Précisons notre **conjecture** :

---

3 « La tendance en paléoanthropologie au cours des deux dernières décennies du XXème siècle était aux équilibres ponctués…L'idée était que les transitions évolutives se font très rapidement par sauts. » (P. Picq, *Sapiens face à Sapiens*, (2019)). Dans ce cadre, les espèces se maintiennent pendant des millions d'années (la stase), dans leur milieu naturel. Ensuite une ou plusieurs espèces se différentient et sont remplacées par une ou plusieurs autres espèces sans intermédiaire transitoire, dérivées par une ponctuation, un saut évolutif, (D. Queiros-Condé, J. Chaline et J. Dubois). (2014).

Et si le modèle des équilibres ponctués se réalisait dans les phénomènes de développement cognitif, comme dans l'arbre de vie, avec sa sensibilité à l'environnement, ses phases de stases d'assimilation et ses sauts de discontinuité conduisant d'un niveau à un autre plus général, plus complexe, plus fonctionnel ? Comment ne pas penser, tout de go, au développement cognitif, constitué de stades successifs, de points d'arrêt, de branches différentielles, décrit par J. Piaget ? C'est aussi ce que suggère Jeanne Guillet-Silvain qui écrit (communication personnelle, 2020) : « *Je suggère de penser au développement cognitif, constitué de stades successifs, de points d'arrêt, de branches différentielles, décrits par J. Piaget* ».

Nous nous proposons donc, pour conforter notre hypothèse de relation étroite entre l'ASI et la notion de développement, de nous appuyer sur 4 situations concrètes et des données qu'elles produisent. Dans les 4 situations seront traitées des variables entre lesquelles préexistent des relations de préordre globales ou locales du type a < b pour signifier, suivant les cas, que a précède b, ou que a évolue en b, ou que a est préféré à b, etc. On verra que ce préordre se traduit dans les graphes de l'ASI en un autre préordre, avatar du développement. D'autre part, cela nous conduira à montrer que les représentations graphiques de l'ASI sont des objets de la géométrie fractale

2 Taxonomie cognitive et graphe implicatif

Une grande partie de ce qui suit a été présentée lors du Colloque ASI 9 (Belfort 2017) dans (Gras, 2017). Son objectif majeur était de montrer que le graphe implicatif relevait de la géométrie fractale et, de ce fait, conférait au graphe, représentation du tableau des implications entre variables, des propriétés imagées et intuitives de l'ASI.

2.1 Taxonomie et implication

À la suite d'un traitement par CHIC des données fournies dans la thèse de Régis Gras (Gras, 1979) sur la base d'un test « **taxonomique** » (411 élèves, 42 variables binaires, correspondant à 42 exercices, voir aussi Gras et Bodin (2017), nous obtenons ci-dessous des graphes implicatifs aux seuils respectifs 0.95 et 0.85.

Rappelons que la **taxonomie** (Annexe 1 et un condensé dans le Tableau 1 ci-dessous)**,** support du test**,** élaborée par R. Gras (1977), puis remodelée par A. Bodin, et appliquée au champ des mathématiques, décompose puis tente d'organiser ce que contiennent, dans l'esprit, les stades dits du développement de l'enfant définis par Piaget.

Niveau A : Connaissance des outils de préhension de l'objet et du fait mathématique	*associer, assembler, simuler, observer, déchiffrer, décrire, organiser, calculer*
Niveau B : Analyse des faits et transposition	*abstraire, prolonger, induire, analyser, comparer, schématiser, traduire, transposer*
Niveau C : Compréhension des relations et des structures	*reconnaître, construire, justifier, déduire, analyser, abstraire, appliquer, interpoler*
Niveau D : Synthèse, et créativité	*Organiser, calculer, optimiser, illustrer, démontrer, valider, créer, inventer, généraliser, induire, prévoir, extrapoler*
Niveau E : Critique et évaluation	*contrôler, optimiser, prévoir, critiquer, questionner, vérifier, tolérer*

Tableau 1 : Taxonomie d'objectifs cognitifs (condensée)

Elle est constituée de 5 grandes classes A, B, C, D et E, ordonnées selon la complexité supposée par le didacticien-mathématicien, c'est-à-dire sur la base du sujet épistémique, donc des structures logico-mathématiques. Elle spécifie, en y répondant assez fidèlement, les stades piagétiens des opérations concrètes et des opérations formelles qui en constituent la théorie du

développement de Piaget (1970). Elle est sensiblement comparable à la taxonomie de Bloom (1969) appliquée aux acquisitions dans le domaine des sciences humaines. La différence ne s'arrête pas là : en effet la taxonomie de Bloom était le résultat d'un consensus de nature psychologique et pédagogique. Celle de Piaget répondait à sa vision logico-mathématique, d'une épistémologie génétique, tandis que celle que nous utilisons ici procède dès le départ d'une réflexion didactique et s'est immédiatement accompagnée de validations empiriques.

Les niveaux d'acquisition d'un concept, éléments de ponctuation du développement, y sont ainsi supposés ordonnés par la complexité croissant du niveau ou classe A vers le niveau ou classe E. Cette complexité, en l'occurrence, se définit sur la base de la plus ou moins grande mobilisation des outils de pensée et de leur intrication pour résoudre un problème donné. Schématiquement dire qu'un problème de la classe C est plus complexe qu'un problème de la classe A sous-entend que des compétences nouvelles ou complémentaires sont nécessaires en C par rapport à celles que l'on possède en A. On sait d'ailleurs que généralement le franchissement d'un niveau se produit comme fruit d'une rupture comme le franchissement d'un obstacle épistémologique. Gérard Vergnaud, à travers sa théorie des champs conceptuels, y apportera sa propre nuance en s'exprimant dans (2007) ainsi : « *Deux problèmes qui ont la même structure, c'est-à-dire qui font appel aux mêmes opérations, mais qui s'appliquent à deux domaines d'expériences différents, pourront représenter des difficultés extrêmement différentes.* » (p.81). Ce qui ne contredit pas Piaget qui parle de stades au sein des développements des contenus de connaissances, mais exprime une réserve à sa théorie.

Le test qui a été présenté à 400 élèves de 13 à 14 ans est constitué de 42 exercices censés illustrer chacun une classe ou une des fractions de la taxonomie. C'est ainsi que nous trouvons ; 9 items de la classe A, 7 de la classe B, 9 de la classe C, 12 de la classe D, 5 de la classe E. Autrement dit, on peut donc attendre de ce test que les réussites des élèves à des items C impliquent généralement les réussites aux items A. On peut donc encore légitimement penser que la transition de A vers E soit un signe de développement considérable des compétences cognitives d'un sujet.

Les items constituant le test sont notés de la façon suivante par deux indices : le 1er indice littéral illustre la classe taxonomique X, le 2ème indice numérique illustre la sous-classe de cette classe X, le 3ème indice correspond à la place arbitraire de cet item parmi les items relatifs à la classe X. Ainsi l'item C21 illustre la classe C, la sous-classe 2 et sa place 1 parmi les items de X.

2.2 Traduction graphique de l'ensemble des règles

L'ensemble des items est structuré en un graphe orienté de haut en bas par une fréquence de réussite croissante à la question posée dans l'item considéré. Par exemple, A11 est mieux réussi (68%) que D16 (24%) et que C10 (51%). Par conséquent si l'ordre de réussite devait respecter l'ordre taxonomique de complexité a priori, le long de chaque chemin, les items se succéderaient de son sommet à son extrémité selon un ordre croissant de réussite des items relevant de la classe E à ceux de la classe A. Le graphe implicatif, obtenu par CHIC, respecte globalement cet attendu comme on peut le constater.

Sur chacune des figures, rappelons que les codes représentent à gauche la classe et sous-classe taxonomiques, et à droite la place de l'item proposé dans le test relativement à cette classe.et sa sous-classe. L'arc ou la règle i⇒j signifie que, sous une certaine intensité d'implication, la réussite à l'item i est accompagnée de la réussite à l'item j, sans que la réciproque ne soit nécessairement vraie. Mais n'apparaissent pas les différentes boucles issues de chaque variable et correspondant à l'intensité d'implication réflexive. Si toutes les variables étaient représentées, on observerait 40 boucles.

Dans la figure 1, on note la présence de 3 arcs dont une boucle, correspondant aux 3 règles : A13⇔A14 et B26 ⇔B27. En filigrane, on pourrait observer la présence des 40 variables, dont la plupart ne sont pas liées aux autres sauf à elles-mêmes.

Dans la figure 2, apparaissent les flèches pointillées indiquant la satisfaction de la transitivité pour laquelle, au seuil donné, si a⇒b et b⇒c alors aussi a⇒c.

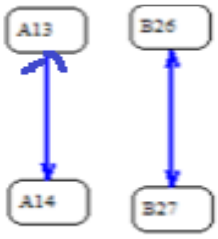


Figure 1 : Graphe au seuil 0.98

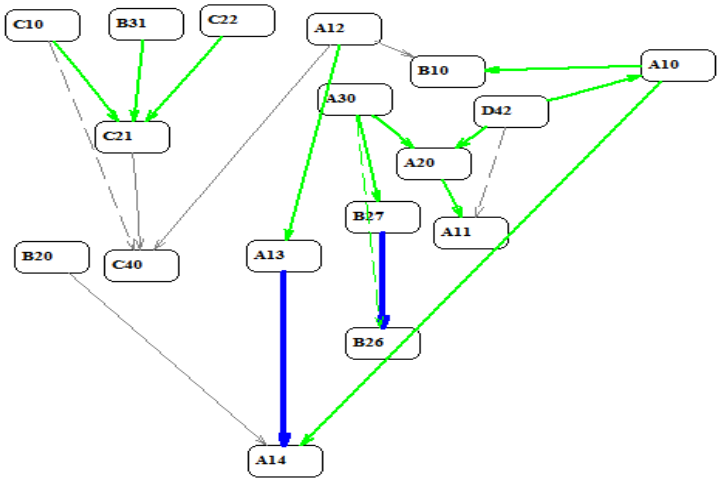


Figure 2 : Graphe au seuil 0.85

2.3 Le graphe implicatif souligne-t-il un développement cognitif du sujet ?

Si nous nous référons à la taxonomie, nous notons, à l'opposé de l'attendu, l'absence de respect total de sa propriété pré-ordinale. En effet, on pouvait s'attendre à sa validation par l'observation de règles vérifiant sauf en de rares exceptions : si E(x) alors D(y), si D(y) alors C(z), si C(z) alors B(t) et si B(t) alors A(u). Or de nombreuses règles ne satisfont pas ce schéma implicatif. Une des raisons tient d'ailleurs, le plus souvent, au contenu même de l'habillage de l'item, à la familiarité de son contenu, à la parcellisation cognitive des items et non pas à sa référence taxonomique donc développementale. Il nous semble, en accord avec G. Vergnaud (1996), qu'il faut chercher les explications des distorsions entre l'attendu taxonomique et la réalité. Pour G. Vergnaud, rappelons-le, un champ conceptuel se définit d'une part, par un ensemble de situations dont le traitement implique les mêmes opérations, et d'autre part, par l'ensemble des concepts et théorèmes (par ex. en acte) qui permettent d'analyser ces situations comme des tâches mathématiques. Il intègre dans son cadre théorique : *perception et action, représentations et processus mentaux et langage*. C'est peut-être plus par ce regard pragmatique qu'il faut expliquer les écarts observés dans la structure des graphes. Par exemple, A12⇒C40 ou B31⇒C21. D'ailleurs, si le test porte sur la notion de symétrie centrale, les items se placent le plus souvent dans un cadre géométrique, mais aussi, pour de rares questions, dans le cadre algébrique de la moyenne de deux nombres.

Finalement, le graphe implicatif ne peut être, en la circonstance de validation de la complexité cognitive, un outil parfaitement adapté. On verra, plus loin, que la hiérarchie cohésitive, en revanche, en possède la propriété.

2.4 Dimension fractale du graphe implicatif

Si l'on s'intéresse à la structure de l'ensemble des règles entre les variables et sa représentation graphique, deux types d'échelles interviennent :

- Celle qui conduit à ne représenter que des règles obtenues par une intensité d'implication minimale donnée,
- Celle qui conduit à l'image elle-même pour une certaine taille imposée sur la feuille.

C'est, bien entendu, la première qui nous intéresse ici, la structure de l'ensemble des règles étant manifestement inchangée quand on fait varier l'échelle de l'image.

Faisant appel à l'intuition du lecteur, il peut remarquer que le graphe de gauche[4] est susceptible d'être placé en surimpression par rapport au graphe de droite (phénomène lié au caractère fractal du graphe). Répéter ce changement d'échelle conduit à recueillir des informations nouvelles, obtenir du *nouveau* par ce changement de *résolution* que l'on peut même produire de façon **continue** par un effet **zoom**. On peut imaginer une sorte de « mille feuilles » ou d'empilements ou d'emboitements de type « poupées russes » obtenus en faisant varier l'indice du changement d'échelle, l'intensité d'implication.

Notons N_α le nombre d'arcs apparaissant lorsque l'échelle donnée par l'intensité d'implication est α.. Pour valider la fractalité du graphe implicatif, nous devons calculer le rapport $D_\alpha = \frac{logN}{log\frac{2}{\alpha}}$ pour une échelle α et en observer l'invariance. Nous renvoyons à l'article précité et intitulé, *Dimension fractale du graphe implicatif* , dans lequel un tableau récapitule les calculs dont nous reverrons la méthodologie dans section 3.2.

Quel que soit le seuil d'implication ($\varphi(a,b)$) on trouve une valeur $\mathbf{D_\alpha}$ qui varie sensiblement autour du nombre 5. Les écarts des valeurs observées au voisinage de 5 sont faibles par rapport à ce que donnerait, par exemple, une relation linéaire entre N_α et $\varphi(a,b)$. Ce qui justifie la nature présumée **fractale** du graphe implicatif. On adopte donc **D** =5 comme dimension du graphe du corpus. Ce nombre est donc considéré comme un **invariant du graphe implicatif**. Il rend compte à la fois de la multitude des arcs et chemins du graphe, de leur intrication et donc du degré important de la brisure du graphe.

Outre la propriété géométrique descriptive du graphe implicatif –structure fractale-, cette propriété d'invariance d'échelle est opératoire. Elle permet d'estimer le nombre de règles qui serait attendu, par exemple, pour $\varphi(a,b) \geq 0.87$. En effet, log N_α devrait être de l'ordre de $5\log(2/0.87)=1.807$ d'où $N_\alpha \cong 64$. Or nous observons à partir du tableau des intensités d'implication figurant en annexe $N_{obs} = 70$, ce qui est une bonne estimation du nombre de règles attendu et conforte le caractère étonnamment opératoire de la dimension.

Dans la situation où l'on dispose d'un ensemble V de variables préordonnées par une taxonomie, représenté par un graphe orienté, sans cycle, pondéré par les intensités d'implication non strictement transitives entre variables, attribuant une relation de préordre sur VxV, alors l'hypothèse suivante est vérifiée : **Le graphe implicatif est un objet de la géométrie fractale.(H).**

En effet, on a vu à partir de deux exemples, que la propriété fractale du graphe implicatif était observée et légitimait l'hypothèse de la généralisation de cette propriété. Celle-ci offrait au graphe une structure géométrique dont la dimension fractale, un invariant du graphe, n'était pas plane, donc non égale à 2 dimension euclidienne, et, en même temps, conduisait à une autre

4 B26 et B27 sont interchangeables. C'est un caprice du logiciel qui les a permutées.

représentation mentale, intuitive et opératoire, d'un graphe dépouillé de son habit linéaire et symétrique.

3 Taxonomie cognitive et hiérarchie cohésitive

Nous reprenons ci-dessous, à partir de la même situation que celle du paragraphe 1, les mêmes données. Nous nous proposons de les traiter avec CHIC selon la modalité « Arbre cohésitif ». Cette modalité calcule les règles de règles (règles généralisées ou métarègles) et en présente une image graphique, ici une hiérarchie. Nous comparerons les informations recueillies respectivement selon les deux modalités de traitement graphique. Car, dans le cas cohésitif, ce ne sont plus les comportements illustrés par une règle, comme le serait un schème au stade inférieur mais plutôt une synthèse, par exemple une conception, elle-même signifiée par une règle de règles ou un ensemble de telles métarègles. À ce titre où perce la structure, non factuelle, s'est alors posée la question de la fractalité de cette hiérarchie relativement à la représentation des données du test (Figures 3 et 4) et, en particulier, du développement cognitif des élèves, à leur conceptualisation, écrirait G. Vergnaud.

Nous proposons ci-dessous deux représentions hiérarchiques de l'ensemble des 42 items : la première traite l'ensemble de tous les items ; la deuxième se limite aux items les plus fréquemment observés, certainement les plus pertinents dans la population, soit seulement 16 variables.

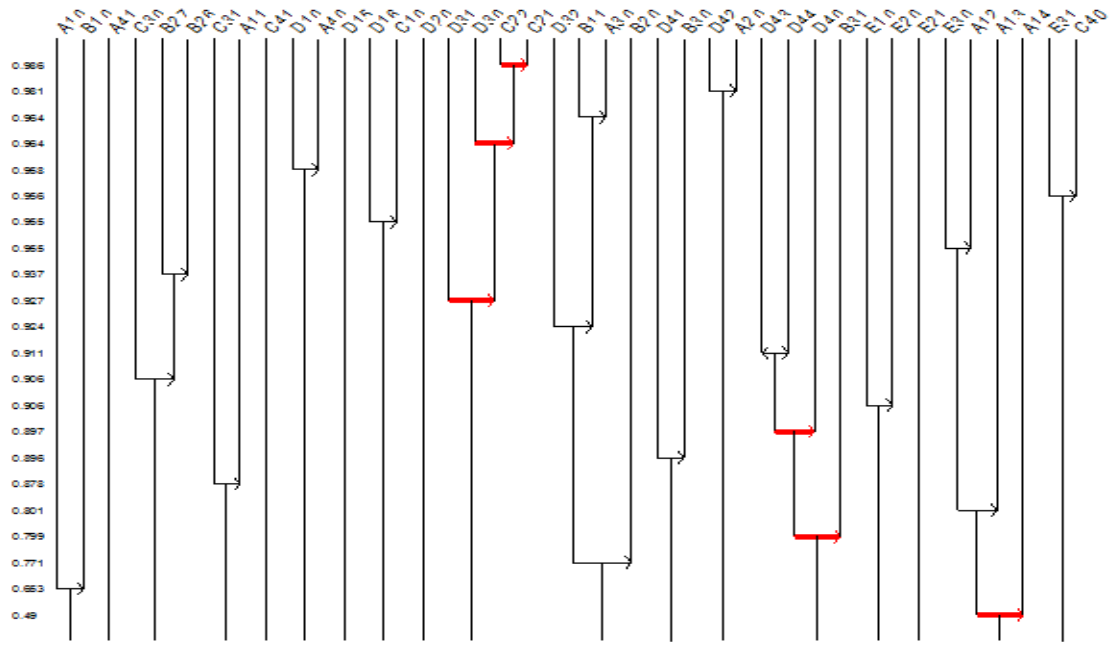


Figure 3 :Hiérarchie cohésitive du test taxonomique 42 variables

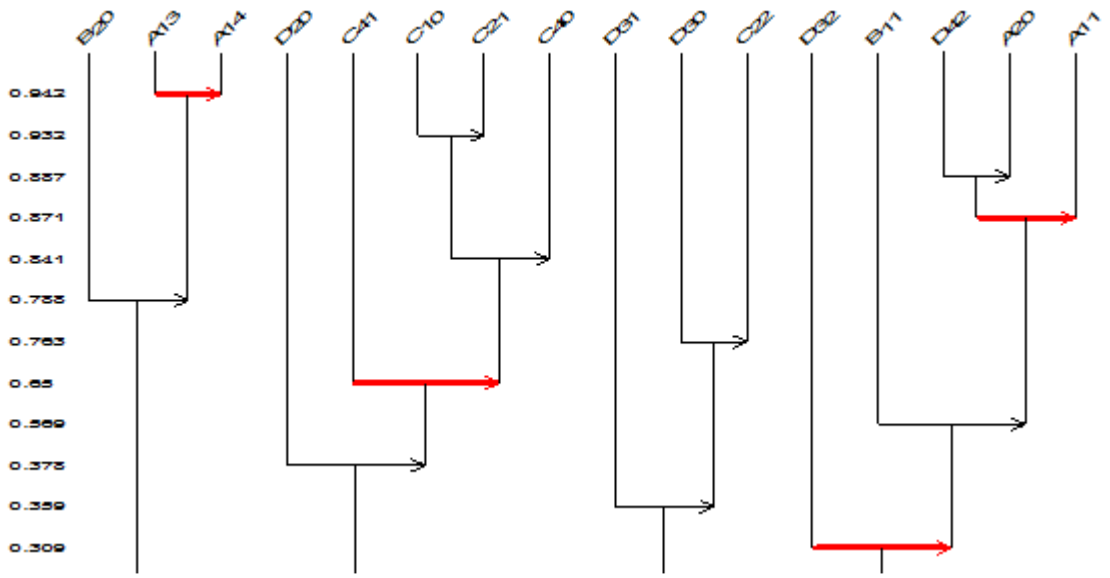


Figure 4 :Arbre cohésitif du fichier de données limité à 16 variables

3.1 La hiérarchie cohésitive souligne-t-elle un développement cognitif du sujet ?

Dans la communication que nous avons présentée à ASI 10, « *Arbre cohésitif et métaphore fractale* » et dans (Gras, 2018), nous avons exploité le caractère orienté de la hiérarchie pour traduire :

- d'une part, la consistance croissante des niveaux d'attente des enseignants de mathématiques vis-à-vis des concepts à enseigner (voir section 4),
- d'autre part, l'hypothèse taxonomique de niveaux d'acquisition cognitive d'un concept mathématique enseigné.

Cette fois, la hiérarchie taxonomique des règles, particulièrement soulignée dans les niveaux significatifs[5] (en rouge sur le graphe) est respectée à une exception près. Par exemple, D(synthèse et créativité)⇒B(analyse de faits et transposition), D⇒C(compréhension des relations et des structures), D⇒A(connaissance des outils de préhension de l' objet et du fait mathématique), (D⇒B)⇒A.

Si l'on fait rentrer dans la même classe taxonomique X tous les items en relevant, sur les 277 implications observées, 235 sont conformes à la taxonomie, soit dans 85% des cas. Les cas non conformes se limitent presque toujours à un échange entre deux classes voisines, par exemple $B_i \Rightarrow C_{j..}$ La permutation E⇒D⇒C⇒B⇒A est vérifiée dans 42% des cas. Les 719 autres permutations se partagent le restant. Par ex., la permutation E⇒C⇒D⇒B⇒A n'est observée que dans 12% des cas. Si nous observons la hiérarchie cohésitive nous retrouvons, à une exception près[6], un respect complet de la taxonomie.

---

5 « Les chemins significatifs de la structure hiérarchique obtenue semblent montrer le fonctionnement interne du savoir, autrement dit un système structuré de la conduite » nous rappelle (communication personnelle) J. Guillet-Silvain dans sa note de synthèse pour son HDR (2020)

6 Une classe de la hiérarchie ne vérifie pas strictement la croissance de niveau cognitif : B11⇒(D42⇒(A20⇒A11)) simplement dû à une mauvaise identification taxonomique de B11 et à un changement de champ conceptuel. Si nous supprimons cette aberration, la classe apparaît parfaitement ordonnée taxonomiquement parlant. De toute façon, nous abritant derrière la sagesse scientifique d'E. Morin (2020), *« il nous faut renoncer à la réduction de la connaissance et de l'action au calcul et il nous faut répudier la raison glacée obéissant inconditionnellement à la logique du tier exclu »*. C'est aussi ce que nous présentons dans (Gras, 2018) sur la logique paracohérente de l'ASI.

Ainsi, les niveaux d'acquisition sont ordonnés par la complexité au sein d'un champ conceptuel. Dans toute classe cohésitive, on observe la croissance du niveau cognitif de l'élève ce qui conforte ici notre conjecture résumée ainsi : *la hiérarchie cohésitive constitue un avatar du développement cognitif.* Ce phénomène, d'une grande importance pour l'ASI, est une conséquence du choix épistémologique fait en analyse cohésitive : c'est l'empilement non symétrique des règles en règles de règles qui importe dans la classification.

3.1.1 Quelques exemples

- D31⇒(D30⇒C22) càd « *compétence à généralise*r » implique « *compétence à justifier par un argument recevable* », soit une amplification ;
- B20⇒(A13⇒A14) càd « *reconnaissance et usage d'une relation implicite* » implique « *reconnaissance de la terminologie* », soit un dépassement ;
- D20⇒((C41⇒((C10⇒C21)⇒C40)) càd « *construction d'une démonstration* » implique « *compréhension d'une raisonnement* » qui implique « *applications à des situations familières* » ;

Ainsi, un raisonnement est l'outil rationnel et **extensionnel** qui ***amplifie*** le champ des propriétés. Dans chacun de ces exemples, on note bien l'élévation du niveau de complexité cognitive. Cette même remarque serait vérifiable dans l'ensemble des métarègles. On peut donc fortement conjecturer que la **hiérarchie cohésitive est un sérieux avatar d'un développement cognitif.** En d'autres termes, appartenant au champ de la cognition, qu'emploierait peut-être Gérard Vergnaud (2007), la hiérarchie cohésitive serait un des **signifian**ts du **signifié qui, à son tour, serait** le développement cognitif.

3.1.2 Remarque

Comparons les deux images obtenues par CHIC du traitement des données du test, limitées aux 16 variables retenues selon l'analyse cohésitive ci-dessus et l'analyse des similarités selon l'algorithme de la vraisemblance du lien de I .-C. Lerman (1981).

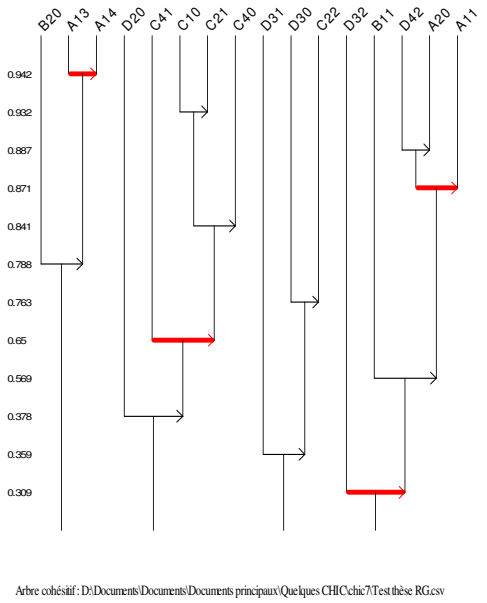


Figure 5 :Arbre cohésitif limité à 16 variables

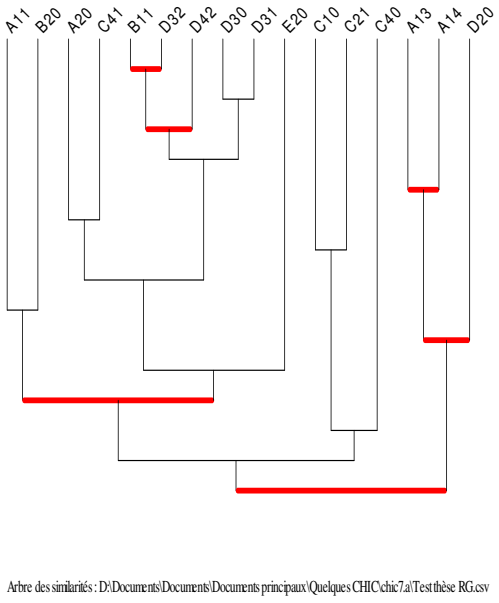


Figure 6 :Arbre des similarités limité à 16 variables

On peut constater que les chaînes cohésitives, signes du développement cognitif, sont rompues sur la figure 5. Les classes semblables terminales, tout comme avec le graphe implicatif, n'ont plus le sens de niveau supérieur atteint comme nous l'observons sur la figure 6. La similarité ne peut être indicatrice de développement.

3.2 Dimension fractale de la hiérarchie cohésitive

Nous allons montrer, ci-après que la hiérarchie cohésitive possède une dimension fractale, propriété que l'on peut comparer à celle de la relativité d'échelle. En effet, à l'instar de ce que nous verrons avec l'évolution darwinienne, nous pouvons faire l'hypothèse d'une autosimilarité des données. Elle est légitimée par la stabilité de sa forme à différentes échelles et d'une loi log-périodique sous-jacente associant le logarithme du nombre N de classes et le niveau de la hiérarchie. Notons que la disjonction des classes de la hiérarchie, l'absence de linéarité de celle-ci sauf au sein de chaque classe, sont parfaitement compatibles avec le modèle des équilibres ponctués en faisant la place aux mutations génératrices du passage d'une classe cohésitive à une autre sans liaison mutuelle comme, en revanche, on le trouverait avec les hiérarchies des similarités classiques (voir ci-dessus).

Rappelons quelle est notre démarche pour parvenir à la mise en évidence de la structure fractale de l'arbre cohésitif présent. À un niveau de formation de celui-ci, en fait relativement à une échelle donnée, correspond une cohésion C qui en donne une valeur relative. Nous savons, comme son nom l'indique, qu'il s'agit d'un indice valorisant la consistance d'un échelon, d'un niveau et la lente régression de cet indice au fur et à mesure de l'extension des classes. Pour ce faire, on dénombre les classes déjà formées, avant ce niveau, en utilisant la représentation hiérarchique donnée par CHIC ainsi que la procédure suivante que nous inspire la formule donnant la cohésion C, tout en respectant sa sémantique :

- à une classe (a,b) formée à ce niveau on ajoutera 1 au nombre de classes formées jusqu'alors ;
- à une classe nouvellement formée de type (a, (b,c)) ou ((a,b),c) on ajoutera 2 au nombre de classes déjà formées correspondant aux cohésions des classes (a,b) et (a,c) ou (a,c) et (b,c) ;
- à une classe de type ((a,b), (c,d)) on ajoutera 4 correspondant aux cohésions de a avec (c,d) et b avec (c,d) ;.

A un niveau significatif choisi, on compte le nombre N de classes formées à un niveau de cohésion supérieure ou égale à une certaine valeur de cohésion C indiquée sur l'arbre et reportée dans le tableau ci-dessous. Nous savons, dans la pratique de l'ASI relativement à l'arbre qui hiérarchise les classes orientées, que le nombre de classes formées croît quand la cohésion décroît. Elle a alors un comportement comparable à une échelle, à une résolution permettant d'observer la forme la plus large. Aussi reprenant la formule de Mandelbrot (1995, 1997) et Richardson (1961), on obtient la valeur instantanée de l'invariant dimensionnel cherché : $D_C = \frac{logN}{\log\left(\frac{2}{C}\right)}$ (1). Cette détermination est en tout point conforme à celle adoptée pour le graphe implicatif (Gras, Gregori, 2019) Ainsi, D_C reste sensiblement invariant si C devient plus petite alors que N devient plus grand. Ce que nous devrions observer.

Revenons à notre fichier du test taxonomique traité ci-dessus. Tout en nous limitant aux variables les plus significatives en termes d'occurrence, c'est-à-dire l'arbre cohésitif de la figure 4, nous obtenons le résultat suivant validant la propriété fractale. Le graphique 6 présente en abscisse le niveau d'une classe cohésitive (il y en a 12) et en ordonnée le log népérien du nombre N de sous-classes déjà formées à ce niveau.

Niveau de l'arbre	Cohésion C à ce niveau	Nombre N de niveaux pour coh≥C	D_C calculé
2	0.932	2	**0.908**
3	0.887	3	**1.35**
4	0.871	5	**1.94**
5	0.841	6	**2.07**
6	0.788	8	**2.23**
7	0.763	9	**2.28**
8	0.65	12	**2.21**
9	0.569	15	**2.15**
10	0.378	19	**1.77**
11	0.359	21	**1.77**
12	0.309	25	**1.72**

Tableau 2 : Test taxonomique niveau-classes

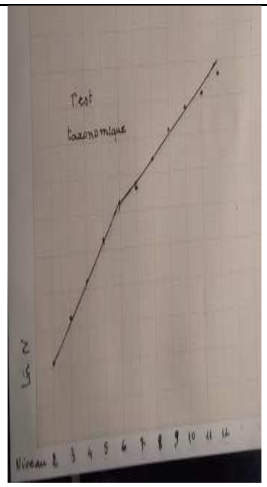


Figure 7 :Test taxonomique
Courbe (N, ln(N))

On observe une quasi-linéarité de l'expression de la loi log-périodique confortant nos hypothèses exprimées plus haut. La moyenne des valeurs prises par D_C est 1.854 et l'écart-type est 0.40. Dans le modèle fractal considéré sous-jacent à l'image de l'arbre cohésitif, on peut considérer, eu égard au faible écart-type, que la dimension fractale de l'arbre est 1.85. Ainsi, la structure géométrique de l'arbre cohésitif semble bien relever de la géométrie fractale au sens de Mandelbrot (1995, 1997). La dimension fractale mesure approximativement la pente de la courbe (N, ln(N)) et donc le degré d'irrégularité de l'arbre, de sa brisure dirait Mandelbrot

Par suite, à travers deux représentations des données taxonomiques, nous confirmons la nature fractale des performances des élèves. Avec la hiérarchie cohésitive et non pas avec l'arbre implicatif, nous obtenons une image des niveaux de conceptualisation préordonnés portant sur la notion de symétrie centrale chez des élèves de 13-14 ans.

On peut faire l'hypothèse que cette propriété serait valable pour tout concept dont l'acquisition se produit selon des étapes du développement cognitif de l'individu.

Une image intuitive de la propriété fractale de l'arbre cohésitif peut se constituer selon un escalier évidé en colimaçon. Le voyez-vous ainsi ?

Figure 8 :Test Escalier en colimaçon

Comment ne pas penser au développement cognitif, constitué de stades successifs, de points d'arrêt, de branches différentielles, décrit par J. Piaget ? Serait-ce ainsi que l'ontogenèse rejoindrait la phylogénèse ? Et si en effet le modèle des équilibres ponctués se réalisait dans les phénomènes de développement cognitif avec sa sensibilité à l'environnement, ses phases de stases assimilatoires et ses sauts de discontinuité conduisant d'un niveau à un autre plus général, plus complexe, plus fonctionnel ?

4 Un questionnaire d'opinions et son arbre cohésitif

Une enquête de l'Association française des Professeurs de Mathématiques de l'Enseignement Public (APMEP, EVAPM1)[7] a été proposée aux professeurs de mathématiques de classes terminales de l'enseignement secondaire dans différentes filières : S (à dominante sciences dures) et ES (à dominante scientifique et sociale), littéraires L et technologique T. Ces variables constitueront des variables supplémentaires de l'analyse, donc n'entrent pas directement dans la constitution des règles. Nous avons recueilli et analysé (Bodin et Gras, 1999) les réponses de 311 professeurs, à des questions portant sur les objectifs (15 ont été retenus) qu'ils assignent à leur enseignement et sur leurs opinions relatives à 11 phrases susceptibles d'être communément énoncées. Des objectifs jugés non pertinents sont ajoutés aux variables supplémentaires « filières ». Nous présentons ci-après le questionnaire dans sa forme textuelle.

QUESTIONNAIRE -Professeurs de Terminale

Q1- Au nom de quelle série répondez-vous ? :…….
(*Si plusieurs séries, utiliser un questionnaire par série*)
Q2-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 ? (*Choisir 6 objectifs et ranger par ordre préférentiel décroissant de 1 à 6*)
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

Les objectifs ci-dessus vous paraissent-ils pertinents (PER): OUI NON (entourez votre choix)
Si non, précisez lesquels en utilisant le codage proposé : .....

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 :

7 EVAPM : http://www.apmep.fr/-Observatoire-EVAPM-

À côté d'EVAPM (Groupe d'enseignants. de Mathématiques travaillant sur l'Evaluation), nous avons développé une base multicritère de questions d'évaluation : la base EVAPMIB. Cette base qui, outre les questions d'EVAPM, contient de nombreuses questions utilisées dans des enquêtes à grande échelle (TIMSS, PISA, …), questions dont on a pu renseigner les conditions de passation et les résultats obtenus, cette base donc contient plus de 2000 questions et son développement se poursuit. Bien sûr le niveau taxonomique est l'un des critères de choix possible des questions.

Mettre une croix dans la case correspondante pour chaque item	d'accord	un peu d'accord	pas d'accord
1-(OP1) C'est vrai que les math constituent un instrument de sélection excessif.			
2-(OP2) An bac, je préfère qu'il y ait un grand problème avec plusieurs parties plutôt qu'un ensemble de petits problèmes indépendants.			
3-(OP3) Dans ma notation, j'attache plus d'importance à la démarche qu'au résultat.			
4-(OP4) Quand je corrige, j'aime bien un barème très détaillé sur les résultats à obtenir.			
5-(OP5) La démonstration est la seule façon rigoureuse de faire des mathématiques.			
6-(OP6) Je préfère des programmes bien définis indiquant ce que je dois et ce que je ne dois pas faire.			
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.			
8-(OP8) pouvoir donner un exemple ou un contre-exemple personnel à l'affirmation : "si deux applications *f* et g sont strictement croissantes sur un intervalle, l'application produit *fxg* y est également croissante".			
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.			
10-(OPX) estimer à vue, à 30% près, le périmètre et l'aire du plancher ainsi que le volume de la salle de classe.			

Les 26 variables correspondantes ne sont pas binaires, mais ordinales ou modales, elles prennent des valeurs décimales sur [0 ;1], comme indiqué ci-dessous. Ainsi l'analyse intègre l'intensité des attitudes, le choix prioritaire d'un objectif pondérant de façon différente un choix plus secondaire, voire non retenu. Pour ce faire, les enseignants font choix de 6 objectifs parmi 15 qu'ils assignent à leurs enseignements (ex : A : « Acquisition de connaissances », B : « Préparation à la vie professionnelle ») et d'opinions relatives à dix phrases communément énoncées (par exemple : OP 1 : « les maths constituent un instrument de sélection excessif ») (Bodin et Gras, 1999).

Le codage est le suivant en tenant compte des modalités ordonnées :

Variables A à O	1 → 1 ; 2 → 0,8 ; 3 → 0,6 ; 4 → 0,4 ; 5 → 0,2 ; 6 → 0,1 ; Vide → 0
Variables OP1 à OPX	D'accord → 1 ; Un peu d'accord → 0 ,5 ; Pas d'accord → 0.
Variable PER (pertinent)	oui → 1 ; non → 0

Tableau 3 : codage numérique des variables selon les contraintes de l'ASI

On trouvera dans l'annexe 2 le tableau récapitulatif des résultats de l'enquête. Rappelons que 311 enseignants y ont répondu

4.1 Étude de la fractalité de l'arbre cohésitif des opinions

Le traitement par CHIC du tableau de données conduit, par l'option arbre cohésitif, à l'arbre suivant (fig. 9).

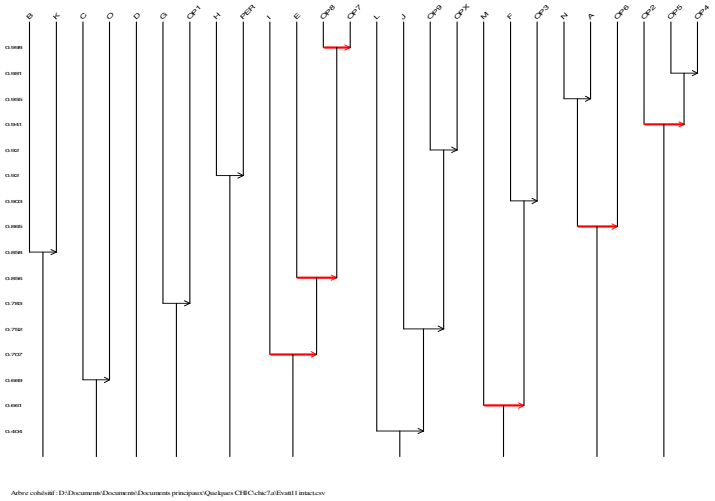


Figure 9 :Arbre cohésitif du fichier de données EVATT1 (26 variables modales et 311 sujets)

Dans l'arbre cohésitif obtenu par CHIC (fig. 9), nous commencerons à étudier l'éventuelle fractabilité, au niveau 4, véritable niveau significatif non trivial. Nous établissons le tableau des estimations des dimensions afférentes aux différents niveaux de 4 à 16, niveau maximal.

Au niveau 5, on a N= 6 et C= 0.92 d'où D_C = 2.31

Au niveau 8, on a N= 10 et C= 0.865 d'où D_C = 2.75

Au niveau 13, on obtient N=19 pour C= 0.707 puis D_C= 2.83, alors qu'au niveau significatif 10 la cohésion est 0.856 pour N=13 ce qui donne D_C=3.02. Au niveau significatif 15, on obtient N = 22 pour une cohésion minimale de 0.661 ce qui conduit à D_C=2.79.

Niveau de l'arbre	Cohésion C à ce niveau	Nombre N de niveaux pour coh≥C	D_C calculé
4	0.941	5	**2.13**
5	0.92	6	**2.31**
6	0.92	7	**2.68**
7	0.903	8	**2.62**
8	0.865	10	**2.75**
9	0.858	11	**2.83**
10	0.856	13	**3.02**
11	0.783	14	**2.81**
12	0.752	16	**2.83**
13	0.707	19	**2.83**
14	0.669	20	**2.76**
15	0.661	22	**2.79**
16	0.404	25	**2.01**

Tableau 4 : Détermination de la dimension de l'arbre cohésitif

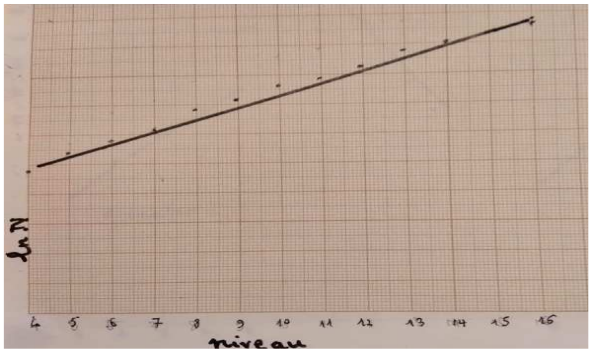


Figure 10 : Courbe opinions (N, ln(N))

La moyenne des valeurs prises par D_C est 2.64 et l'écart-type est 0.324. Dans le modèle fractal considéré sous-jacent à l'image de l'arbre cohésitif, on peut considérer, eu égard au faible écart-type, que la dimension fractale de l'arbre est 2.6. Ainsi, la géométrie de l'arbre cohésitif semble bien apparentée à la géométrie fractale au sens de Mandelbrot (1995, 1997). La dimension fractale mesure le degré d'irrégularité de l'arbre, de sa brisure dirait Mandelbrot.

4.1.1 Remarque 1

Cette valeur de dimension peut être utilisée de façon inverse, à savoir : retrouver à quel niveau correspond une certaine valeur de cohésion. Par ex. utilisant la formule inverse : $\log N_C = D_C.\log (2/C)$, on obtient pour C= 0.783, $N_C = 11.45$ alors que le niveau de hiérarchie est 11. Mais notons que cette propriété opératoire d'inversion est quelque peu locale puisque nous observons que la décroissance de la cohésion n'est accompagnée de la croissance de D_C que jusqu'aux niveaux médians. Elle pourrait cependant nous alerter de l'existence d'un niveau intermédiaire qui aurait été oublié ou négligé comme pourrait l'être un sous-niveau cognitif. De nouvelles explorations nous paraîtraient nécessaires pour consolider la conjecture d'opérativité de l'arbre cohésitif comme nous l'avons exprimée avec le graphe implicatif. Dans notre communication intitulée *Evolution darwinienne et classification orientée A.S.I*, nous avions une autre approche de la fractalité de la hiérarchie qui conforte cette fractalité.

4.1.2 Remarque 2

Une autre procédure de décompte N des classes formées à un niveau hiérarchique quelconque consiste récursivement à ajouter au nombre de classes une unité correspondant à la nouvelle classe, c'est-à-dire aucune des classes générées par l'apparition de cette nouvelle. Par exemple, la formation de la classe (a,(b,c)) ne conduit plus à majorer N-1 de 2 unités Ainsi, au niveau 10, on disposera de N=10 classes et non pas 13 obtenues dans le précédent décompte. Cette procédure nous conduit à une dimension fractale D=2,35 avec un écart-type de 0.23 qui permet d'améliorer sensiblement des prédictions inverses.

4.2 La hiérarchie cohésitive traduit une évolution mentale des opinions des enseignants

Nous avons déjà mis en évidence dans des travaux précédents (Gras, Gregori, 2019) et, en particulier lors du colloque ASI 9 de Belfort en 2019, la restitution préordonnée du graphe implicatif en même temps que sa métaphore fractale. Nous connaissons la propriété hiérarchique et non linéaire au niveau conceptuel de l'arbre cohésitif. Nous pouvons le noter ici encore en effectuant l'analyse des 3 classes significatives :

4.2.1 (si N alors A) alors OP6.

Cette R-règle peut être lue ainsi : « si l'acquisition de savoir-faire (N) doit s'accompagner de celle des connaissances (A), alors le professeur demande que les programmes soient bien définis (OP6). On peut faire l'hypothèse que cette relation soit de type causal : « je suis très attaché aux contenus d'enseignement, mais, acteur dans l'institution, je demande que celle-ci définisse et précise ses choix ». Autrement dit, toute centration sur les savoirs exige un encadrement préalable de l'institution à travers des programmes. On observe alors que la règle généralisée permet de donner une signification plus synthétique aux règles qui la constituent : on passe des comportements, au sens behavioriste, à une conduite supérieure qui piloterait le comportement. On peut dire, de façon contraposée que ce sont les conditions drastiques du milieu qui vont favoriser l'acquisition de savoirs solubles, voire précipités (sens chimique) en savoir-faire. L'évolution du milieu vers le sujet est manifeste. Elle conduit à la construction

d'un modèle d'enseignant s'appuyant sur un modèle psycho-social. S'ensuit alors une certaine **évolution.**

4.2.2 si OP2 alors (si OP5 alors OP4).

Explicitement, cette R-règle se lit de la façon suivante : si l'on considère un grand problème indispensable à l'examen (OP2), alors, considérant que la démonstration est la seule façon rigoureuse de faire des mathématiques (OP5), le barème de correction (OP4) doit servir d'échelle. Comme ci-dessus, il y a **élargissement** d'une spécification à une fonction générative.

4.2.3 si I alors (si E alors (si OP8 alors OP7)) (classe significative).

En effet, on focalise de comportements calculatoires spécifiques OP7 puis OP8 vers une propriété cognitive E (la créativité) appliquée au champ conceptuel de la divisibilité, puis vers la compétence de la mise à distance critique I des concepts enseignés ce qui représente un niveau conceptuel plus élevé. Dans la taxonomie d'objectifs cognitifs dont nous avons déjà parlé, ces deux objectifs E et I relèvent respectivement de classes cognitives rangées dans cet ordre (cf. Annexe 2). On retrouve, dans cette classe, la propriété extensive, propre à un **développement cognitif** déjà validé plus haut.

4.2.4 Hypothèse sur la nature de la classification cohésitive

Cet exemple illustre clairement la citation de Pascal rappelée par E. Morin dans (Morin, 2015): « *Je tiens pour impossible de connaître les parties en tant que parties sans connaître le tout, mais je tiens pour non moins impossible la possibilité de connaître le tout sans connaître singulièrement les parties* ». D'où des structures gigognes où, à chaque niveau d'échelle, on passe d'un « quoi » ou un « pourquoi » à un « comment » explicatif ou commentatif.

En effet, cet exemple permet de répondre à des suites de questions préordonnées en termes de « pourquoi » puis de « comment ». Par exemple, pourquoi un enseignement de mathématiques ? pourquoi un médiateur (l'enseignant) ? comment satisfaire la finalité de l'enseignement ? quelles en sont les articulations ? On voit alors que la relativité d'échelle est la théorie qui articule, pas à pas, cette scansion et confirme de ce fait la structure fractale de l'arbre cohésitif.

Insistons sur l'accroissement, dans chacun de ces cas, de la richesse de l'analyse obtenue par l'association des règles d'association en des R-règles. Ce ne sont plus seulement des faits ou des comportements isolés qui sont extraits, mais plutôt des conduites générales, révélatrices elles-mêmes de phénomènes plus globaux, moins singuliers ou de représentations psychologiques profondes selon trois types de profils, de modèles d'enseignants de mathématiques. On parlerait, métaphoriquement, d'**espèces** comme dans le paragraphe suivant, aboutissement de crises. Une typologie comme en fournissent les classifications traditionnelles (donc symétriques), ne pourrait pas rendre compte de la dynamique des faits ou comportements sous-jacents débouchant sur des structures stables à la suite d'une **genèse**, ce que justement produit l'arbre cohésitif. C'est pourtant cette **dynamique** restituée par les règles généralisées qui, appuyée sur des nécessités (les prémisses), conduit à des élucidations vives d'un fragment de théorie, éventuellement en voie de construction.

Notre hypothèse sur la nature de la classification cohésitive comme avatar du développement psycho-social est renforcée.

5 Une situation sans référence à un développement

Se pose alors la question : existe-t-il une relation entre la hiérarchie cohésitive et la fractalité ? Autrement dit, la relation cohésitive est-elle redondante de la fractalité ?

Il est bien évident que tout tableau de données sujets x variables analysé par l'A.S.I. et dont les ensembles de règles orientées sont représentées par une hiérarchie cohésitive, n'ont pas vocation à traduire l'existence d'un développement fractal.

Nous partons cette fois d'un fichier de données dont les variables modales sont les 20 traits de personnalité tel que les définissent les psychologues du travail. Les données ont été recueillies dans dix pays (France, Italie, Inde, etc.) sur un total de 2450 sujets à raison d'un échantillon de 245 sujets dans chacun de ces pays. A un sujet quelconque est attachée une valeur numérique entre 0 et 1 en fonction du degré d'intensité de réponse à l'un des 20 traits.

Voici la signification des codages des 20 traits du modèle étudié :

AFF	Affirmation	DIN	Dynamisme intellectuel	POU	Pouvoir
ANX	Anxiété	DTN	Distanciation	PRO	Protection
APP	Appartenance	EXT	Extraversion	REA	Réalisation
CCL	Conciliation	FAC	Facilitation	REC	Réceptivité
CIN	Conformisme intellectuel	IMP	Improvisation	RIG	Rigueur
COM	Combativité	IND	Indépendance	RMC	Remise en cause
DET	Détente	INT	Introversion		

Tableau 5 : Codage des 20 variables « traits »

À partir de la matrice des 2450 x 20 données principales, croisant sujets et variables actives, il a été pratiqué une analyse des règles orientées, accompagnées de l'apport explicatif fourni par les variables supplémentaires comme l'âge, le sexe, la profession, le niveau d'études, etc.

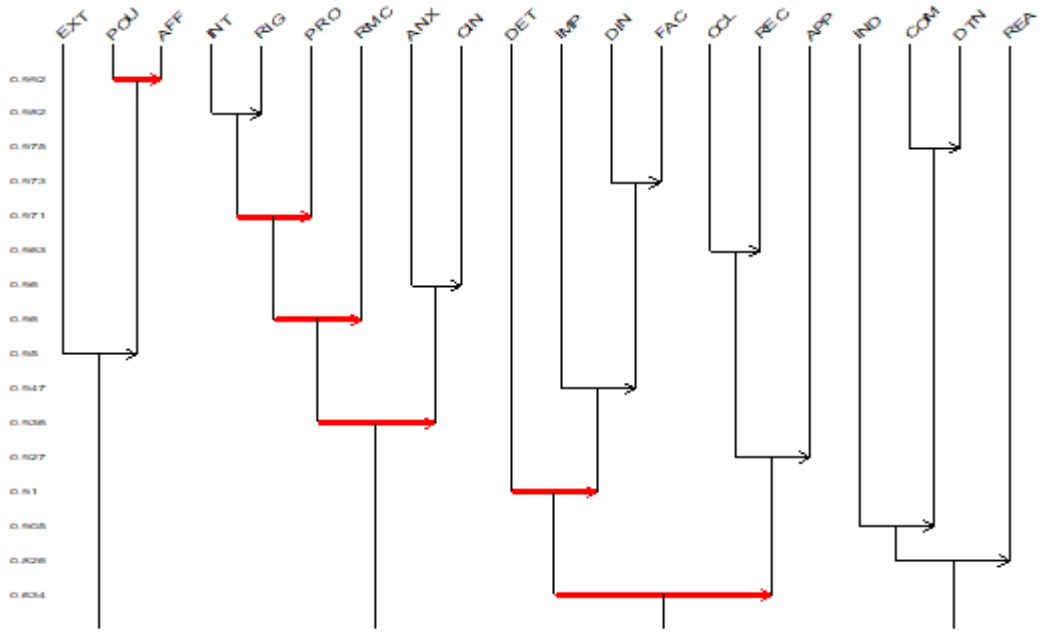


Figure 11 : Hiérarchie cohésitive des 16 « traits » en France

Examinons, à travers le tableau suivant s'il existe un soubassement fractal dans le phénomène des traits de caractère que dévoilerait l'arbre cohésitif.

Niveau de l'arbre	Cohésion C à ce niveau	Nombre N de niveaux pour coh≥C	DC calculé
2	0.982	2	0.974
3	0.975	3	1.53
4	0.973	4	1.92
5	0.971	5	2.23
6	0.963	6	2.45
7	0.960	7	2.65
8	0.960	8	2.83
9	0.950	9	2.95
10	0.947	10	3.08
11	0.936	11	3.16
12	0.927	12	3.23
13	0.910	13	3.27
14	0.905	14	3.32
15	0.826	15	3.06
16	0.634	16	2.41

Tableau 6 : Détermination de la dimension de l'arbre cohésitif

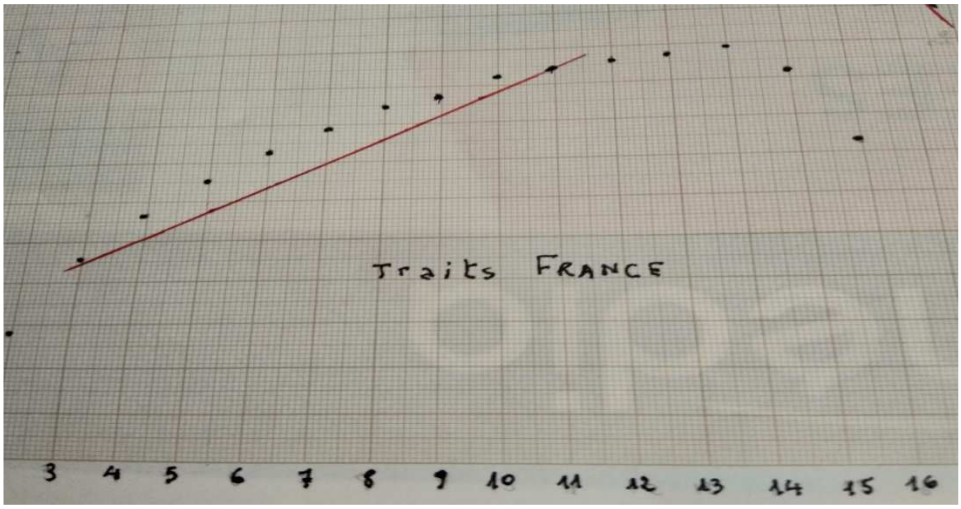


Figure 12 : Représentation logarithmique des données des traits en France

Le questionnaire ici traité ne rend pas compte d'une évolution, d'un développement, anthropologique en l'occurrence. En effet, l'ensemble des variables ne sous-entend pas l'existence d'un ordre ou d'un préordre sous-jacent comme on le rencontre dans la comparaison des temps perçus par rapport aux temps réels ordonnés par l'échelle de l'âge dans la vie. Comme on le rencontre aussi dans une ontogenèse ou la phylogenèse du vivant. Il n'est donc pas étonnant que la propriété liée à la géométrie fractale, à savoir un graphe linéaire où la pente révèle la dimension D, ne soit pas présente dans cet exemple. En revanche, nous en avions observé la réalisation dans la représentation hiérarchique cohésitive des résultats du test taxonomique dont nous rendons compte dans (Gras, Gregori, 2019) et plus haut dans cet article.

En résumé, retenons que la représentation d'un tableau de contingences, selon une hiérarchie cohésitive d'un phénomène ne présume pas, nécessairement, une propriété fractale du phénomène représenté.

6 Hiérarchie des primates

Dans le texte des Actes ASI 9, (Gras, 2017) nous avions souligné l'intérêt opératoire de la propriété fractale du graphe implicatif : inférer la connaissance empirique de la dimension fractale à des graphes construits selon des seuils implicatifs variables, inférence portant sur l'estimation du nombre de règles non encore extraites. Comme nous l'avions dit, avec Laurence Ndong, nous avons lors d'une communication de ASI 9[8], puis de ASI 10[9], entrepris de présenter un autre type de classification : au lieu de classifier les êtres vivants nous avons choisi de classifier leurs propriétés, en l'occurrence les propriétés morphologiques phénotypiques, faute de disposer d'autres informations de type biologique. Nous avons obtenu une hiérarchie cohésitive pour représenter la classification des propriétés des primates. Grâce à l'ASI et son outil logiciel CHIC nous pouvons en outre identifier les sujets qui ont contribué, préférentiellement à la constitution des classes de propriétés. Il s'agit donc d'une autre façon de classer les espèces et les éléments qui en relèveraient.

Nos connaissances limitées en paléontologie et en biologie ne nous permettent pas de présenter un autre type d'arbre de vie complet sous-jacent à l'évolution au sens de Darwin, sous la forme d'un arbre cohésitif alors que, comme le montre Darwin et la *généalogie génétique*, et comme nous l'avons constaté ci-dessus, l'arbre de vie n'est manifestement pas linéaire ou composé de classes emboîtées. Mais, tout comme nous le voyons avec le phénomène de conceptualisation, éminemment dynamique, les classes s'organisent aussi selon un préordre. Celui-ci peut alors être représenté par un graphe orienté ou par une hiérarchie orientée. C'est aussi pour cette raison que les solutions graphiques de l'ASI sont préférables aux solutions fournies généralement dans les ouvrages qui traitent les lois de l'évolution en laissant penser à une symétrie ou une réversibilité des propriétés, ce qui est manifestement faux. L'échantillon de singes retenu est composé des animaux suivants : Toupaïe, Maki, Indri, Tarsier, Saki, Macaque, Babouin, Homme, Gorille, Bonobo, Orang-outan et Chimpanzé.

Les caractères retenus sont les suivants :

- Présence ou absence d'ongle et de pouce opposable ;
- Orbites fermées ou ouvertes ;
- Présence ou absence de nez ;
- Présence ou absence de narines rapprochées ;
- Présence ou absence de queue.

Les variables binaires qui caractérisent les sujets, sont les suivantes :

Ongpop1 :	Présence d'ongle / pouce opposable	Orfer2 :	Pas d'orbites fermées
Ongpop2 :	Absence d'ongle / pouce opposable	Narap1 :	Présence de narines rapprochées
Nez1 :	Présence de nez	Narap2 :	Absence de narines rapprochées
Nez2 :	Absence de nez	Queue 1 :	Présence de queue
Orfer1 :	Orbites fermées	Queue 2 :	Absence de queue

Tableau 7 : Codage des variables binaires des « traits ».

L'arbre phylogénétique ci-dessous a été obtenu à partir du logiciel Phylogène[10].

8 *Classification des primates et classification hiérarchique orientée de l'A.S.I.* par Régis Gras et Laurence Ndong in Actes ASI9.

9 *Évolution darwinienne et classification orientée*, par Régis Gras et Laurence Ndong in Actes ASI10.

10 http://acces.ens-lyon.fr/evolution/logiciels/phylogene

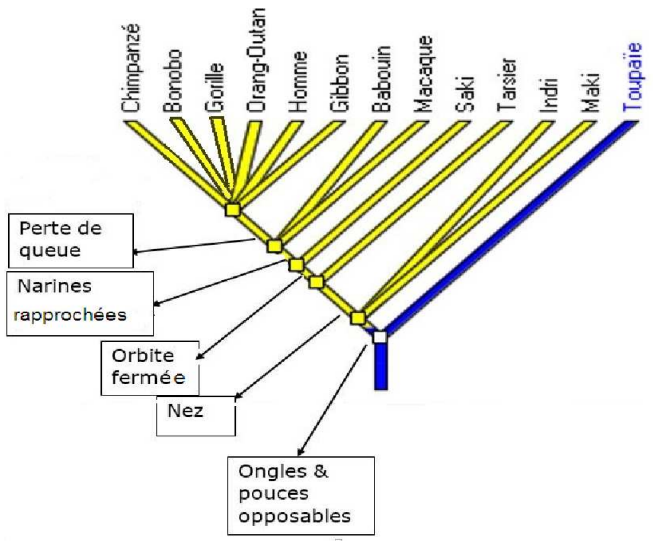


Figure 13 : Arbre phylogénétique de certains primates et archonte non-primate.

En croisant caractères retenus et animaux de notre échantillon les possédant ou non à l'aide de variables binaires, nous avons construit une matrice de variables binaires que nous avons traités avec le logiciel CHIC. Nous avons obtenu l'arbre cohésitif ci-dessous :

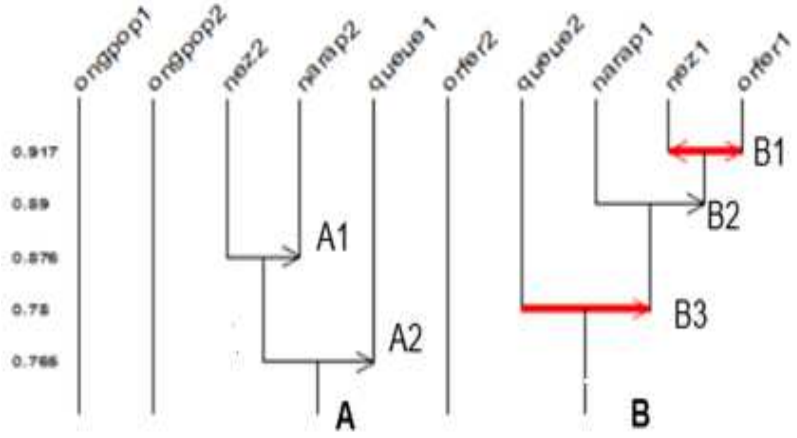


Figure 14 : Arbre cohésitif des primates

Les deux grandes classes **A** et **B** sont disjointes par l'algorithme cohésitif ASI, signant une **séparation nette au niveau chronologiq**ue dans l'évolution entre deux grandes familles, observation plus nette que celle fournie par la Fig. 11. Après des recherches nombreuses sur internet, il ne nous pas été possible de préciser les étapes datées de l'apparition de sous-espèces telles que le présente l'arbre cohésitif. Mais il nous paraît évident que les différentes étapes B3⇒B2⇒B1et A1⇒A2 marquent des sauts évolutifs datés par les paléontologues comme, par exemple, la disparition de la queue (queue 2). En conclusion, la hiérarchie cohésitive se présente ici aussi comme avatar d'un développement, en l'occurrence l'évolution des espèces. Une question intéressante se pose au regard de l'arbre cohésitif : peut-on donner une interprétation en termes évolutifs aux niveaux significatifs de l'arbre ?

Vérifions maintenant le respect de la loi log-périodique sur la hiérarchie cohésitive comme il a été fait plus haut. On observe, en effet, la quasi-linéarité du graphe de la fonction niveau N→ln N.

Niveau de l'arbre	Nombre N de classes	ln N
1	3	1.1
2	4	1.39
3	5	1.61
4	7	1.95
5	8	2.08

Tableau 8 : Primates niveau-classes

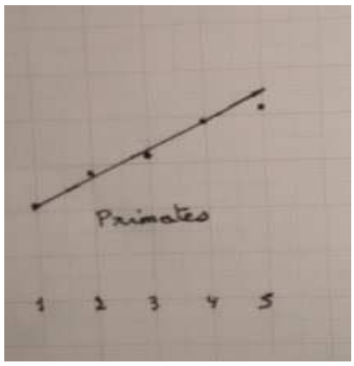


Figure 15 : Courbe primates (N, ln N)

Le travail présenté ci-dessus n'a pas été établi selon les formules de L. Nottale et J. Chaline. Il serait intéressant de revenir sur les données fournies par L. Ndong et sur la méthodologie hiérarchique employée en utilisant le rapport d'échelle g fourni par la recherche soit 1.736 $\mp$ 0.013 et le temps critique propre aux primates T_c = + 2.25 $\mp$ 0.80 Ma.

7 Conclusion

Il s'avère donc que, non seulement l'Analyse Statistique implicative permet de mettre en évidence des relations causales pondérées entre des phénomènes, mais aussi traduit, au moyen de représentations graphiques, certaines structures de l'ensemble des règles (graphe implicatif) et plus encore d'un ensemble généralisé de règles de règles (arbre cohésitif). On a vu que, bien souvent, cette dernière structure, représentée par une hiérarchie orientée, permettait de mettre au jour des genèses, des développements, des évolutions constitués d'équilibres et de changements orientés selon une même orientation. On a montré à travers plusieurs exemples que cette orientation pouvait être

- celle du développement cognitif.
- mais aussi celle d'une opinion en psycho-sociologie, élément d'un rapport social,
- ou encore des lois de l'évolution darwinienne.

Gérard Vergnaud, et nous avec lui maintenant, parlerions d'**homomorphisme** de structures entre ces trois domaines. Piaget lui-même (1998) affirme que « *le constructivisme, en épistémologie, est beaucoup plus comparable -c'est une conviction profonde- à l'évolution vitale organique de la vie sur le plan biologique qu'à tout autre développement* ». Cette évolution vitale se plie certes aux caprices du hasard mais aussi respecte des règles déterministes probabilistes fonctions des organismes comme le fait la physique.

Il existe bien des points communs entre ces trois domaines comme nous l'avons exprimé dans l'introduction et portant sur l'univers, la vie et les sociétés :

- ils évoluent dans le temps ;
- ils connaissent des ruptures ;
- ils sont hiérarchisés en niveaux selon une structure fractale pour les arbres évolutifs ;
- le franchissement d'étape est comparable à une rupture à la suite d'une crise dans le développement ;

- la prédictibilité de l'existence de structure à venir est réelle.

Il est donc apparu que la hiérarchie cohésitive, appliquée aux données de ces domaines était un excellent avatar de l'ASI pour représenter un développement, ce qui est plus faiblement le cas du graphe implicatif. Mais cette hiérarchie, à elle-seule n'est pas suffisante pour traduire une évolution, un développement. Cependant, elle le révèle. La variété et les caractères étrangers des exemples nous permettent de supposer l'extension de la propriété de la hiérarchie à d'autres situations dont elle décèlerait la composante développementale exprimée ou sous-jacente. Elle pourrait même dans un corpus indéfini tenir lieu de révélateur d'un développement inconnu ou caché, total ou partiel. Les exemples choisis ont montré en outre que dès que les phénomènes associés relevaient clairement d'un développement, les représentations étaient des objets de la géométrie fractale. De ce fait, elles en héritaient la richesse des références intuitives ainsi qu'une certaine prédictibilité des propriétés évolutives.

Références

[7] Bachelard G. (1967) *La Formation de l'esprit scientifique*. Paris : Librairie philosophique J. Vrin.

[8] Bloom B.(1969), *Taxonomie des objectifs pédagogiques,* Tome 1*: Domaine cognitif*, Education Nouvelle.

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

[10] Chaline J. (2006) *Quoi de neuf depuis Darwin ?* Paris : Ellipses.

[11] 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'État, Université de Rennes I.

[12] Gras R., Kuntz P. (2005), Discovering R-rules with a directed hierarchy, *Soft Computing, A Fusion of Foundations, Methodologies and Applications,* Volume 1, p.46-58, ISSN 1432-7643*,* Springer Verlag.

[13] Gras R., Bodin A. (2017) L'ASI, analyseur et révélateur de la complexité cognitive taxonomique. J.-C. Régnier, R. Gras, R. Couturier, A. Bodin (Eds) *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines* Université de Bourgogne Franche-Comté, p.128-142.

[14] Gras R. (2017) Dimension fractale d'un graphe implicatif. In J.-C. Régnier, R. Gras, R. Couturier, A. Bodin. (Eds) *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines,* Université de Bourgogne Franche-Comté, p.143-153.

[15] Gras R (2018) La logique paracohérente de l'Analyse Statistique Implicative. In R. Gras, *La théorie de l'Analyse Statistique Implicative*. Toulouse : Éditions Cépaduès.

[16] Gras R. (2018). Dimension fractale d'un graphe implicatif. In R. Gras, *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*, Toulouse : Éditions Cépaduès.

[17] Gras R., Gregori P. (2019) Arbre cohésitif et métaphore fractale. In J-C. Régnier, R. Gras, M. Henry, R. Couturier, G. Brousseau (Eds), *Analyse Statistique Implicative, Cadre*

théorique en relation étroite et au service de multiples disciplines. Belfort : Université de Bourgogne Franche-Comté p.90-105

[18] Gras R., Ndong L. (2019) Evolution darwinienne et classification orientée In J-C. Régnier, R. Gras, M. Henry, R. Couturier, G. Brousseau (Eds), *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines*. Belfort : Université de Bourgogne Franche-Comté. p.106-122.

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

[20] Mandelbrot B. (1995) *Les objets fractals*, *Forme, hasard et dimension*. Paris : Flammarion.

[21] Mandelbrot B. (1997) *Fractales, hasard et finance*. Paris : Flammarion.

[22] Morin E. (2015) *Penser global.* Paris : R. Laffont.

[23] Morin E. (2020) *Changeons de voie*. Paris : Denoël.

[24] Nottale L., Chaline J. et Grou P. (2000) *Les arbres de l'évolution*. Paris : Hachette.

[25] Piaget J. (1970) *L'épistémologique génétique*, Paris, P.U.F.

[26] Piaget J. (1998) Le constructivisme épistémologique, *Bulletin de Psychologie*, 51-3, 435, 226-232.

[27] Picq Pascal, (2019). *Sapiens face à Sapiens,* Flammarion,

[28] Querros-Conde D., Chaline J. et Dubois J. (2015) *Le monde des fractales.* Paris Ellipses

[29] Régnier J.-C., Gras R., Henry M., Couturier R., Brousseau G. (Eds) (2019) *Analyse Statistique Implicative : cadre théorique en relation étroite et au service de multiples disciplines*. Actes du colloque ASI 10, 2-5/10/19 à IUT Belfort-Montbéliard , ISBN : 978-2-9562045-2-7, 604 p.

[30] Richardson L. F. (1961) The problem of contiguity : an appendix of statistics of deadly quarrels, *General Systems Yearbook, 6* ; 139-167

[31] Vergnaud G. (2007) Héritages, *Activité humaine et conceptualisation, Questions à Gérard Vergnaud,* Coordonné par Maryvonne Merri, Toulouse : Presses Universitaire du Mirail

[32] Vergnaud G. (1996) La théorie des champs conceptuels. In J. Brun (Ed). *Didactique des Mathématiques*. Lausanne : Delachaux et Niestlé

Annexe 1

Taxonomie de Régis GRAS (1979)

TABLEAU D'OBJECTIFS COGNITIFS

CATEGORIES	RUBRIQUES	OBJECTIFS	ACTIVITES ATTENDUES
(A) Connaissance des outils de préhension de l'objet et du fait mathématiques	A_1	Connaissance de la terminologie et du fait spécifique	associer assembler
	A_2	Capacité d'action intériorisée sur l'évocation d'une forme physique du concept	simuler observer
	A_3	Capacité de lire des cartes, des tableaux, des graphiques, des notices	déchiffrer décrire
	A_4	Effectuation d'algorithmes simples	organiser, calculer
(B) Analyse de faits et transposition	B_1	Substitution d'une démarche représentative à une manipulation Anticipation graphique	abstraire prolonger induire
	B_2	Reconnaissance et usage d'une relation implicite simple où intervient l'objet mathématique connu	analyser comparer
	B_3	Traduction d'un problème d'un mode dans un autre avec interprétation	schématiser, traduire transposer
(C) Compréhension des relations et des structures	C_1	Compréhension du concept, de ses relations avec les autres objets mathématiques	reconnaître construire
	C_2	Compréhension d'un raisonnement mathématique : justification d'un argument	justifier
	C_3	Choix et ordonnancement d'arguments	déduire
	C_4	Application dans des situations familières	analyser, abstraire appliquer, interpoler
(D) Synthèse et Créativité	D_1	Effectuation et découverte d'algorithmes composites et de nouvelles relations	organiser, calculer optimiser
	D_2	Construction de démonstrations et d'exemples personnels	illustrer, démontrer, valider, créer, inventer
	D_3	Découverte de généralisations	généraliser, induire, prévoir, extrapoler
	D_4	Reconnaissance du modèle et applications dans des situations non routinières	modéliser, identifier, différencier, classifier, résumer
(E) Critique et Evaluation	E_1	Distinction du nécessaire et du suffisant	formuler des hypothèse, déduire
	E_2	Critique de données et de méthodes ou de modèles résolvants	contrôler, optimiser, prévoir, critiquer, questionner, vérifier
	E_3	Critique d'argumentation et construction de contre-exemple	critiquer, tolérer contredire

Annexe 2

Tableau des résultats de l'enquête

	opinion	Occurrence	Moyenne	Ecart-type
1	A	105,70	0,34	0,40
2	B	8,80	0,03	0,13
3	C	9,70	0,03	0,15
4	D	140,00	0,45	0,41
5	E	21,80	0,07	0,20
6	F	138,70	0,45	0,38
7	G	19,50	0,06	0,18
8	H	44,80	0,14	0,24
9	I	83,10	0,27	0,32
10	J	108,40	0,35	0,36
11	K	77,60	0,25	0,32
12	L	4,60	0,01	0,09
13	M	90,20	0,29	0,33
14	N	66,60	0,21	0,31
15	O	33,20	0,11	0,23
16	OP1	81,50	0,26	0,33
17	OP2	147,50	0,47	0,43
18	OP3	242,50	0,78	0,27
19	OP4	229,00	0,74	0,35
20	OP5	190,00	0,61	0,39
21	OP6	240,00	0,77	0,32
22	OP7	200,00	0,64	0,43
23	OP8	165,00	0,53	0,44
24	OP9	98,00	0,32	0,40
25	OPX	207,00	0,67	0,38
26	PER	254,50	0,82	0,36

FRACTALITÉ DU RYTHME DU TEMPS SUR L'ARBRE DE VIE

Régis Gras[1], Raphaël Couturier[2], Saddo Ag Almouloud[3]

FRACTALITY OF THE RHYTHM OF TIME ON THE TREE OF LIFE

RÉSUMÉ

Le texte que nous présentons ci-dessous se charpente selon plusieurs facettes renvoyant toutes à la notion de fractalité, concept mathématique géométrique imaginé par Mandelbrot. Nous avons fréquenté ce concept en A.S.I. lors de l'étude des représentations graphiques des relations causales (graphe et arbre). Cet article vise à renforcer la familiarisation du lecteur avec cette géométrie en lui montrant un autre champ d'application et sa puissance créative. Nous partons d'une question vive qui perturbe chacun d'entre nous et de façon accentuée au fil de la vie, à savoir la différence numérique entre le temps réel, c'est-à-dire celui de l'horloge et le temps perçu par un sujet existant. Moyennant deux hypothèses plausibles sur la relation entre ces deux mesures, l'une physique, l'autre psychologique, nous modélisons une formule générique reliant ces deux notions. Celles-ci constituent une distorsion, autrement dit un biais cognitif. Nous montrons qu'il s'agit d'un phénomène fractal plongé dans l'espace-temps de la relativité restreinte. En effet, il apparaît que la représentation graphique logarithmique des covariations de ces deux notions est linéaire. Nous savons que cette dernière propriété, comme nous l'avons observé par ailleurs avec les images graphiques de l'A.S.I., est caractéristique d'un phénomène fractal

Mots-clés : *biais cognitif, temps réel, temps ressenti, temps perçu, relativité, espace-temps, échelle logarithmique, fractal.*

ABSTRACT

This paper is structured according to several facets all referring to the notion of fractality, a geometric mathematical concept imagined by Mandelbrot. We have come across this concept in A.S.I. when studying graphical representations of causal relationships (graph and tree). This article aims at strengthening the readers' familiarity with this geometry by showing them another field of application and the creative power of the latter. This reflection springs from a controversial issue which gets more and more pressing as people age, namely the digital difference between real time and perceived time by an existing subject. Using a plausible hypothesis on the relationship between these two measures, one physical, the other psychological, this study models a generic formula linking these two notions. These notions constitute a distortion, in other words a cognitive bias. This is a fractal phenomenon immersed in the space-time of special relativity. Indeed, it appears that the logarithmic graphical representation of the covariations of these two concepts is linear. This last property, as can be observed, with the graphic images of A.S.I., is characteristic of a fractal phenomenon.

Keywords*: cognitive bias, real time, felt time, perceived time, relativity, space-time, logarithmic scale, fractal.*

[1] Ecole Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire LS2N, UMR 6241, Site de la Chantrerie, rue C.Pauc, BP 44306, Nantes cedex 3, e-mail : regis.gras@univ-nantes.fr, regispor@hotmail.fr.

[2] Univ Bourgogne Franche-Comté (UBFC), Institut FEMTO-ST, Belfort, raphael.couturier@univ-fcomte.fr

[3] Université Fédérale de Pará, Brésil, saddoag@gmail.com

Figure 1 : D'après W. Shakespeare

Préambule

Mise en bouche : *Le temps des tout petits*. Extrait du magazine SLATE

Mais pourquoi mesurer notre perception à l'aide d'une échelle logarithmique ? Parce que notre rapport au temps est relatif à l'âge que nous avons : pour un petit de 2 ans, une année représente la moitié de sa vie, ce qui explique que la distance entre deux anniversaires semble si longue à un enfant. Pour un marmot de 10 ans, une année ne représente que 10% de sa vie (il devient alors un peu moins difficile d'attendre son prochain anniversaire), et pour un jeune de 20 ans, une année ne représente plus que 5% de sa vie. Voilà pourquoi le temps semble s'accélérer avec l'âge. Habituellement, nous envisageons la vie par tranches de dix ans – la vingtaine, la trentaine, etc.– ce qui revient à accorder autant de poids à chaque période. Cependant, selon l'échelle logarithmique, nous percevons des périodes de durées différentes comme si elles étaient de même durée. En termes de perception du temps, passer de 10 à 20 ans, de 20 à 40 et de 40 à 80 revient exactement au même. Ce qui veut dire – attention, c'est un peu déprimant – que les années vécues entre l'âge de 5 et 10 ans vous paraissent aussi longues que les années qui séparent vos 40 ans de vos 80 ans.

Alors, activez-vous ! Le temps passe vite… et ça ne s'arrange pas avec…. le temps[4].

1 Introduction

Qui ne s'est pas interrogé sur la perception de l'écoulement du temps dans son propre vécu ? Qui ne s'est pas étonné que la fin du mois se soit précipitée par rapport à ce qu'il savait de son activité récente ? Qui ne s'est pas étonné que la photo qu'il a en main ne soit pas d'il y a 2 ans mais plutôt 4 ans ? Nous avons fait nous-mêmes de fausses estimations d'un passé et nous nous sommes trouvés surpris de l'éloignement dans le temps de ce que notre mémoire erronée plaçait récemment.

Nous avons rencontré fréquemment ce phénomène au fil de notre vie, de nos travaux et de nos lectures et l'avons accepté comme naturel et inexorable, inhérent à la nature humaine. Nous en avons déduit que le monde dans lequel nous vivions n'était pas

[4] Selon Aristote, le temps est la mesure du changement. Nous *appelons « temps » la mesure, la comptabilité de ce changement.* » (C. Rovelli, p.80) C'est cette mesure que l'on modélise. Newton, comme Leibniz ne pensent pas qu'il existe un temps absolu qui s'écoule en soi, indépendamment des choses, comme une entité qui s'écoule uniformément et imperturbablement.

nécessairement celui que nous percevions. En particulier, existait une distorsion, un biais cognitif dirait le psychologue, entre le temps ressenti et le temps réel[5]. On y affirme que le premier s'exprime dans un rapport au second, et, tout au moins, respecte les relations d'ordre dans chacune des deux familles qu'elles soient observées ou vécues. Dans l'actualité des années 2020 et 2021, confrontés aux confinements face au Covid 19 , les décalages entre ces deux mesures du temps ont été plu vives encore.

Dans la vie courante, nous rencontrons d'autres types de biais cognitif, par exemple, l'effet de rareté (valorisant un produit d'autant qu'il est rare), l'effet de supériorité de l'image par rapport aux mots, la perception de la température extérieure, ou encore le biais de complaisance.

Le lecteur, et *a fortiori* l'utilisateur de l'A.S.I., ne sont pas étonnés de cette ambivalence des concepts filtrés, biaisés par la philosophie de l'A.S.I. Nous l'avons étudiée en mettant en évidence une logique paracohérente LSI, sous-jacente à l'A.S.I. et fondement de cette philosophie (Gras, 2018). Logique paracohérente, cognitivement opposée à la logique aristotélicienne du tiers exclu. Nous retrouvons dans l'ouvrage *Changeons de voie* d'E. Morin (2020) un accent comparable dans la phrase : *« Il faut renoncer à la réduction de la connaissance et de l'action au calcul et il nous faut répudier la raison glacée obéissant inconditionnellement à la logique du tiers exclu* ». (p. 134)[6].

Dans l'article présent, il s'agit de la notion de temps et de sa mesure paradoxale dans le déroulement de l'arbre de vie. Nous puiserons l'inspiration dans l'ouvrage *Les arbres de l'évolution* de L. Nottale, J. Chanine et P. Grou (2000), les lignes suivantes : « *Le résultat le plus stimulant de cette recherche est sans aucun doute la suggestion de l'existence possible d'une structure mathématique fractale pour les arbres évolutifs qui semble s'appliquer à une partie au moins de l'arbre de vie lui-même; Il lui correspond une loi d'accélération temporelle log-périodique (c'est-à-dire périodique non pas en fonction du temps lui-même, mais du logarithme de son écart à une époque critique... ».*

Dans ces phrases, nous avons retrouvé plusieurs éléments plus ou moins familiers en écho avec différentes recherches antérieures : « *évolutifs », « structure mathématique fractale », « log-périodique », « logarithme ».* Mais, sur quelle loi prendre le premier appui ?

Pour cela, dans un premier temps, rejoignant Galilée (L'essayeur, 1623) qui affirme *« Le livre de la Nature est écrit en langage mathématique* », nous avons voulu examiner la relation fonctionnelle entre le temps réel représenté par l'âge du sujet et le temps perçu respectant l'observation coutumière. Pour ce faire, on devra introduire la monotonie entre ces deux variables qui expriment la remarque anthropologique : « le temps perçu paraît de plus en plus court quand l'âge avance ». La fonction à retenir devra donc être décroissante. De plus, elle devra être invariante lors de tout changement d'échelle.

Dans un second temps, on examinera un choix exprimant la décroissance du temps perçu pour un âge donné afin d'obtenir la durée ressentie pour une durée réelle.

---

[5] « Le temps ne se comporte pas comme il nous apparaît » (C. Rovelli, p.12).

[6] Au fil de la lecture d'un magazine, nous avons observé aussi, d'un article au suivant, la dualité entre les registres de décodage de leur contenu. Par exemple, si un premier article est écrit sur le ton de l'ironie provoque chez moi un décodage relativement durable sur le même ton, lors de la lecture des premières lignes de l'article suivant. Or celui-ci, en général, n'admet d'aucune façon le registre ironique du précédent. Ce transfert de décodage, ce biais cognitif, nous paraît ressembler à celui du temps. La conscience est trompée par l'inertie cérébrale. Au cours d'un entretien dans une émission télévisée (« La grande librairie ») le 10 mars 2021, Robert Badinter, fin pianiste jadis, a déclaré entendre « faux » les notes sorties d'un piano (juste).

Enfin, rassemblant les propriétés de la relation ressenti⇔réel, nous l'avons identifiée comme objet fractal. Le phénomène « fractal » apparaît alors comme objet de l'espace-temps dans lequel s'expriment deux variables de la même façon que nous l'avons étudié pour l'arbre de vie donné par les lois de l'évolution (Gras , Ndong, 2019). C'est donc cette relation métaphysique dans le monde de la géométrie fractale qui nous a conduits à apporter un certain et nouvel éclairage à un phénomène sensoriel bien connu mais peu formalisé jusqu'à sa lecture rationnelle[7].

2 Relation logarithmique entre temps perçu et temps réel

Nous nous inspirons ici d'un article de T. L. Freeman, intitulé *Why it's later than you think*, publié dans J. Irr. Res. en 1983 dont nous avons trouvé la partie principale en consultant le web à l´adresse[8]. Partant de l'expression connue et correspondant à un vécu bien partagé contenu dans la phrase : « plus je vieillis, plus le temps me paraît court », il semble aller dans le bon sens de faire l'hypothèse H_1 suivante, respectant les dilatations, avancée par Freeman sans motivation.

2.1 Un premier modèle

H_1 : les durées ressenties sont inversement proportionnelles aux durées réelles.

Notons, à un instant déterminé (une année donnée), x = âge réel, y(x) = durée ressentie d'une année et pour l'âge x. Pour satisfaire l'hypothèse H_1 pour laquelle la durée ressentie doit être plus faible que la durée réelle, nous devons choisir une fonction décroissante y, vérifiant en outre la propriété dite d'invariance d'échelle, c'est-à-dire préservant les propriétés quelle que soit la distance à laquelle on se place (Wikipedia) comme les poupées russes :

pour tout h réel : y(a/b) = y(ha/hb) (Formule : 1)**,**

Cette propriété homomorphique permet donc de respecter une **loi d'échelle** [9](« même structure de près ou de loin » écrirait B. Mandelbrot). De ce fait, la structure de l'antécédent est respectée dans son image et, en particulier sont respectés les ordres correspondants dans les deux ensembles de temps. Cette remarque est fondamentale car elle nous assure que les relations internes aux deux espaces sont en général bien conservées comme la relation d'antériorité ou de successivité. Mandelbrot dira alors que les phénomènes obéissent à une **invariance d'échelle** ou à une **autosimilarité**. C'est le cas recherché dans le modèle mathématique mais pas nécessairement dans le ressenti psychologique en raison des paramètres à prendre en considération. En fait nous recherchons une loi, comparable et de moindre portée, bien sûr, à la loi de relativité restreinte d'Einstein, entre deux mesures du temps.

Alors, pour respecter totalement H_1, on choisit une fonction simple, intégrable, sur intervalle privé de 0 et vérifiant la condition (1), mais d'autres fonctions peuvent la satisfaire ;

---

[7] « Mais ce ne sont pas les arguments rationnels qui nous motivent dans la vie. La raison sert à avoir des idées claires, à débusquer des erreurs » écrit C. Rovelli dans l'ouvrage *L'ordre du temps*.

[8] https://everything2.com/user/Professor+Pi/writeups/Why time appear stospeedup with age

[9] « le principe de relativité d'échelle postule que les lois de la nature doivent être valides quel que soit l'été d'échelle du système. Ce principe complète et élargit celui de Galilée, Poincaré et Einstein… ». Note p. 471 de l'ouvrage de Queiros cité en référence.

y est de la forme y(x) = k/x [10] (Formule : 2),

y défini par (2) vérifie (1), k étant un paramètre qui dépend du sujet examiné. Par exemple, si l'âge de ce sujet est 20 ans et que ce sujet ressent à cet âge la durée d'une année comme étant égale à 1, k sera choisi égal à 20. La durée ressentie pour x variable devient y(x) = 20/x en supposant k constant au fil de l'âge. Ainsi, une demi-année serait ressentie comme une année à 40 ans et un quart d'année ou un trimestre le serait à 80 ans. Ce qui nous semble vraisemblable[11].

Intégrons au fil du temps la durée ressentie y(x) de l'âge 1 à un âge réel donné X par rapport à la variable x sur l'intervalle [1, X] afin d'obtenir la durée ressentie pour une durée réelle X :

$$Y(X)= \int_1^X y(x)dx = \int_1^X \frac{k}{x}dx$$

Y= k ln (X) (Formule : 3)

Ainsi, les durées ressenties sont liées au temps réel par une **relation logarithmique**. Notons que Y, défini par (3) vérifie l'équation (1).

2.2 Un modèle raffiné

Freeman a introduit un modèle plus raffiné, susceptible de prendre mieux en compte les personnalités des sujets. En effet, nous savons, par l'usage, que chacun ne réagit pas de la même façon aux événements, à la nature elle-même. Introduisons le temps t représentant la durée d'une observation à partir de l'origine des temps qui, pour un sujet quelconque, serait égale à 0. Introduisons une deuxième hypothèse H_2.

H_2 L'âge réel d'un sujet est une fonction affine du temps qui passe et se présente sous la forme bt + c. La durée ressentie Y(t) à l'instant t pour une durée ressentie d'une année, lui est inversement proportionnel, s'exprime cette fois par **:**

$$\mathbf{Y(t) = \frac{k}{bt+c}\ .}$$

Les paramètres ont le sens suivant pour T.L. Freeman :

1) b indique, par exemple, le degré de vitalité du sujet ou l'existence d'un handicap
2) c indique le degré de précocité du sujet
3) k deviendra plus clair dans l'examen des conditions aux limites.

Intégrons de 0 à T où T est l'âge réel donné.

Alors $Y(T) = \int_0^T \frac{k}{bt+c}dt$ = d.ln (T+ε) - d.ln ε. en posant d = k/b et ε = c/b

[10] A Jacquard, dans (Jacquard, 2001) écrit cette hypothèse de la façon suivante : dS = dE/E où dS représente l'accroissement de la sensation et dE celui de l'excitation, relation qui est, dit-il, une propriété des logarithmes. Notre conscience compare chaque durée nouvellement écoulée à la totalité déjà vécue.

[11] A. Jacquard dans (Jacquard, 2001, p.148) suggère de prendre pour unité la durée de la gestation, qui elle-même est plus proche du rythme du développement. Ainsi à sa naissance l'enfant a pour mesure l'âge « un » dont le log ou le ln est 0. Il compare ce moment au big bang en écrivant : « *La métaphore du big bang pour décrire le point de départ est particulièrement pertinente car, tant que la fusion de l'ovule et du spermatozoïde n'avait pas eu lieu, cet individu n'avait pas plus d'existence que le cosmos tant que l'explosion primordiale ne s'était pas produite* ».

Si l'on fait l'hypothèse, pour ce sujet, qu'il dispose d'une espérance de vie de 90 ans (choix minimal personnel souhaité), il est possible de déterminer les paramètres d et ε. Par exemple, pour T=0 (naissance), on a Y(T)=0 qui est satisfait pour ε = 1. Si l'espérance de vie (90) est satisfaite jusqu'à la mort, on doit avoir pour T=90, Y(90) = 90 d'où d = 90/ln (91).

Si 90 ans est l'espérance de vie d'un sujet, la formule générale donnant le temps ressenti Y en fonction de l'âge réel X est :

Y(X)= 90 ln (X+1) / ln(91) (Formule : 4)

2.2.1 Exemple 1

Pour un enfant de 11 ans, le ressenti de la durée réelle d'une année est : 90/ln (91).ln (1+1) soit 90/ln(91). ln (1+1) # 14 ans. Pour un enfant de 11 ans, Y vaut 49.5 ans, ce qui signifie que son ressenti de temps vécu équivaut à la moitié de la durée de sa vie espérée environ 49.5 : 90= 0.55 soit 55%. C'est la durée de vie où les apprentissages ont été énormes : marcher, courir, faire du vélo, lire, apprendre une autre langue, un instrument de musique, faire des mathématiques, résoudre une équation, avoir de premiers sentiments amoureux, etc.

2.2.2 Exemple 2

À la transition 40 ans→42 ans, la formule générale conduit au calcul suivant : $\int_{40}^{42} \frac{k}{bt+c}$dt = d[ln (42+1) – ln (40+1)]= 0.95 # 1 an. La durée ressentie en 2 ans est moitié moindre à 40 ans que la durée réelle. A 65 ans, la même durée ressentie correspond à une durée réelle de 0.61 an, soit environ le tiers de la durée réelle. Quant au temps de vie ressenti il est égal à Y(65) soit 83.2 ans. Autre exemple, à 87 ans, une année est ressentie comme étant équivalente à ¼ année réelle.

2.3 Conclusion partielle

Cette situation est de nature fractale car il s'agit de mesurer le même objet (un fait historique) avec des échelles différentes, liées à l'observateur, comme nous l'avions vu avec le mesurage de la côte bretonne. Les faits, dans l'axe du temps, peuvent être vus aussi comme les poupées gigognes, en relation homothétique. « *La fractalié, en sciences, se définit par la persistance du motif à toutes les échelles* », rappelle l'astrophysicien philosophe, A. Barrau dans *Le plus grand défi de l'histoire de l'humanité* (2020, p.166).

La fractalité se retrouve aussi dans la correspondance algébrique logarithmique entre la durée réelle et la durée perçue. En conformité avec la métaphore précédente, dans ce transfert de structures, en l'occurrence, il y a conservation des relations d'ordre. Deux évènements différents seront perçus avec un décalage comparable par rapport à leur réalité.

En définitive, pour une même situation physique, et pour le même observateur, dans cette problématique de décalage entre temps réel et temps perçu, nous sommes en présence de deux points distincts de l'espace-temps au sens d'Einstein. Leur composante, sur l'axe du temps, étant différentes, ils appartiennent à deux plans différents, une loi logarithmique les reliant. On peut même enrichir l'intuition en imaginant une famille de plans engendrés par une famille d'interlocuteurs face à la même situation temporelle. Nous sommes en présence d'un exemple que Diego Queiros-Condé dans (Queiros-Condé, Chaline, Dubois, 2015) pense inobservable dans la vie courante et seulement dans des cas ultra relativistes

de particules à hautes énergie[12]. C'est pour cette raison que nous allons nous préoccuper des mêmes relations entretenues ci-dessus entre géométrie fractale et relativité.

3 Fractalité de l'arbre de vie

Nous avons montré, à travers de multiples travaux, certaines propriétés de l'ASI à la lumière de son étude dans le cadre de la géométrie fractale de B. Mandelbrot.

3.1 Rappel de travaux antérieurs sur l'évolution de l'arbre de vie cognitif

Dans *Dimension fractale du graphe implicatif* (Gras 2018), nous mettons en évidence la nature fractale du graphe implicatif en exhibant un invariant du graphe équivalent à une dimension comme il est coutumier en géométrie fractale. Dans *Arbre cohésitif et métaphore fractale* (Gras, Gregori, 2019), nous montrons de la même façon que l'arbre cohésitif possède également cette propriété qui se trouve ainsi associée à son rôle de révélateur de la construction cognitive de l'enfant. Enfin, dans *Évolution darwinienne et classification orientée* (Gras, Ndong, 2019), après avoir illustré et confirmé que l'arbre de vie, au sens biologique du terme, avait une structure fractale, nous désignons l'arbre cohésitif comme une des métaphores de l'arbre de vie. Revenons à la problématique de la relativité de la perception du temps.

3.2 Validation graphique de la propriété fractale des durées ressenties du temps qui passe.

On représente dans le plan, la relation entre X, représentant la durée réelle du temps pour passer de 20 à 100 ans et de la forme

Y= k ln (X) + h, (Formule : 5)

Soit en reprenant l'exemple 1 de la section 2.2 :

Y= 90 ln (X+1) / ln (91)

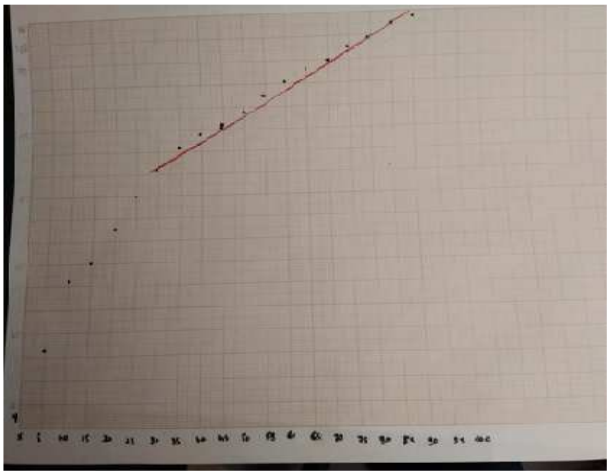


Figure 2 :Courbe de la relation entre âge réel (X) et âge perçu (Y)

[12] « Ces notions de relativité ont été appliquées à la physique, mais n'ont pas été prises en compte dans l'étude du phénomène particulièrement important de la nature qu'est l'évolution du monde vivant » Le monde des fractales, p.27.

On note, au-delà de 25 ans où la perception du temps est correcte, la quasi-linéarité de l'image de la relation (1), significative la quasi-parfaite corrélation entre l'âge réel et la perception anthropique du temps. Il existe une loi log-périodique entre ces deux variables. Cette propriété est une caractéristique de la nature **fractale** du phénomène et présente donc une propriété prédictive (Gras et *al*, 2017, 2018). La pente de la droite n'est autre que la dimension associée à cette propriété rappelant l'existence d'une loi d'échelle. En effet, du temps réel au temps ressenti la traduction ressemble fort à un **changement d'échelle sur l'axe du temps**. Celui-ci est de nature logarithmique ce qui confère à cette étude des propriétés prédictives Intuitivement encore, on peut imaginer deux canaux superposés dont l'un possède des graduations régulières et l'autre des graduations ayant subi un changement d'échelle. On peut même imaginer deux trajectoires distinctes, de même origine parcourant un plan de l'espace-temps.

Autant de traits qui nous confirment notre conjecture : la perception des durées temporelles par un sujet humain est un objet fractal indexé par la flèche du temps. Le modèle mathématique retenu, soutenu par les hypothèses 1 et 2, porte en lui-même certes la propriété de fractalité, encore fallait-il prouver cette adéquation.

En toute hypothèse, on peut penser, comme on l'étudie et le confirme dans « *La hiérarchie cohésitive, avatar fractal de théories du développement* » (2021), que cet objet est signe d'un développement, à savoir ici, l'affaiblissement continu de la perception anthropologique du temps. En outre, cette conclusion se renforce de l'évocation intuitive de l'image d'une cascade d'interlocuteurs 1 2 …n-1, n…chez lesquels un son ou/et un souvenir datés seraient successivement filtrés du n-1 ^{ème} réel au n^{ième} ressenti.

4 Conclusion

Cet article s'inscrit dans le cadre de nos travaux sur l'évolution sur l'évolution des êtres vivants et sur la géométrie fractale. Partis d'une situation familière, ressentie progressivement au cours de la vie, à savoir la différence entre le temps réel et le temps perçu par toute personne vivante, nous en avons formalisé une relation calculable. On a vu alors qu'elle était un objet de l'espace-temps de la théorie de relativité. Relativité que l'on retrouve également convoquée et corrigée entre le temps présent et le temps à venir. B. Pivot, évoquant et soutenant la procrastination, la sienne en particulier, la justifie, à 85 ans, dans « *Mais la vie continue* » (2020) en écrivant : « *À long terme, je n'ai plus beaucoup de temps ; à court terme, j'ai tout mon temps* ». Par le long terme, il évoque, semble-t-il, le temps qui dans le passé s'est écoulé lentement, s'éloignant toujours ou dans le futur s'écoulera non moins lentement alors que le temps immédiat, présent est plus vif, plus rapide. Et sa procrastination en vise la correction contre la perception. De plus, cette relation, entre temps réel et temps perçu, en adoptant une représentation logarithmique, nous a montré qu'elle possédait une propriété caractéristique des phénomènes fractals ainsi que leurs capacités prédictives.

Références

[1]. Barrau A. (2020) *Le plus grand défi de l'histoire de l'humanité*. Neuilly- sur-Seine : Michel Lafon.

[2]. Gras R., et *al.* (2017), Hiérarchie de règles en ASI et conceptualisation, in R. Gras, JC. Régnier, D. Lahanier-Reuter, C. Marinica, F. Guillet (Ed) *L'Analyse Statistique*

implicative, Des Sciences dures aux Sciences Humaines et Sociales, Toulouse : Cépaduès.

[3]. Gras R. (2018)*, La théorie de l'Analyse Statistique Implicative*, *ou l'invraisemblance du faux* Toulouse : Cépaduès.

[4]. Gras R. (2018). Dimension fractale d'un graphe implicatif, dans R. Gras, *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*, Toulouse : Cépaduès Éditions.

[5]. Gras R., Gregori P. (2019) Arbre cohésitif et métaphore fractale. In J-C. Régnier, R. Gras, M. Henry, R. Couturier, G. Brousseau (Eds), *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines*. Belfort : Université de Bourgogne Franche-Comté p.90-105

[6]. Gras R. Ndong L. (2019) Évolution darwinienne et classification orientée In J-C. Régnier, R. Gras, M. Henry, R. Couturier, G. Brousseau (Eds), *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines*. Belfort : Université de Bourgogne Franche-Comté p. 41-57.

[7]. Jacquard A. (2001) *La science à l'usage des non-scientifiques*. Paris : Calmann-Lévy.

[8]. Mandelbrot B. (1997) *Fractales, hasard et finance*. Paris : Flammarion

[9]. Morin E. (2020) *Changeons de voie*. Paris : Denoël.

[10]. Ndong L., Gras R. (2017), Classification des primates et Classification hiérarchique orientée de l'A.S.I., in R. Gras, JC. Régnier, D. Lahanier-Reuter, C. Marinica, F. Guillet (Ed) *L'Analyse Statistique implicative, Des Sciences dures aux Sciences Humaines et Sociales*, Toulouse : Cépaduès.p.453-458.

[11]. Pivot B. (2020) *Mais la vie continue*. Paris : Albin Michel.

[12]. Queiros-Condé D., Chaline J. et Dubois J. (2015) *Le monde des fractales.* Paris Ellipses

[13]. Rovelli C. (2017) *L'ordre du temps* Paris : Flammarion

METODOLOGÍA DE APLICACIÓN DEL ANÁLISIS ESTADÍSTICO IMPLICATIVO EN LAS INVESTIGACIONES MÉDICAS DE CAUSALIDAD

Nelsa María Sagaró Del Campo[1], Larisa Zamora Matamoros [2]

MÉTHODOLOGIE D´APPLICATION DE L´ANALYSE STATISTIQUE IMPLICATIVE DANS LES ENQUÊTES MÉDICALES DE CAUSALITÉ

METHODOLOGY OF APPLICATION OF IMPLICATIVE STATISTICAL ANALYSIS IN MEDICAL INVESTIGATIONS OF CAUSALITY

RESUMEN

El perfeccionamiento de las técnicas estadísticas es un problema de investigación que no pierde vigencia. Las investigaciones sobre factores de riesgo iniciadas a mediados del siglo pasado, las cuales sentaron las bases del modelo multicausal, se enmarcan en el enfoque lineal que ha predominado en las ciencias naturales. Sin embargo, el proceso salud-enfermedad es complejo, lo que conlleva a la necesidad de aplicar técnicas estadísticas renovadoras como es el caso del análisis estadístico implicativo creado para solucionar problemas didácticos de las matemáticas. Con el objetivo de diseñar una metodología para el empleo de esta nueva forma de análisis en las investigaciones médicas de causalidad, se llevó a cabo una investigación documental, donde se aplicaron métodos teóricos para determinar los fundamentos que la sustentan y definir las etapas y los procedimientos necesarios por cada etapa, así como las recomendaciones para su empleo. Luego se concibió la forma de analizar e interpretar los resultados. Además, se revisaron las investigaciones realizadas en la Universidad de Ciencias Médicas de Santiago de Cuba en colaboración con la Universidad de Oriente, conducentes a la formación de especialistas en Bioestadística, bajo la dirección de las autoras de la presente investigación. La metodología propuesta quedó constituida por nueve etapas: transformación de las variables, codificación de los datos, análisis exploratorio, análisis principal, presentación de los resultados, interpretación de los resultados, análisis a posteriori, selección de variables a incluir en el modelo de regresión logística binaria y discusión de los resultados. Cada una de estas etapas fue seleccionada teniendo en cuenta la literatura, el criterio de expertos y las regularidades y contradicciones demostradas en los estudios antes mencionados. Se reconoce entre las particularidades más importantes para la aplicación de este análisis: la duplicación de la variable dependiente y el empleo de las nuevas variables dependientes en el cono del grafo implicativo como la vía para el reconocimiento de los supuestos factores causales, ya sean pronósticos o de riesgo. En la interpretación se tienen en cuenta los índices, las tablas y los gráficos: árboles de similaridad y cohesitivo y el grafo implicativo, a partir de los cuales se identifican los posibles factores causales, las variables confusoras, las modificadoras del efecto y otras intervinientes. Se recomienda aplicar esta nueva perspectiva de análisis e interpretación como complemento de las técnicas multivariadas empleadas habitualmente en los estudios clínico-epidemiológicos para el estudio de la causalidad.

Palabras clave: *análisis estadístico implicativo, investigaciones médicas de causalidad, causalidad en salud*

RÉSUMÉ

L'amélioration des techniques statistiques est un problème de recherche permanent. Les recherches sur les facteurs de risque initiées au milieu du siècle dernier, qui ont jeté les bases du modèle multicausal, s'inscrivent dans le cadre de l'approche linéaire qui a prédominé dans les sciences naturelles. Cependant, le processus santé-maladie est complexe, ce qui entraîne la nécessité d'appliquer de nouvelles techniques statistiques, comme c'est le cas de l'analyse statistique implicative créée pour résoudre des problèmes didactiques en mathématiques. Dans

[1] Universidad de Ciencias Médicas, Calle E s/n e/ 6ta y Ave Americas, Sueño, nelsa@infomed.sld.cu
[2] Universidad de Oriente, Patricio Lumumba s/n Quintero, Larisa@uo.edu.cu

le but de concevoir une méthodologie pour l'utilisation de cette nouvelle forme d'analyse dans la recherche médicale de la causalité, une recherche documentaire a été réalisée, où des méthodes théoriques ont été appliquées pour déterminer les fondements qui la soutiennent et pour définir les étapes et les procédures nécessaires pour chaque étape, ainsi que les recommandations pour son utilisation. Ensuite, une méthode d'analyse et d'interprétation des résultats a été élaborée. En outre, les recherches menées à l'Université des sciences médicales de Santiago de Cuba en collaboration avec l'Université d'Oriente, menant à la formation de spécialistes en biostatistique, sous la direction des auteurs de cette recherche, ont été examinées. La méthodologie proposée était constituée de neuf étapes : transformation des variables, codification des données, analyse exploratoire, analyse principale, présentation des résultats, interprétation des résultats, analyse a posteriori, sélection des variables à inclure dans le modèle de régression logistique binaire et discussion des résultats. Chacune de ces étapes a été sélectionnée en tenant compte de la littérature, des critères des experts et des régularités et contradictions démontrées dans les études susmentionnées. Parmi les particularités les plus importantes pour l'application de cette analyse figurent la duplication de la variable dépendante et l'utilisation des nouvelles variables dépendantes dans le cône du graphique implicatif comme moyen de reconnaître les facteurs causaux supposés, qu'il s'agisse de facteurs de pronostic ou de risque. L'interprétation prend en compte les indices, les tableaux et les graphiques: arbres de similarité et de cohésion et le graphique implicatif, à partir desquels sont identifiés les éventuels facteurs de causalité, les variables de confusion, les modificateurs d'effet et les autres variables intervenantes. Il est recommandé d'appliquer cette nouvelle perspective d'analyse et d'interprétation en complément des techniques multivariées couramment utilisées dans les études clinico-épidémiologiques pour l'étude de la causalité.

***Mots clés**: analyse statistique implicative, recherche sur la causalité médicale, causalité dans la santé.*

ABSTRACT

The improvement of statistical techniques is an ongoing research problem. Research on risk factors initiated in the middle of the last century, which laid the foundations of the multicausal model, is framed within the linear approach that has predominated in the natural sciences. However, the health-disease process is complex, which leads to the need to apply new statistical techniques, such as the case of the statistical implicative analysis created to solve didactic problems in mathematics. In order to design a methodology for the use of this new form of analysis in medical research on causality, a documentary research was carried out, where theoretical methods were applied to determine the foundations that support it and to define the necessary stages and procedures for each stage, as well as the recommendations for its use. Then, a way of analyzing and interpreting the results was devised. In addition, the research carried out at the University of Medical Sciences of Santiago de Cuba in collaboration with the University of Oriente, leading to the training of specialists in Biostatistics, under the direction of the authors of this research, was reviewed. The proposed methodology consisted of nine stages: transformation of variables, data coding, exploratory analysis, main analysis, presentation of results, interpretation of results, a posteriori analysis, selection of variables to be included in the binary logistic regression model and discussion of results. Each of these stages was selected taking into account the literature, expert criteria and the regularities and contradictions demonstrated in the studies mentioned above. Among the most important particularities for the application of this analysis are the duplication of the dependent variable and the use of the new dependent variables in the cone of the implicative graph as the way to recognize the supposed causal factors, whether prognostic or risk factors. The interpretation takes into account the indexes, tables and graphs: similarity and cohesive trees and the implication graph, from which the possible causal factors, confounding variables, effect modifiers and other intervening variables are identified. It is recommended to apply this new perspective of analysis and interpretation as a complement to the multivariate techniques commonly used in clinical-epidemiological studies for the study of causality.

***Keywords**: implicative statistical analysis, medical causality research, causality in health.*

1 Introducción

El perfeccionamiento de las técnicas estadísticas es un problema de investigación que no pierde vigencia. La sociedad actual genera datos de manera exponencial por lo que se requieren nuevos métodos de análisis que permitan extraer el conocimiento subyacente en dichos datos y ajustarlos a modelos cada vez más cercanos al comportamiento real del proceso salud-enfermedad.

1.1 Estudio de la causalidad en medicina

En la actualidad se presentan diferentes enfoques de causalidad, lo cual obliga a los investigadores a estudiar el proceso salud-enfermedad a través de modelos complejos, dinámicos, jerárquicos que tienen en cuenta el papel creativo del desorden, de las inestabilidades, del azar, del caos, de las asimetrías (Bacallao, 2016; Ramis, 2004).

Sin embargo, las investigaciones sobre factores de riesgo iniciadas a mediados del siglo pasado, las cuales sentaron las bases del modelo multicausal, se enmarcan en el enfoque lineal que ha predominado en las ciencias naturales (Morin, 2004). Así el nexo causa-efecto se ha analizado como una relación lineal, no compleja, unívoca y dimensionable, en el cual la ocurrencia del efecto es una función de la causa.

Es por esta razón que en la actualidad el estudio de la causalidad se torna polémico y controvertido, pues las investigaciones médicas emplean técnicas estadísticas que parten del supuesto de que los problemas de salud son simples y lineales, lo cual conduce a un cierto grado de inexactitud y muchos de los modelos empleados tradicionalmente presentan contradicciones, al no lograr explicar satisfactoriamente determinados procesos. Todo lo cual conduce al empleo de técnicas renovadoras que no partan de supuestos de linealidad y estén acorde con los principios actuales de la investigación, como es el caso del análisis estadístico implicativo.

1.2 Análisis estadístico implicativo para el estudio de la causalidad en salud

El análisis estadístico implicativo, conocido por la sigla ASI de A*nalyse Statistique Implicative* del idioma francés donde se originó, es una herramienta de la minería de datos, basada en las técnicas estadísticas multivariadas, la inteligencia artificial y el álgebra booleana, para modelar la cuasi-implicación entre los sucesos y variables de un conjunto de datos (Sagaró y Zamora, 2019).

Esta técnica fue creada por el francés Régis Gras, (2008, 2013, 2014) profesor emérito de la Universidad de Nantes, Francia, quien comenzó sus trabajos en este campo en 1980, y desde entonces ha venido estudiando el fenómeno de la creación de reglas inductivas no simétricas y de la cuantificación de la probabilidad de que se presente la característica b si se ha observado la característica a en la población, tratando de modelar relaciones del tipo "si a entonces casi b".

El ASI se ha empleado con éxito en el diagnóstico y solución a problemas propios de la Didáctica de las Matemáticas, que fue su objetivo inicial, así como en otras áreas. Sagaró y Zamora (2019) identificaron los aspectos teóricos y prácticos que justifican el empleo de esta técnica en los estudios de causalidad en medicina. Sin embargo, para lograr su aplicación eficiente en los estudios de causalidad, propios de las ciencias médicas, se hizo necesario contextualizar su análisis e interpretación. Surgió entonces, la interrogante ¿cómo lograr una aplicación eficiente del ASI en las investigaciones médicas de causalidad?

Para dar solución a dicha interrogante en este trabajo se formuló como objetivo: diseñar una metodología de contextualización del ASI a las investigaciones médicas de causalidad.

2 Métodos

Se realizó una investigación aplicada cuya primera fase conllevó una investigación documental entre julio y diciembre de 2018, en la cual se diseñó un plan de la investigación; se realizó una amplia revisión bibliográfica donde se aplicaron métodos teóricos para analizar los fundamentos que sustentan la creación de la metodología que se propone, así como los procedimientos de cada una de sus etapas y las recomendaciones para su empleo adecuado; se elaboraron fichas resumen donde se identificó el estado de desarrollo alcanzado en torno al tema, se obtuvieron datos relevantes acerca de los enfoques teóricos y disciplinares dados a las técnicas estadísticas, de las tendencias y de las perspectivas metodológicas.

Luego, se evidenciaron los hechos que hacen posible el empleo del ASI en las ciencias biomédicas, para lo cual se realizó una comparativa entre esta técnica y la regresión logística binaria, como la mejor propuesta para la identificación de factores pronósticos y de riesgo, donde se tuvieron en cuenta 14 aspectos (Sagaró y Zamora, 2019).

Se revisaron las investigaciones realizadas en la Universidad de Ciencias Médicas de Santiago de Cuba en colaboración con la Universidad de Oriente, en las cuales se aplicó el ASI, conducentes a la formación de especialistas en Bioestadística bajo la dirección de las autoras de la presente investigación (García, 2015; Moraga, 2016, 2017, 2018; Paez, 2018; Galano, 2019; Pardo, 2019). Esta revisión reveló las regularidades que permitieron definir los procedimientos necesarios para un mejor análisis de los datos e interpretación de los resultados en las investigaciones médicas de causalidad.

Para la identificación de las posibles etapas y sus procedimientos se consultaron otras metodologías, se tuvo en cuenta lo planteado en la literatura, el criterio de expertos, y se elaboró un esquema resumen de la misma. Se seleccionaron los autores destacados en la temática, cuyos trabajos se tomaron como referentes para las propuestas de las interpretaciones de los resultados.

3 Metodología ASI-IMC

Se partió de considerar que la aplicación del ASI requiere de dos conjuntos finitos, uno formado por los individuos o unidades de análisis y otro por las variables del estudio, que conforman la matriz de datos a analizar. Las variables pueden ser medidas en cualquier escala, pero en este primer trabajo solo se tuvieron en cuenta las dicotómicas (presencia o ausencia de la característica). Para su aplicación se empleó como procesador estadístico el CHIC (del francés: *Classification Hiérarchique Implicative et Cohésitive* que significa Clasificación Jerárquica, Implicativa y Cohesitiva), el cual genera los índices y gráficos propios de este análisis, por lo que las modificaciones y recomendaciones propuestas se basan en los resultados generados por este procesador.

En correspondencia con los diseños metodológicos empleados en las ciencias médicas para el estudio de la causalidad, mencionados en Sagaró y Zamora (2019b), se propone el empleo del ASI en estudios analíticos, cuya finalidad es la identificación de los posibles factores causales de un efecto o desenlace dicotómico y que sigan un diseño de tipo caso y

control, ya sea en su forma clásica retrospectiva o en alguna variante prospectiva como el anidado en una cohorte, o estudios de cohorte clásicos (dos cohortes prospectivas).

3.1 Diseño de la metodología

El diseño de la metodología tuvo en cuenta que en el ASI: 1) no existe una división de las variables en dependientes e independientes, 2) el algoritmo empleado en el procesamiento de los datos solo analiza la variable codificada con '1', 3) es compleja la interpretación de las implicaciones entre variables más que dicotómicas, 4) no está establecida la manera de identificar los posibles factores causales de un desenlace, ni las posibles variables confusoras, intermedias o colisionadoras, 5) no existen las categorías que permitan clasificar los factores identificados de acuerdo con la influencia que ejercen en la aparición del desenlace.

Para la contextualización del ASI a las investigaciones médicas de causalidad, se concibieron una serie de procedimientos desde la preparación de los datos, la forma de análisis y la interpretación de los resultados, los cuales se explican a continuación.

3.1.1 Procedimientos para la preparación de los datos

La primera modificación que deben tener los datos está dada por la división de las variables en dependientes e independientes. La variable dependiente, de respuesta o desenlace, que tradicionalmente es única con dos categorías, se duplicará contando para el análisis con dos variables dependientes binarias que representan el desenlace peor y mejor, respectivamente.

Por ejemplo, en los estudios para la identificación de factores de riesgo se crearán dos variables, "Enfermo" y "Sano", y para la identificación de factores pronóstico una variable para el desenlace favorable y otra para el desfavorable, que según el tipo de desenlace que se escoja podría ser "Fallecido" y "Vivo" o "Complicado" y "No complicado", etc. Cada variable creada se codificará con el valor 1 si el individuo analizado posee la característica de interés y con el 0 en caso contrario.

Esta duplicación se sustenta por el hecho de que el algoritmo empleado en el procesamiento de estos datos sólo analiza la variable codificada con 1, por ejemplo, en el caso de la identificación de factores pronósticos de mortalidad si se declarase una sola variable "estado", donde se representase el fallecido con 1 y el vivo con 0, nunca sería posible analizar las relaciones de causalidad asociadas al estado vivo, es decir identificar los factores de buen pronóstico que pudieran existir.

Otra modificación es la dicotomización de las variables independientes que sean politómicas, para ganar en eficiencia, aun cuando parezca que se pierde en información. Este proceso permitirá visualizar mejor el cambio de una categoría a otra y facilitará su interpretación, se puede realizar de dos formas: creando tantas variables nuevas como categorías posean, o creando dos variables, en las cuales se agrupen, siempre que sea posible, varias categorías. En el caso de las covariables cuantitativas, estas se transformarán en dicotómicas, considerando la de peor y mejor desenlace. Este proceso de dicotomizar consiste en buscar un punto de corte apropiado en el recorrido de la variable, para lo cual es posible emplear diferentes estrategias basadas, sobre todo, en la literatura y la opinión de expertos que conlleven a una coherencia de la relevancia clínica con la estadística, según se explica en Sagaró y Zamora (2019c).

Por último, se guardará la base de datos en formato CSV de Excel, requisito indispensable para ser reconocida por el procesador automatizado CHIC.

3.1.2 Procedimientos para el procesamiento de los datos

El procesamiento de los datos se realizará de dos formas, en dependencia de cómo se visualizan las variables de respuesta creadas: incluyéndolas en el análisis como variables principales, al igual que todos los posibles factores causales o considerándolas como variables suplementarias para reconocer y cuantificar su contribución a la formación de cada una de las implicaciones o relaciones causales entre el resto de las variables.

Primeramente, se calcularán las frecuencias absolutas y relativas para determinar las características de los grupos en estudio (casos y controles). Esta debe concluir con el cálculo y presentación de las frecuencias de cada variable independiente en los dos grupos de estudio.

Luego se hará el análisis clasificatorio, cuyo propósito es agrupar aquellas variables que presentan características similares, revelando las agrupaciones naturales dentro de la colección de datos. La estrategia de agrupación se basará en medir la distancia entre pares de variables, formando clústeres con aquellas menos distantes según el índice de Lerman, lo cual es facilitado por el CHIC. Se obtendrá y analizará el árbol de similaridad, así como los índices de similaridad entre las variables con comportamiento semejante en los grupos de casos y controles.

En el árbol de similaridad se apreciarán los diferentes grupos o clases que se forman en torno a las dos variables dependientes (generalmente, los casos y los controles) según los factores que más se asemejen en su comportamiento a cada una de ellas, así como los subgrupos con mayor o menor grado de similaridad dentro de estos grandes grupos. Este análisis es útil también para identificar relaciones entre variables no consideradas al inicio del estudio.

En el análisis implicativo y para un cierto nivel de confianza, elegido por el investigador, se determinarán las reglas admisibles a partir del valor de su intensidad implicativa. Se obtendrán y analizarán los gráficos que brinda el sistema.

Se sugiere obtener el grafo implicativo normal, que muestra las relaciones causales entre el conjunto de variables, donde se debe analizar la formación de caminos en los grafos o subgrafos que se formen, acompañado por la intensidad implicativa en el sentido clásico o entrópico según la opción elegida.

Además, es imprescindible obtener dos grafos implicativos en modo cono, uno para las relaciones causales de los factores que influyeron en el desenlace peor, tomando como cono la variable dependiente de peor desenlace, y otro para los que influyeron en el mejor desenlace, tomado como cono la variable que representa el mejor desenlace.

En cuanto a las relaciones implicativas que se desean visualizar en el grafo implicativo, estas serán la que determine el investigador a partir de fijar valores umbrales para el nivel de confiabilidad. El software permite elegir hasta cuatro umbrales diferentes de confianza a la vez. Según sea el valor mínimo de los umbrales elegidos, podrá identificar una mayor o menor cantidad de factores asociados con las variables de respuesta. Se sugiere emplear los dos enfoques que brinda el software, clásico y entrópico, para reafirmar los hallazgos y emplear los cuatro niveles mayores que brinda el sistema, de manera tal que aparezcan enlaces con cuatro colores en el gráfico, uno para cada umbral, excepto en los casos que con gran intensidad implicativa, del 100 o 99 %, aparezcan casi todas las variables representadas en el grafo.

En el análisis cohesitivo se obtendrá y analizará el árbol cohesitivo, un gráfico jerárquico orientado de manera ascendente, siguiendo el índice decreciente de las

cohesiones que brinda el sistema, formado por las reglas y metarreglas donde se destacan los nodos significativos, acompañado por los índices de cohesión en el sentido del análisis implicativo.

Por cada tipo de análisis, el CHIC calcula y muestra resultados numéricos como: la frecuencia de ocurrencia de cada variable, su media y desviación estándar, la frecuencia de ocurrencia de cada uno de los pares de variables que se pueden formar y el coeficiente de correlación para cada par de variables.

La contribución y tipicalidad de cada individuo para la clase de variables elegida (un clúster de variables en el análisis de similaridad o una regla o camino de reglas en el análisis cohesitivo) deben ser calculadas, así como la lista de los riesgos asociados a la contribución y a la tipicalidad de cada variable suplementaria (de existir), junto con la variable más típica y más contributiva. Uno de los procedimientos propuestos en el análisis es la implementación de una segunda forma de análisis, denominado a posteriori, para que una vez identificados los sujetos responsables de la formación de una regla que motiva el interés del investigador, defina nuevas variables a partir de ciertas características que tengan en común estos sujetos y que no se tuvieron en cuenta en el primer análisis, las cuales podrían ayudar a esclarecer las relaciones encontradas. Por último, se debe realizar la selección de las variables a incluir en el modelo de regresión logística binaria a partir de los resultados de la aplicación de la metodología. Se sugiere que las primeras variables seleccionadas sean las que se enlazan con el cono de los dos grafos implicativos en modo cono, comenzando por las de mayor intensidad implicativa y las de enlace continuo, que son las que muestran mayor intensidad de asociación y relación directa, respectivamente.

3.1.3 Procedimientos para la presentación de los resultados

En este apartado se muestran los procedimientos para la presentación de los resultados, los cuales se auxilian, para su ilustración, de las salidas obtenidas en varias investigaciones realizadas en la Universidad de Ciencias Médicas de Santiago de Cuba, en las cuales se estudiaron diferentes tipos de canceres. Los resultados deben ser presentados en forma tabular y gráfica. La primera tabla a presentar debe ser la que emerge de la descripción de la muestra de estudio, donde se mostrarán las frecuencias de cada variable en ambos grupos de estudio (Figura 1).

Factor de riesgo	Casos		Controles	
	No	%	No	%
Edad (EDAD)	49	55,06	40	44,94
Raza blanca (PRB)	18	46,15	21	53,85
Raza mestiza (PRM)	23	74,19	8	25,81
Raza negra (PRN)	9	30,00	21	70,00
Antecedentes genéticos (PAG)	31	91,18	3	8,82
Antecedentes patológicos personales (PAP)	12	63,16	7	36,84
Factores endocrinos endógenos (PEEn)	8	50,00	8	50,00
Factores endocrinos exógenos (PEEx)	1	100,00	0	0,00
Consumo de bebidas alcohólicas (BEBE)	27	56,25	21	43,75
Hábito de fumar (FUMA)	32	47,76	35	52,24
Vida sedentaria (PVS)	38	53,52	33	46,48
Dieta inadecuada (PDI)	21	70,00	9	30,00
Radiación ionizante (PRI)	2	66,67	1	33,33
Nivel socioeconómico bajo (PS1)	5	45,45	6	54,55
Nivel socioeconómico medio (PS2)	34	46,58	39	53,42
Nivel socioeconómico alto (PS3)	11	68,75	5	31,25

Fuente: García (2015)

Figura 1 – Tabla de casos y controles según factor de riesgo

Otra de las tablas debe crearse a partir del análisis de similaridad, donde se presentarán las clases o aglomeraciones de variables formadas en orden decreciente del índice de similaridad, el valor del índice de similaridad y si la clase forma parte de una clase mayor que contiene al peor desenlace o al mejor, como se muestra en la figura 2, en cuyo ejemplo los desenlacen son "Vivo" y "Fallecido". Si no fuera necesaria la exhaustividad en esta tabla, se pueden mostrar solo las clases de mayor similaridad.

Variables similares	Índice de similaridad	Clase a la que pertenece
(META PVASCULAR)	1	Fallecido
(ETAPAC (META PVASCULAR))	0.999999	Fallecido
((ETAPAC (META PVASCULAR)) INFCAPSULAR)	0.999995	Fallecido
(TTUMORAL QUIMIO)	0.999966	Fallecido
((TTUMORAL QUIMIO) GRADO3)	0.99987	Fallecido
(((ETAPAC (META PVASCULAR)) INFCAPSULAR) ((TTUMORAL QUIMIO) GRADO3))	0.999792	Fallecido
((((ETAPAC (META PVASCULAR)) INFCAPSULAR) ((TTUMORAL QUIMIO) GRADO3)) HISTO)	0.998306	Fallecido
(((((ETAPAC (META PVASCULAR)) INFCAPSULAR) ((TTUMORAL QUIMIO) GRADO3)) HISTO) RADIO)	0.989997	Fallecido
(FALLECIDO HEMOG)	0.956826	Fallecido
(VIVO GRADO1)	0.941407	Vivo
((FALLECIDO HEMOG) (((((ETAPAC (META PVASCULAR)) INFCAPSULAR) ((TTUMORAL QUIMIO) GRADO3)) HISTO) RADIO))	0.897747	Fallecido

Fuente: Tabla 3 del Trabajo de Terminación de Especialidad de Bioestadística de Galano Vázquez K. Universidad de Ciencias Médicas de Santiago de Cuba (2018)

Figura 2 – Tabla de los factores pronósticos con mayor nivel de similaridad

Además, se puede presentar la tabla que contiene los índices de cohesión para las reglas o metarreglas formadas. También se mostrarán los tres gráficos que brinda el sistema: el árbol de similaridad, el grafo implicativo y el árbol cohesitivo y en el caso del grafo implicativo se obtendrá uno general y dos en modo cono como se explicó anteriormente.

3.1.4 Procedimientos para la interpretación de los resultados

Se interpretarán las frecuencias absolutas y relativas más relevantes en cada grupo de estudio, destacando las características que predominen y los indicadores estimados para cada variable como sus medias, desviaciones estándares, frecuencias de ocurrencias de cada par de variables, coeficientes de correlación, los índices de similaridad, de cohesión implicativa y de implicación inclusión, las tipicalidades y contribuciones de los individuos.

En el árbol de similaridad aparecerán agrupados aquellos factores que reúnan características similares, atendiendo a la medida de similaridad definida. Asociadas a estos grupos pueden estar o no las variables dependientes definidas, de forma tal que:

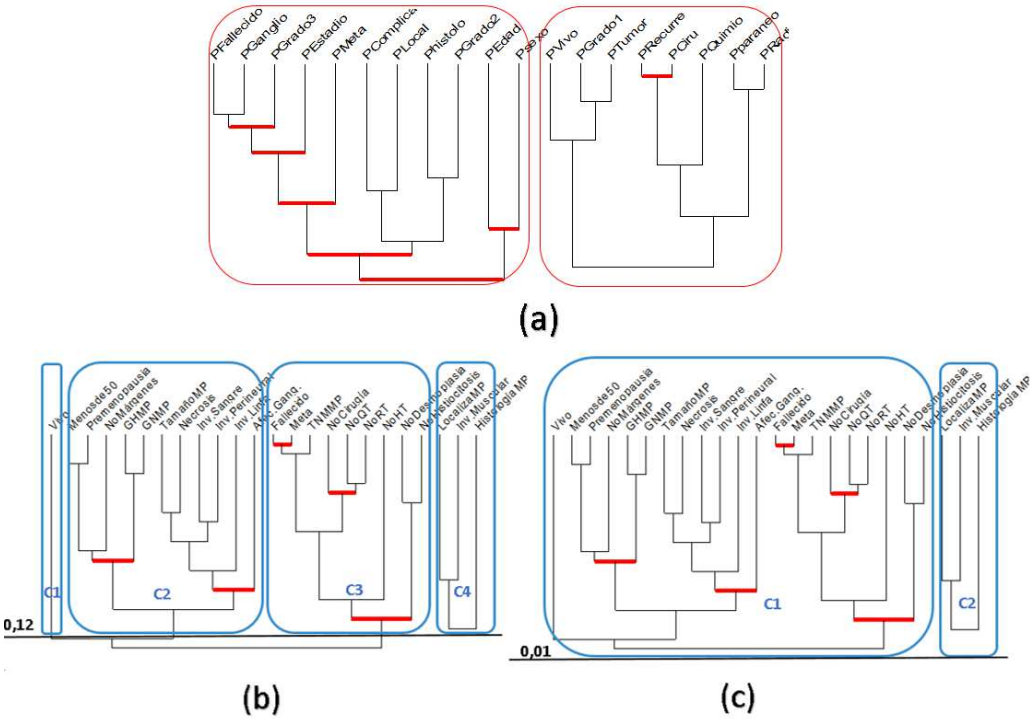
1. Si la variable dependiente "Vivo" ("Fallecido") aparece como miembro de un conglomerado de factores, estos factores se interpretan como que son los predominantes para la condición de vivo (fallecido).
2. Si las variables dependientes no resultan ser miembros de ningún conglomerado de factores, esta situación se interpretaría en el sentido de que no existe similitud entre ellos.

A modo de ejemplo, en la figura 3(a) se muestra el árbol de similaridad del estudio para la identificación de factores pronósticos de mortalidad del cáncer de pulmón (Moraga, 2016), donde se aprecia la formación de dos grandes grupos, uno que incluye la variable

PFallecido y un grupo de factores como la metástasis (PMeta), las complicaciones (PComplica), la edad avanzada (PEdad), el estadio avanzado (PEstadio), entre otros, interpretándose como que estos factores predominan entre los fallecidos. Del segundo grupo se puede concluir que el grado menor de diferenciación del tumor (PGrado1), el mayor tamaño tumoral (PTumor), los tratamientos con cirugía (PCiru), la radioterapia (PRadio) y la quimioterapia (PQuimio), fueron los factores más frecuentes entre los vivos (PVivo).

La formación de los grupos o conglomerados dependerá del nivel escogido por el investigador como significativo para poder considerar la existencia de similitud entre sus miembros, así por ejemplo, Pardo (2019) en su estudio consideró un índice de similaridad de 0,12, como se aprecia en la figura 3(b), donde se forman cuatro clases, quedando el vivo no asociado a ningún factor. Si se hubiera considerado el nivel mínimo, se tendrían dos clases donde se unen vivos y fallecidos, lo cual se puede apreciar en la figura 3(c).

En el grafo implicativo normal se debe reconocer el conjunto de relaciones multivariadas y determinar la intensidad de estas.



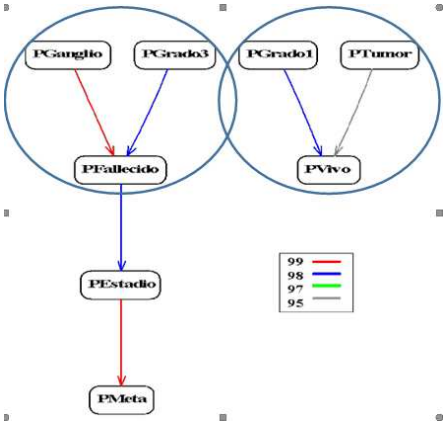
Fuentes: (a): Moraga (2016), (b) y (c): Pardo (2019)
Figura 3 - Árboles de similaridad

Los dos grafos implicativos en modo cono permitirán identificar los posibles factores causales como los factores de riesgo o de mal pronóstico y los factores protectores o de buen pronóstico, las relaciones mutuas y las posibles variables confusoras, intervinientes y colisionadoras.

La interpretación se basó en la similitud estructural del grafo implicativo con un grafo acíclico orientado conocido como DAG del inglés *Directed Acyclic Graph*, diagrama causal que plantea posibles rutas de asociación entre causas y efecto, así como otras rutas alternativas generadoras de sesgos, por lo que facilitan también el control de estos. Además, guían al investigador en cuanto a cuáles variables deben incluirse en el análisis (Cortes, 2016; Brewer, 2017; Greenland, 2017; Williamson, 2014).

3.1.5 Reconocimiento del conjunto de relaciones multivariadas

En el grafo implicativo normal es posible observar las relaciones que se establecen entre todo el conjunto de variables definidas, así como la intensidad y el sentido de cada implicación. En el grafo implicativo de la figura 4, perteneciente al estudio de Moraga (2016), antes mencionado, se aprecia que un grado histológico I (PGrado1) implicó estar vivo, con una intensidad implicativa igual o superior al 98 %, aunque inferior al 99 %. Un grado histológico III (PGrado3) y la presencia de ganglios (PGanglio) implicaron haber fallecido, el primero con una intensidad implicativa igual o superior al 98 % e inferior al 99 % y el segundo igual o superior al 99 %.



Fuente: Moraga (2016)
Figura 4 – Grafo implicativo

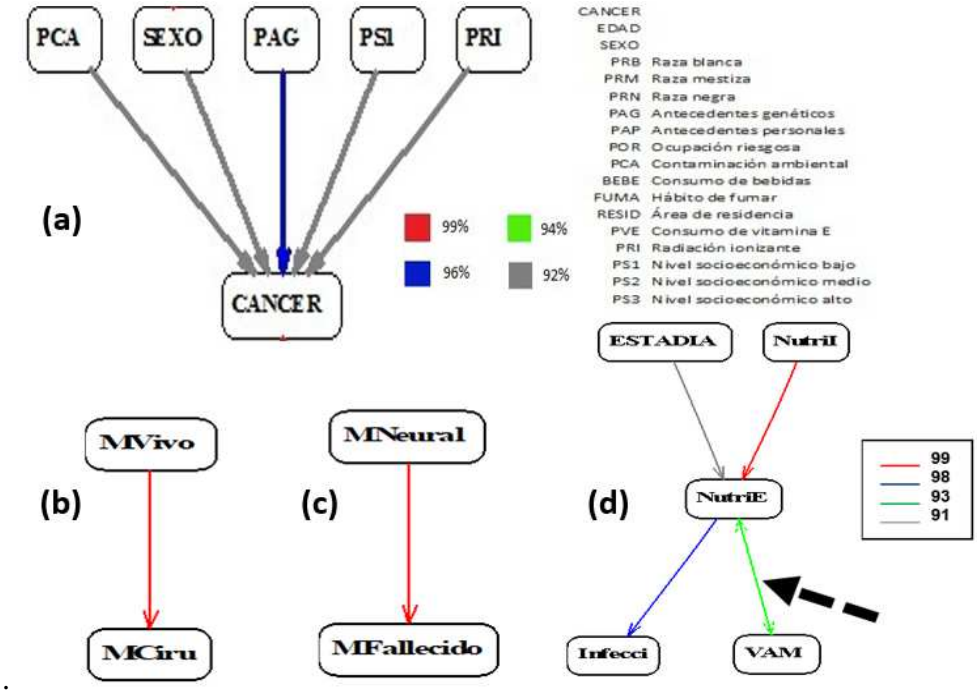
3.1.6 Identificación de los posibles factores causales

La identificación de los posibles factores causales se realiza a partir del modo cono del grafo implicativo, de manera que en un estudio de factores de riesgo, el modo cono permite identificar los factores de riesgo cuando en el cono del gráfico se coloca la variable "Enfermo" y los factores protectores cuando se sitúa la variable "Sano", mientras que en un estudio de factores pronósticos se identifican los factores de buen pronóstico cuando en el cono del gráfico se ubica el evento final favorable, por ejemplo "Vivo", y los factores de mal pronóstico cuando se ubica el evento final desfavorable, por ejemplo "Fallecido". La variable dependiente puede tomar otra denominación acorde con el desenlace esperado en el estudio. Esta modificación es reconocida por las autoras como el nexo fundamental entre este análisis y la causalidad en las ciencias médicas y se ha ejemplificado en la figura 5 empleando resultados de los estudios realizados por las autoras.

En la figura 5 (a) el cono está representado por la variable que identifica tener cáncer (CANCER), y las variables que aparecen unidas a él por flechas se consideran los factores de riesgo asociados al cáncer de pulmón: la contaminación ambiental (PCA), el sexo (SEXO), los antecedentes genéticos (PAG), el nivel socioeconómico bajo (PS1) y la radiación ionizante (PRI).

La figura 5 (b) permite identificar que recibir tratamiento quirúrgico (MCiru) es un factor de buen pronóstico porque se asocia con el evento favorable, estar vivo (MVivo) y la 5 (c) que la invasión perineural (MNeural) es un factor de mal pronóstico porque se asocia con el evento desfavorable fallecer (MFallecido).

La figura 5 (d) indica una doble implicación entre la ventilación aérea mecánica (VAM) y la malnutrición al egreso (NutriE), o sea existe una relación recíproca entre estas variables. Esto también puede ser interpretado como la existencia de un factor confusor.



Fuentes: (a): García (2015), (b) y (c): Moraga (2016) y (d): Paez (2018)
Figura 5 - Grafos implicativos en modo cono. Ejemplos de identificación de los posibles factores causales

3.1.7 Identificación de posibles variables confusoras

La identificación de posibles variables confusoras se propone por tres vías:

1. la presencia de dos nodos adyacentes unidos por un enlace bidireccional en el grafo implicativo, lo cual se fundamenta en la teoría de grafos, según la cual un enlace bidireccional se debe a la existencia de una variable adicional (variable confusora) con enlaces dirigidos hacia esos nodos (Prado, 2012). Para corroborar la existencia de una variable confusora se debe incorporar la supuesta variable confusora en el análisis y si en el nuevo grafo implicativo se mantiene el enlace bidireccional entonces hay causación recíproca (Reyna, 1971).
2. un cambio significativo (superior al 10 %) en la intensidad implicativa en el análisis al incluir o no la supuesta variable confusora, tomando como referente lo que plantean Manterola (2015), Requena (2017) y Bacallao (2012), quienes plantean que es posible detectar una variable confusora durante el análisis si se obtiene una diferencia relativa en el efecto mayor del 10 % mediante el análisis de regresión logística.
3. la obtención de reglas absurdas (sin sentido desde el punto de vista médico) visualizadas en el árbol cohesitivo.

3.1.8 Identificación de posibles variables colisionadoras

La identificación de posibles variables colisionadoras (sesgo de selección) en el grafo implicativo, tanto en el normal como en el modo cono, es posible cuando se observa una estructura en “V” (dos nodos con flechas apuntando a un mismo nodo).

3.1.9 Identificación de la interacción

Existe interacción o modificación del efecto cuando la asociación entre exposición y efecto varía según las categorías de una tercera variable, la variable modificadora de efecto. Tomando como referentes a Hernán y Robins (2020), autores que consideran útil el modelo de Rothman para descubrir la interacción, al considerar que dos causas componentes de una causa suficiente única tienen una interacción mutua y el grado de interacción depende de la prevalencia de los otros factores que la integren; se ha considerado la posible interacción entre las denominadas causas cuasi-componentes, lo cual debe ser confirmado, en cada estudio, por el investigador experto, que en ese caso incluirá el término de interacción en el modelo de regresión logística que se haga posteriormente. Estas variables se describen e interpretan sin necesidad de controlarlas como ocurre con las variables confusoras, ya que las variables modificadoras de efecto reflejan una característica de la relación entre exposición y efecto cuya intensidad depende de esta variable. No tiene sentido calcular una medida ajustada, ya que no sería representativa de la asociación global entre exposición y efecto y tampoco tendría sentido estratificar y hacer una media entre los estratos.

3.1.10 Clasificación de las causas

Otra de las propuestas a considerar es la creación de categorías que permiten clasificar los factores identificados en cuatro tipos de causas, tomando como referentes el modelo de Rothman (2005) y la teoría cuasiimplicativa. Estos tipos son las causas: cuasicomponentes, cuasisuficiente, cuasinecesaria y cuasiequivalente (cuasisuficiente y cuasinecesaria a la vez). Las mismas se identifican a partir de las reglas obtenidas del árbol cohesitivo (Figura 6).

1. **Causas cuasicomponentes:** No existe un único factor que pueda constituirse como causa cuasicomponente de un desenlace (D), debido a ello se define como causas cuasicomponentes a los factores que forman parte de una regla (R) o metarregla (R´) entre factores, que se relacionen directamente con el desenlace, estando este último cuasiimplicado por R o R´ (R´ ⇒ D o R ⇒ D), donde: R puede ser del tipo: F´ ⇒ F, F´ ⇐ F o F´ ⇔ F; y R´ puede ser del tipo: F´ ⇒ R, F´ ⇐ R o F´ ⇔ R, siendo F y F´ las causas cuasicomponentes.

En la figura 6 (a) una causa cuasicomponente es la regla R_2: (PrMeta ⇒ Ganglio) es del tipo F´ ⇒ F, lo cual significa que tener metástasis implicó con cierto índice de cohesión tener afectación ganglionar, por lo que se puede decir que la presencia de metástasis y afectación ganglionar son causas cuasicomponentes de mal pronóstico en el cáncer de próstata. También son causas cuasicomponentes los factores presentes en R_1 (del tipo F´ ⇔ F), R_3 (R_2 ⇒ R_1) y R_4 (F´ ⇒ R).

2. **Causa cuasisuficiente:** La causa cuasisuficiente es el conjunto formado por las causas cuasicomponentes. En el caso particular de que exista un factor que implique directamente al desenlace (F´⇒ D), este constituirá una causa cuasisuficiente, situación que se presenta muy pocas veces. Entre las causas cuasicomponentes que forman una causa cuasisuficiente existe una interacción biológica que puede depender de otras causas cuasicomponentes, de manera que el hallazgo de una regla de este tipo será el punto de partida de otro análisis para probar la hipótesis de interacción encontrada en dicha regla.

Una causa cuasisuficiente es la regla R8: (PrCiru ⇒ PrVivo) de la figura 6 (b), la cual se interpreta de la siguiente forma: ‘casi todos los individuos operados siguieron con vida’,

con un índice de cohesión de 0.836, ‘operarse es cuasisuficiente para estar vivo’. Esta interpretación puede también ser expresada de otras formas: ‘la cirugía predispone al paciente al evento favorable’ o ‘la casi totalidad de los pacientes que se operan se mantienen vivos’.

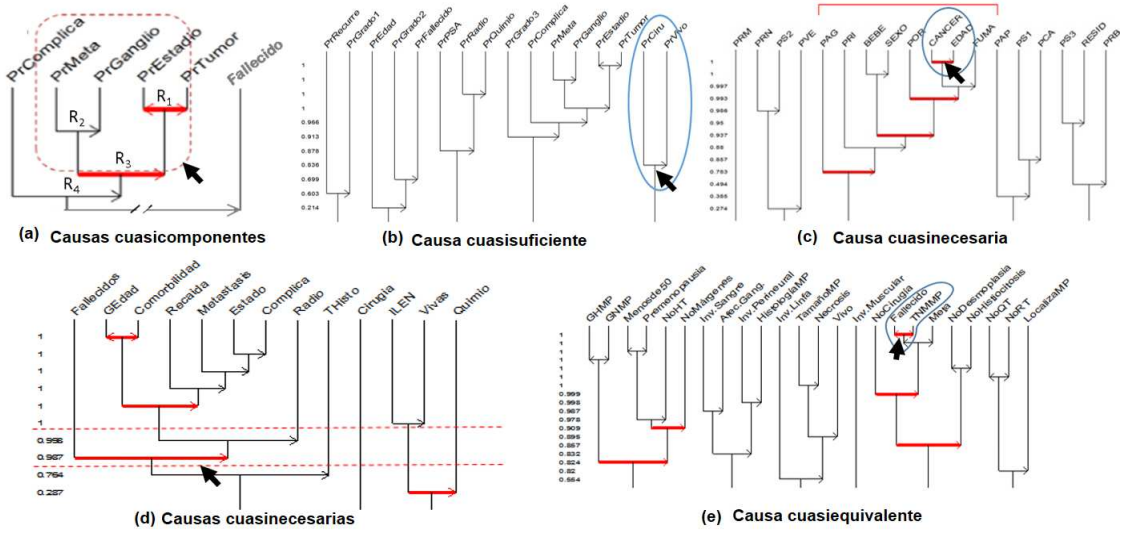
3. **Causa cuasinecesaria:** Se dirá que el factor F´ es una causa cuasinecesaria cuando se relaciona con el desenlace y está cuasi implicado por este (D ⇒ F´). También se denomina causa cuasinecesaria al conjunto formado por todos los factores que integran reglas o meta-reglas que estén relacionadas con el desenlace, y estén cuasi implicadas por este.

Una causa cuasinecesaria es la regla R_1: (CANCER ⇒ EDAD) de la figura 6 (c), la cual se interpreta: ‘tener cáncer implicó siempre (con un índice de cohesión de 1) tener una edad avanzada’, la edad avanzada es una causa cuasinecesaria para tener cáncer de pulmón, o sea, que deben estar presentes otros factores coincidiendo en el individuo para que este desarrolle cáncer de pulmón.

En la figura 6 (d) se muestra un conjunto de causas cuasinecesarias, en el cual el desenlace (Fallecidos) cuasi implica una metarregla R´, esto es, Fallecidos ⇒ R´, donde: R´: (R3 ⇒ radio), R3: (R1 ⇒ R2), R1: (GEdad ⇒ Comorbilidad), R2: (Recaida ⇒ (Metastasis ⇒ (Estado ⇒ Complica))).

4. **Causa cuasiequivalente:** Se dirá que el factor F´ es una causa cuasiequivalente si es cuasinecesaria y cuasisuficiente a la vez, o sea, si este factor cuasiimplica y está cuasiimplicado por el desenlace (D ⇔ F´).

Una causa cuasi equivalente es la regla R_1: (Fallecido ⇔ TNMMP) de la figura 6 (e) que se interpreta: ‘fallecer implicó siempre tener un estadio avanzado de la enfermedad (TNMMP), y tener un estadio avanzado de la enfermedad implicó siempre fallecer (con un índice de cohesión igual a 1)’.



Fuentes: (a): Modificado de Moraga (2018), (b): Moraga (2018), (c): García (2015), (d): Lambert (2021), (e): Pardo (2019)

Figura 6 - Árboles cohesitivos. Ejemplos de la clasificación de las causas.

La interpretación de los resultados que se proponen en este trabajo se basó en los criterios, que hasta el momento las autoras consideraron más plausibles, todo lo cual pudiera cambiar en un futuro a la luz de nuevos conocimientos y la experiencia de otras investigaciones.

3.2 Estructura de la metodología

Los procedimientos descritos anteriormente se integraron en una secuencia de etapas junto con otros procedimientos propios del método estadístico para conformar la metodología denominada ASI-IMC, acrónimo de Análisis Estadístico Implicativo en las Investigaciones Médicas de Causalidad.

La misma quedó compuesta por nueve etapas, según su finalidad, algunas de ellas con subetapas, y dentro de cada etapa o subetapa, los procedimientos a desarrollar para lograr el objetivo. Cada etapa sucede a la anterior. Al llegar a la etapa 6 si alguna regla despierta el interés del investigador pasa a la etapa 7 que remite al investigador a la etapa inicial para repetir el proceso, en caso contrario pasa a la etapa 8 y luego a la 9 para finalizar el proceso. La fundamentación de la propuesta de cada una de estas etapas se discute en Sagaró y Zamora (2020). Un esquema general donde se aprecia la secuencia de estas etapas y los procedimientos a ejecutar en cada una de ellas se presenta en la figura 7.

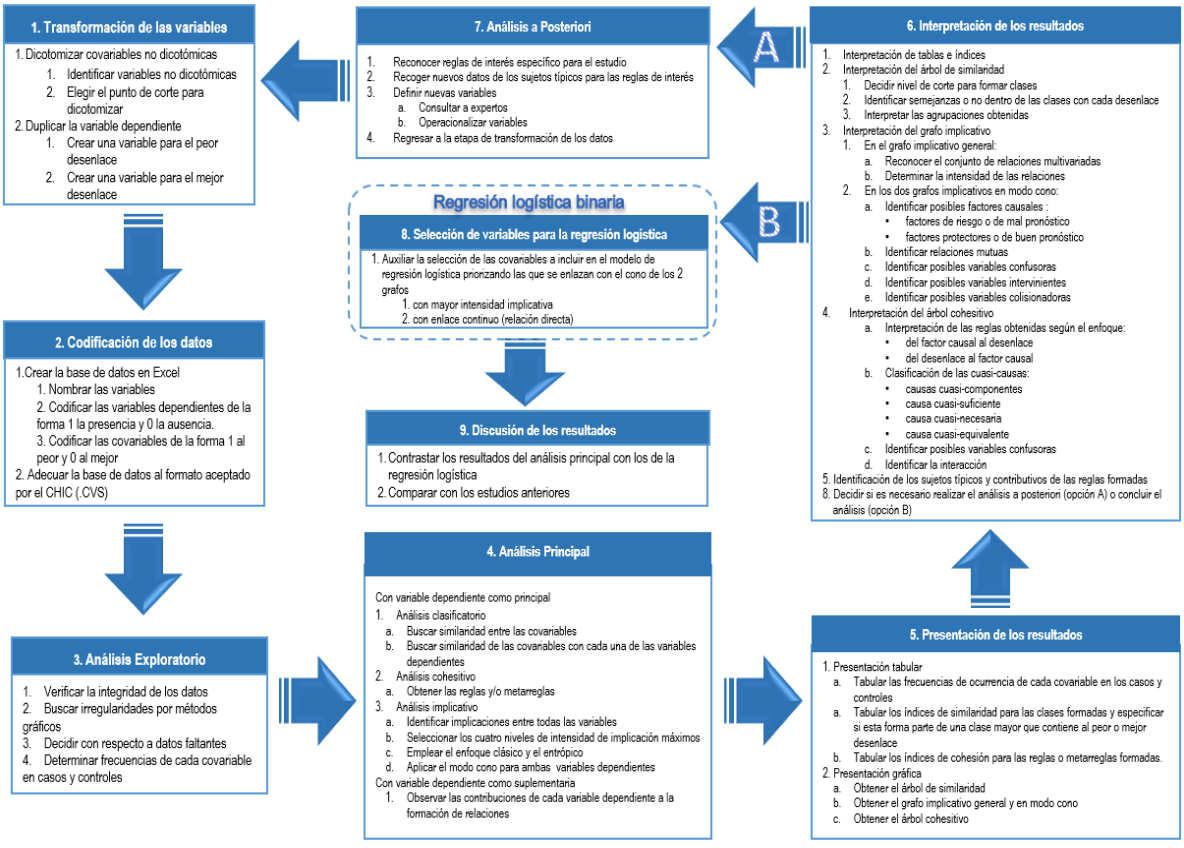


Figura 7 - Esquema de la metodología ASI-IMC

3.3 Recomendaciones para el empleo de la metodología

En el árbol de similaridad, el investigador seleccionará el nivel de similaridad a partir del cual se considerarán las clases formadas en dependencia del grado de homogeneidad que desee alcanzar en ellas. A medida que este nivel disminuya, también decrecerá la cantidad de clases que se forman y su grado de homogeneidad.

En el grafo implicativo, las relaciones implicativas que se muestran dependen del valor umbral fijado por el investigador. El software CHIC permite elegir hasta cuatro valores umbrales diferentes a la vez, asignándole a cada uno de ellos colores diferentes para su representación. Según sea el valor mínimo de los valores umbrales elegidos, se podrá identificar una mayor o menor cantidad de factores asociados con las variables de

respuesta. Se sugiere emplear los cuatro niveles mayores que brinda la aplicación, para obtener la mayor cantidad de relaciones existentes entre las variables, excepto en los casos que con gran intensidad implicativa, del 100 o 99 %, aparezcan casi todas las variables representadas en el grafo. A medida que se disminuye el valor umbral aparecerán, de existir, nuevas relaciones implicativas.

En el análisis de la tipicalidad y contribución de las variables suplementarias a las clases formadas, el CHIC calcula el riesgo de cada variable de no ser típica, por lo que se tendrá presente que, las variables con menor riesgo son las más típicas o que más contribuyen, respectivamente, a la formación de las clases.

Se sugiere emplear los dos enfoques que brinda la aplicación, el clásico y entrópico, para reafirmar los hallazgos; este último cuando el tamaño de muestra supere los 100 individuos.

Otra recomendación es no usar las variables dependientes como suplementarias si los controles no están balanceados con los casos.

4 Conclusiones

El reconocimiento e implementación de ciertas modificaciones al análisis estadístico implicativo permitieron diseñar una metodología para su empleo contextualizado a las investigaciones médicas de causalidad, brindando una nueva forma de análisis y una visión renovadora en la interpretación de los resultados, en correspondencia con los enfoques actuales del estudio de la causalidad en las ciencias médicas.

Dicha metodología fue aplicada en investigaciones de causalidad en diferentes problemas de salud, en los cuales demostró su buen desempeño en la identificación de factores pronósticos y de riesgo, por lo que se considera que su introducción entre las herramientas estadísticas de uso frecuente en estas investigaciones, contribuirá substancialmente a incrementar la calidad de las mismas.

Referencias

[1] Bacallao, J. (2012). Enfoques modernos del sesgo y la causalidad en la investigación epidemiológica, *Revista Cubana de Salud Pública*, 38, 686-701.

[2] Bacallao, J., Alerm, A. y Ferrer, M. (2016). *Paradigma del curso de la vida: implicaciones en la clínica, la epidemiologia y la salud pública*, Editorial Ciencias Médicas, La Habana.

[3] Brewer, L. E., Wright, J. M., Rice, G., Neas, L. y Teuschler, L. (2017). Causal inference in cumulative risk assessment: The roles of directed acyclic graphs, *Environment international*, 102, 30-41.

[4] Cortes, T. R., Faerstein, E. y Struchiner, C. J. (2016). Use of causal diagrams in Epidemiology: application to a situation with confounding, *Cadernos de saude publica*, 32(8), e00103115-e00103115.

[5] Galano, K., Sagaró, N. D., Zamora, L., Lambert, Y. y Mingui E. (2019). Análisis estadístico implicativo en la identificación de factores pronósticos de mortalidad del cáncer renal, *Rev. inf. Cient,* 97(4).

[6] García, Y. M., Zamora, L. y Sagaró, N. D. (2015). Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón, *Medisan*, 19(08), 947-957.

[7] Gras, R., Suzuki, E., Guillet, F. y Spagnolo, F. (2008). *Statistical implicative analysis*, Springer, New York.

[8] Gras, R., Régnier, J. C., Marinica, C. y Guillet, F. (2013). *L'analyse statistique implicative Méthode exploratoire et confirmatoire à la recherche de causalités* (p. 522), Cépaduès Editions, France.

[9] Gras, R. (2014). Genese et developpement de l'analyse statistique implicative: retrospective Historique, *Educ Matem Pesq São Paulo*, 16(3), 645-661.

[10] Greenland, S. y Pearl, J. (2014). *Causal diagrams*, Wiley StatsRef: Statistics Reference Online, New Jersey.

[11] Hernán, M. A., Robins, J. M. (2020). Part I: Causal inference without models. In: *Causal Inference*, Chapman & Hall/CRC, Boca Raton.

[12] Manterola, C., & Otzen, T. (2015), Los sesgos en investigación clínica, *International Journal of Morphology*, 33(3), 1156-1164.

[13] Moraga, A., Zamora, L., Sagaró, N. D., Moraga, A. y Rodríguez, A. (2016), Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de pulmón, *Medisan*, 20(3), 344-353.

[14] Moraga, A., Zamora, L., Sagaró, N. D., Moraga, A. y Rodríguez, A. (2017). Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de mama, *Medisan*, 21(4), 395-406.

[15] Moraga, A., Zamora, L., Sagaró, N. D., Moraga, A. y Rodríguez, A. (2018). Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata, *Medisan*, 22(1), 48-56.

[16] Morin, E. (2004). La epistemología de la complejidad, *Gazeta de Antropología,* 20(2).

[17] Paez, Y. C., Sagaró, N. D., & Zamora, L. M. (2018), Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso, *Medisan*, 22(6), 431-440.

[18] Pardo, S. S., Sagaró, N. D., Zamora, L. y Viltre, D. M. (2019). Utilidad del análisis estadístico implicativo para identificar factores pronósticos en pacientes con cáncer de mama, *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, 44(4).

[19] Prado, M. D. (2012). *Búsqueda de relaciones causales para aplicaciones en cardiología y psiquiatría*, Tesis Doctoral, Universidad Carlos III de Madrid.

[20] Ramis, R. M. (2004). La causalidad compleja: un nuevo paradigma causal en Epidemiología, *Revista Cubana de Salud Pública*, 30(3).

[21] Requena, C. C. (2017). *Métodos estadísticos para evaluar la causalidad en estudios observacionales*. Tesis de maestría, Universidad Complutense de Madrid.

[22] Reyna, J. L. (1971). Modelos causales y tipo ideal: una estrategia teórico-metodológica para la sociología del desarrollo latinoamericano, *Demografía y economía*, 5(1), 1-12.

[23] Rothman, K. J. y Greenland, S. (2005). Causation and causal inference in epidemiology, *American journal of public health*, 95(S1), S144-S150.

[24] Sagaró, N. D., & Zamora, L. M. (2019), ¿Por qué emplear el análisis estadístico implicativo en los estudios de causalidad en salud?, *Revista Cubana de Informática Médica*, 11(1), 88-103.

[25] Sagaró, N. D. y Zamora, L. (2019). Métodos actuales para asegurar la validez de los estudios de causalidad en medicina, *Gaceta Médica Espirituana*, 21(2), 146-160.

[26] Sagaró, N. D. y Zamora, L. (2019). Análisis estadístico implicativo versus Regresión logística binaria para el estudio de la causalidad en salud, *Multimed*, 23(6), 1416-1440.

[27] Sagaró, N. D. y Zamora, L. (2020). Propuesta metodológica de contextualización del análisis estadístico implicativo a las investigaciones médicas de causalidad, *Revista Cubana de Informática Médica*, 12(1), 31-43.

[28] Williamson, E. J., Aitken, Z., Lawrie, J., Dharmage, S. C., Burgess, J. A. y Forbes, A. B. (2014). Introduction to causal diagrams for confounder selection. *Respirology*, 19(3), 303-311.

LES STRATÉGIES DE RIPOSTES, DONT LES PLANTES TRADITIONNELLES, À LA PANDÉMIE DE COVID-19 CHEZ DES ÉTUDIANTS UNIVERSITAIRES MALGACHES :CAS DE L'UNIVERSITÉ D'ANTSIRANANA

Christian Razanatsoavina[1], Bakys Bruno Ralahady[2], Andrianasy Angelo Djistera[3], Jeanne Ravaosolo[4], Jean-Claude Laberche[5]

RESPONSE STRATEGIES, INCLUDING TRADITIONAL PLANTS, TO THE COVID-19 PANDEMIC AMONG MALAGASY UNIVERSITY STUDENTS: CASE OF THE UNIVERSITY OF ANTSIRANANA

RÉSUMÉ

L'objectif de cet article est de montrer les différentes protections pratiquées chez les étudiants d'Antsiranana, confinés, lors de la pandémie de Covid-19. Face à l'absence d'un traitement officiel et efficace pour cette maladie, les étudiants ont adopté, soit un traitement préventif officiel basé sur les tisanes de CVO et la prise de vitamine C, soit un traitement préventif traditionnel à base de fumigations de plantes dont l'eucalyptus, le gingembre et le citronnier. Sont ainsi rapportées dans cet article : 1) L'acquisition des données au moyen d'un questionnaire complété par 204 étudiants, 2) L'analyse des résultats et des implications que ceux-ci peuvent engendrer. Pour disposer de données précises et normées, un questionnaire semi-directif a été utilisé. Les résultats de cette analyse précisent bien les ressentis du terrain. D'abord, la majorité des étudiants ont opté plus pour la médecine traditionnelle à bases de plantes traditionnelles plutôt que pour le CovidOrganics et la vitamine C. L'analyse statistique implicative a permis de mettre en évidence des relations très intéressantes comme le souci des étudiants de protéger la poursuite de leur études et l'influence des conditions précaires dans lesquelles ils vivent

Mots-clés : *Codiv-19, riposte, plante, médecine traditionnelle, Madagascar, ASI*

ABSTRACT

The objective of this article is to show the different protections practice by the confined students of Antsiranana, during the Covid-19 pandemic. In the absence of an official and efficient treatment for this disease, the students adopted either an official preventive treatment based on CovidOrganics tea and vitamin C intake, or a traditional preventive treatment based on fumigation of plants such as eucalyptus, ginger and lemon tree. The following are reported in this article: 1) The acquisition of data by means of a questionnaire completed by 204 students, 2) The analysis of the results and the implications that these may have. In order to have precise and standardized data, a semi-directive questionnaire was used. The results of this analysis clearly indicate the feelings of the field. First, the majority of students opted for traditional herbal medicine over CovidOrganics and vitamin C. The Implicative Statistical Analysis revealed very interesting relationships such as the concern of the students to protect their further education and the influence of the precarious conditions in which they live.

Keywords : *Codiv-19, response, plant, traditional medicine, Madagascar, ASI*

[1]Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, cmrazanatsoavina@gmail.com

[2]Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, ralahadybru@yahoo.fr

[3]Faculté de Droit, d'Economie, de Gestion et de Mathématiques, Informatique et Applications, Université de Toamasina, angelo.djistera@yahoo.fr

[4]Arts, Lettres et Sciences Humaines, Université de Tuléar, ufr.ravaosolo@moov.mg

[5]Professeur des Universités émérite, jeanclaudelaberche@gmail.com

1 Introduction

Depuis près de deux ans, le Monde affronte une pandémie qui n'a pas épargné l'Afrique et encore moins Madagascar. En attendant le développement généralisé de la vaccination qui seule devrait pouvoir y mettre un terme, chaque pays a adopté un système de protection originale. Si les pays occidentaux ont mis en place des gestes barrières basés sur l'isolement, le port du masque, la distanciation et l'usage de gel hydro-alcoolique pour la désinfection, d'autres pays comme Madagascar, ont choisi d'utiliser les vertus des plantes médicinales locales soit pour se protéger, soit pour se soigner. La raison en est simple et correspond à ce que l'on est en droit d'attendre d'un pays qui ne bénéficie pas d'un système sanitaire de modèle occidental et qui recourt encore à la médecine traditionnelle, basée sur l'utilisation de plantes médicinales locales accompagnées éventuellement d'incantations avec les esprits des Anciens (Dandouau , 1922).

Dès l'apparition des premiers cas importés de Covid-19 à Madagascar, le gouvernement malgache a immédiatement pris des mesures pour gérer cette crise sanitaire (Instat, 2020), et a ordonné la fermeture des écoles et des universités, sans pour autant fermer les campus. De ce fait, de nombreux étudiants sont restés confinés dans les résidences universitaires. A Antsiranana celle-ci est vétuste avec des cloisons en planches ou en rideau pour séparer les occupants d'une même chambre, initialement prévue pour un seul occupant alors que bien souvent actuellement trois ou quatre s'y entassent. L'espace pour chacun est plus que réduit et l'accès aux installations sanitaires très précaire. Exceptés les *vatsy tsinjo* (soutien alimentaire d'urgence) (Plan Multisectoriel D'urgence Madagascar, 2020) et l'officielle solution de Covid Organics, CVO, (décoction à base de plantes qui serait composée à 62% d'*Artémision annua*, mais dont la composition complète n'a jamais été dévoilée), les étudiants ont été abandonnés et exclus de tout service de protection sanitaire. Quand ils ne voulaient pas suivre, pour x raisons, les directives officielles d'utiliser les infusions de COV associées à des prises de vitamine C, ils se sont rappelés alors les gestes séculaires et traditionnels utilisés dans leurs familles pour prévenir ou guérir les maladies (Debray et *al*, 1975) et ont riposté au covid-19 en utilisant plusieurs traitements exhaustifs censés chasser le virus. C'est cette stratégie de ripostes contre le covid-19 que nous avons étudiée à Antsiranana, l'une des six universités malgaches, située au nord de la grande Île. Sur un échantillon de 204 étudiants confinés, nous avons effectué une enquête sur leur comportement. Leurs réponses ont été analysées par ASI et plusieurs conclusions s'en dégagent

2 Matériels et méthodes

Dans cette étude, deux méthodes sont successivement utilisées : recueillir l'information, puis traiter cette information.

2.1 Mesurer la perception des étudiants universitaires

2.1.1 Un questionnaire d'enquête

L'étude a été effectuée au premier trimestre 2021, sur 204 étudiants, tous cursus confondus, au sein de campus universitaire d'Antsiranana. Cette étude a été effectuée sans distinguer l'âge et le sexe des étudiants, ni leur niveau d'études ou leur lieu habituel de résidence.

Nous avons opté pour un type de questionnaire semi-directif, sur le modèle défini par Razanatsoavina *et al.* (2019) avec une réponse ouverte ou fermée.

Les grands points abordés dans ce questionnaire (Annexe 1) sont :

- les caractéristiques de l'étudiant,
- son environnement matériel,
- son ressenti vis à vis de la maladie,
- l'éventuelle vaccination.

Ce questionnaire a été distribué le fin mars 2021 aux étudiants présents alors sur le campus et manifestement confinés. Il leur a été présenté dans leurs appartements et pour avoir leurs réponses, il leur a été demandé de le remplir directement et immédiatement.

2.1.2 Les limites d'enquête

L'enquête est du type photographie d'une population à un instant donné. En effet, cette enquête était destinée aux étudiants confinés en résidence et présents le jour de l'enquête. Il est incontestable que la population cible ne couvre pas l'ensemble des 2000 étudiants de l'Université d'Antsiranana. Toutefois, le nombre de questionnaires remplis dépassant les deux cents (204 exactement) sur près de 400 étudiants confinés sur le campus et susceptibles d'être remplis, il représente un échantillon représentatif surtout si l'on tient compte des difficultés matérielles de distribution, d'explication et de collecte d'un questionnaire destiné à des confinés et ce en pleine pandémie.

2.1.3 Dépouillement et Analyses statistiques

Le dépouillement a été effectué selon la technique présentée précédemment par Ralahady *et al. (*2019c) et qui distingue trois phases : collecte de données, traitement des données et analyses des résultats. C'est un groupe d'étudiants du parcours informatique de l'Ecole Normale Supérieure pour l'Enseignement Technique à l'Université d'Antsiranana qui avait la responsabilité de dépouiller les données recueillies. Le codage et le contrôle de qualité de données ont été réalisés à l'aide du logiciel bureautique tableur Excel. Ce dépouillement permet déjà de recueillir un certain nombre d'observations.

Le traitement des données a été réalisé ensuite avec l'outil d'analyse Analyse Statistiques Implicative ASI-MGK. Néanmoins par rapport aux techniques décrites par Ralahady *et al.*(2019a) un traitement préliminaire a établi (au seuil minimal de support égal 30%) une génération successive de règles valides en étudiant le nombre des règles et la compacité du graphe (nombre des composantes connexes) selon une variation du risque d'erreur.

3 Présentation des résultats et interprétations

3.1 Caractéristiques de la cohorte d'étudiants après le dépouillement des données

Lors du dépouillement, il a été immédiatement constaté que l'immense majorité des sondés utilisent des plantes traditionnelles dans leur vie courante pour se prévenir de la maladie ou la combattre (Figure 1).

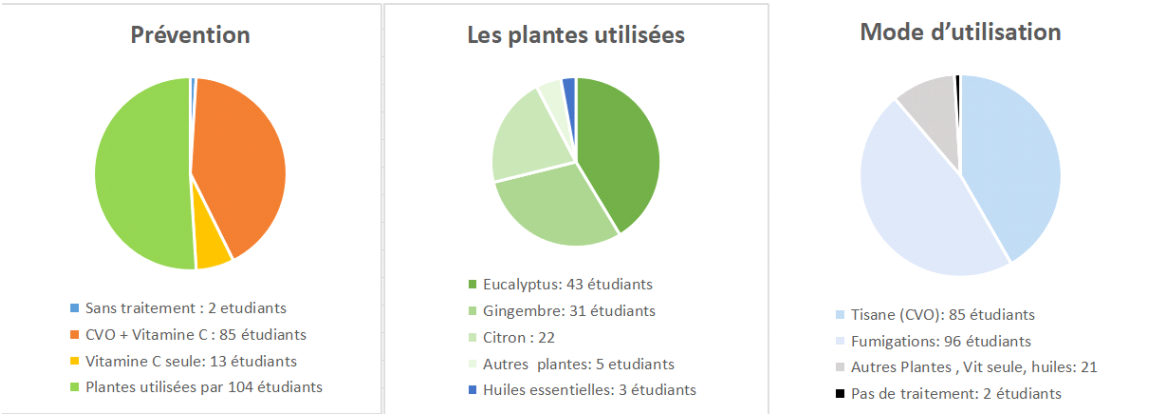


Figure 1. Effectifs observés lors du dépouillement des données et classés selon les mesures de préventions adoptées (a), le détail des plantes utilisées pour cette prévention (b), enfin le mode d'utilisation (c).

Plusieurs points importants sont à retenir de ces figures.

Une très grande majorité d'étudiants se traitent préventivement contre le Covid-19 signe de leur inquiétude devant la pandémie. Seuls deux étudiants ont déclaré ne point se traiter. Parmi les 99 % d'étudiants qui se protègent un peu plus de 4 étudiants sur 10 adoptent le mélange officiel CVO distribué dans chaque village et qui contient de l'artémisinine (dérivé de l'*Artemisia annua*) connue comme étant un anti-paludé en, fébrifuge et protégeant les voies respiratoires des infections (Tu Youyou, 2019) ainsi que de l'extrait de Ravintsara, (*Ravensara aromatica*) une plante agissant contre les infections respiratoires (Ramiandrasoa D. 2005) De même elle est reconnue comme antivirale et constituant une alliée utile pour la prévention de la grippe et pour lutter contre toute forme d'infection virale comme l' herpès ou les hépatites (Giraud-Robert, 2005).

En même temps que la prise du CVO, le ministère de la sante malgache préconisait la prise de vitamine C censée protéger contre les infections pulmonaires et donc contre le Covid. Il reprenait très certainement les théories du Prix Nobel Julius Pauling qui dés 1970 conseillait la prise de 1000 mg quotidiens de vitamine C pour diminuer le risque de contracter une infection virale (Pauling L.et Roth B., 1988) même si actuellement l'efficacité de la vitamine C contre le Covid-19 n'a pas été vraiment démontrée (Feyaerts A.F. et Luyten W., 2020). Près de la moitié des étudiants (98) ont absorbé cette vitamine le plus souvent en association avec le CVO, plus rarement seule (13 étudiants).

L'autre partie des étudiants (la majorité 104/204) a préféré mettre en œuvre leur système de défense personnelle. Ils se sont tournés essentiellement vers des fumigations d'Eucalyptus, et de Gingembre et de Citronnier. Quelques rares étudiants ont préféré les massages aux huiles essentielles de plantes ou à des mélanges de sept plantes, qu'ils n'ont point identifiées lors de l'enquête.

Le tableau 1 précise les caractéristiques des trois plantes les plus utilisées par les étudiants et leurs usages reconnus.

Nom malgache	**Kininim-potsy**	**Sakay tany**	**Voasary makirana**
Nom français	Eucalyptus	Gingembre	Citronnier
Nom latin	Eucalyptus globulus Labill.	Zingiber, officinale Roscoe	Citrus limon (L.) Burm.
		Partie utilisée	
Écologie	feuilles	rhizome	fruit et écorce
Exigences	faibles	sols bien drainés	situation ensoleillée
Sols	sols acides	sols riches, acides	sols riches
Besoins en eau	+ 600 mm/an	craint les excès d'eau	importants
Sensibilité	au froid au feu	fusariose, froid	sensible au froid
Domestication	cultivé	cultivé	cultivé
		Propriétés thérapeutiques	
Propriétés générales	effet purifiant sur l'air, désinfectant	anti-inflammatoire, antipyrétique anti-nauséeux	Présence de flavonoïdes antioxydants
Sur les poumons	bronchites, toux	asthme	vertus expectorantes
Reins	sur les infections rénales		sur l'insuffisance rénale
Cholestérol		anti cholestérol	
Action sur le foie et le pancréas		combat les insuffisances biliaires et pancréatiques	stimule la production de bile
Rhumatismes		anti-inflammatoire	ralenti l'évolution des maladies articulaires.

Origine des données: Debuigne G.et al. (2019) et Fleurentin J. Wenige B., (2018)
Tableau 1. Les trois principales plantes utilisées par les étudiants interrogés.

Les trois plantes les plus utilisées par les étudiants en fumigations ont toutes des propriétés positives sur les voies respiratoires. Ceci est en accord avec le fait que, du moins les premiers mois, l'infection par le Covid-19 a été grossièrement assimilée à une grippe entraînant des difficultés respiratoires.

Il est remarquable que l'Eucalyptus, bien que d'introduction récente (1857) à Madagascar (Verhaegen *et al.*, 2011), soit la première plante choisie par les étudiants pour se protéger contre le covid-19. Ses bienfaits pour la santé des organes et des voies respiratoires sont bien reconnus. La brume d'eucalyptus est souvent utilisée pour soulager la toux, la sinusite. Cette plante est considérée comme l'une des meilleures espèces pour le traitement des processus grippaux ou de toute maladie affectant les voies respiratoires. (Lassak et Mc Carthy, 1983) L'utilisation de l'Eucalyptus comme prévention au covid-19 semble d'ailleurs, assez générale en Afrique comme le rapporte Chaachouay N. (2021). Mais dans le traitement direct du Covid-19 c'est plutôt vers l'utilisation directe des huiles essentielles que semblent s'orienter les expérimentations en cours (Valussiun M.*et al.*, 2021)

Trouver le gingembre comme seconde plante utilisée pour la protection des étudiants contre le covid-19, n'est pas en elle-même une surprise, puisque le gingembre est connu dans de nombreux pays du monde pour ses vertus thérapeutiques, depuis plus de 5000 ans et il est très utilisé à Madagascar. On s'en sert pour combattre les infections, la fatigue, les douleurs musculaires et, surtout, les problèmes digestifs (vomissements, diarrhées...) (Odukoya J.*et al.* 2021). Contre le covid-19, il apparaîtrait que le gingembre pourrait réduire les symptômes qui se produisent lors d'une infection virale mais les études précliniques et cliniques sont nécessaires pour le confirmer (Musdja M.Y.2021).

Le citron est aussi utilisé par quelques étudiants. Or le citron est à la fois un anti-virus et un antiseptique de choc. Il combat naturellement les virus et les bactéries aussi bien en

externe qu'en interne, pour protéger de la grippe ou du rhume. Grâce à sa richesse en vitamine C, il va, de plus, stimuler l'immunité (Debuigne G. *et al.*, 2019). Contre le covid-19 plusieurs essais sont développés dans le monde, mais c'est surtout comme désinfectant des mains que le citronnier est éventuellement préconisé (Alghamdi, 2021).

Il apparaît ainsi que les étudiants, à défaut de disposer de traitements efficaces et reconnus ou de vaccins, se protègent avec des produits qu'ils soient officiels ou traditionnels. Ceux-ci leur permettent d'assainir leur environnement grâce aux effets antiseptiques mais aussi de stimuler leurs défenses immunitaires à défaut d'avoir une action directe sur le virus.

Il est probable que la présence de plantations d'eucalyptus sur le campus universitaire, près du lieu de confinement des étudiants, en plus de son utilisation familiale très fréquente a favorisé utilisation de ce matériel végétal comme moyen de protection contre le Covid-19 à Antananrivo.

3.2 Les résultats obtenus par l'ASI

Les informations générales apportées par l'observation directe des résultats de l'enquête vont se trouver précisées par l'outil d'analyse ASI-MGK. Un calcul préliminaire sur le nombre de règles est d'abord effectué.

3.2.1 Nombre de règles valides en fonction du risque d'erreur

Dans un premier temps, nous avons effectué un calcul préliminaire permettant de connaître le nombre de règles valides en fonction du risque d'erreur que nous pourrions retenir. Nous obtenons ainsi un graphe permettant de connaître le nombre de règles valides en fonction du risque d'erreur choisi (figure 2)

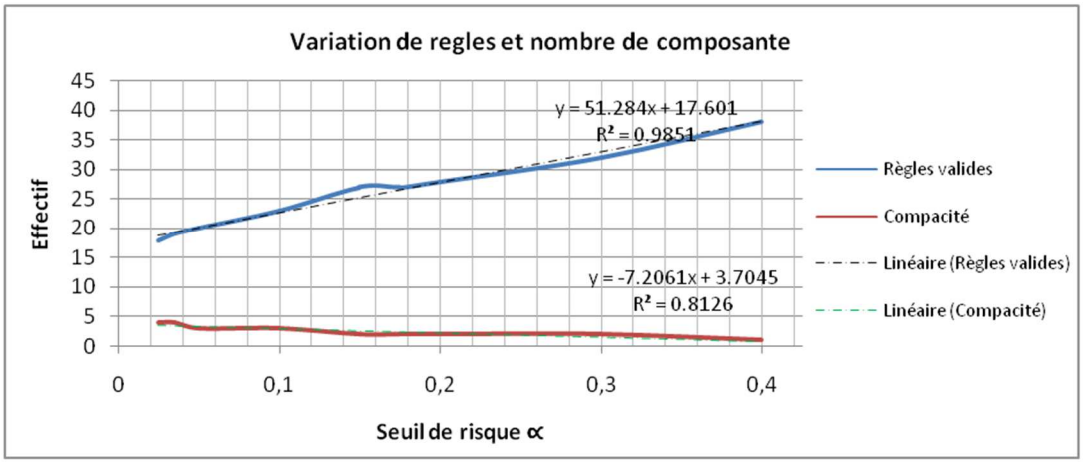


Figure 2: Nombre de règles et nombre des composantes connexes du graphe en fonction de risque d'erreur

La figure 2, montre la variation croissante et linéaire (r =0.99) du nombre des règles d'association valides et la variation décroissante et linéaire (r = 0.90). Elle permet une augmentation du nombre des composantes connexes du graphe implicatif dans l'intervalle de confiance [0,60 à 0,975]. Ainsi au seuil du risque d'erreur 0.025 (seuil de confiance) nous avons 18 règles (figure 3-a) d'associations valides selon la mesure de qualité M_{GK} (Ralahady *et al.*, 2019b) et le graphe implicatif présent quatre composantes connexes. En augmentant ce seuil à 0.95, nous avons une augmentation de deux règles valides et une combinaison de deux composantes, c'est-à-dire, 20 règles et 3 composantes (figure 3-b).

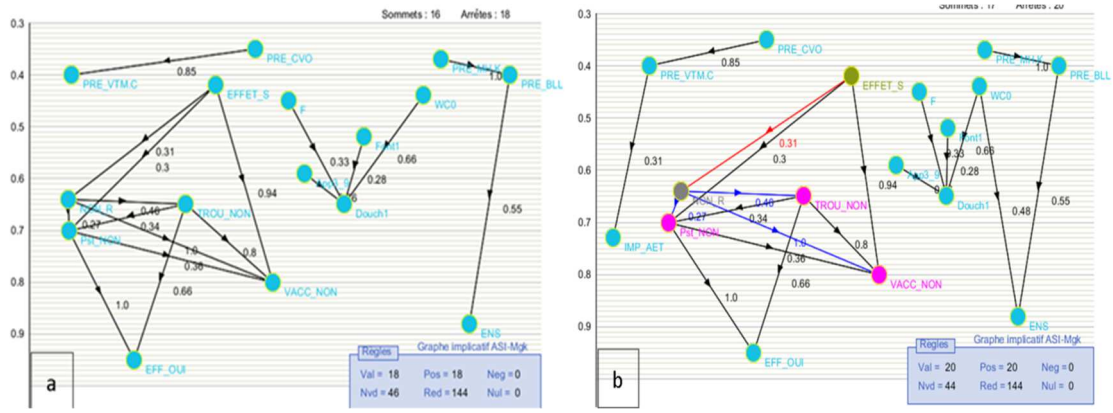


Figure 3: L'ensemble des grands axes des résultats obtenus par l'ASI ; a) obtenu au risque d'erreur ∝=0.025 et b) obtenu au risque d'erreur ∝=0.05.

Données générales communes à tous les graphes ASI. Pour tous les graphes, les nœuds sont les items étudiés, les arcs sont les relations implicatives entre les items et les poids sont les indices ou les forces des relations. Les noms des items seront précisés sous chaque graphe.

Nous avons détaillé des relations mises en évidence par l'ASI-MK. Les deux premiers points confirmeront d'abord ce qui a été observé dans les premières analyses tout en montrant des relations nouvelles. Le dernier point fera apparaître des relations entre la crainte de la maladie (donc la mise en œuvre de précautions) et l'environnement social des étudiants durant le confinement.

3.2.2 Prévention avec les traitements officiels: CVO et Vitamine C.

La figure 4 présente les relations entre les traitements officiels et leur impact sur l'enseignement.

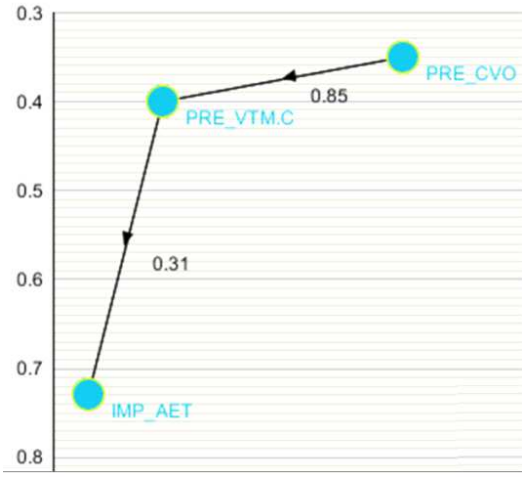


Figure 4 : Relation entre l'utilisation du protocole officiel (CovidOrganics et Vitamine C) et impact sur l'enseignement (risque d'erreur , a = 0,05)

Items PRE_{VTMC}: prévention avec COV, PRE_{VTMC}: prévention avec la prise de Vitamine C, IMP_{AET}: impact sur l'arrêt de cours. Dans cette figure on constate d'abord : a) une règle significative ou l'item est l'impact sur l'arrêt de cours. Au risque d'erreur 0.025 la règle significative. La prévention avec le CovidOrganics implique la prévention avec la vitamine C. Tous les étudiants qui ont utilisé le CovidOrganics ont pris également de la

vitamine C. L'indice n'est cependant pas égal à 1 puisque quelques étudiants (13 sur 204) ont pris seulement de la Vitamine C comme l'a montré l'analyse des résultats bruts Le risque d'erreur à 0,05, a révélé une nouvelle règle valide, Ceci signifie que la prise vitamine C implique pour les étudiants une protection contre un arrêt d'études. Il est intéressant de noter que si les étudiants lors de ce confinement étaient anxieux pour leur santé, puisque 99 % se sont protégés, les étudiants qui ont pris le remède officiel étaient aussi soucieux de leur devenir et craignaient pour leurs études.

3.2.3 Traitement traditionnel et crainte pour la poursuite des études

Remarquons d'abord que les relations entre les items liés à l'utilisation du citron (PRE_M.C.), du gingembre (PRE_MG), aux huiles essentielles (PRE_BR) ou aux mélanges de 7 feuilles (PRE_MG7) n'apparaissent pas dans la figure 3b. Nous présentons donc la seule relation présente à savoir celle pour les étudiants de faire des fumigations avec de l'eucalyptus et leur crainte sur leurs enseignements (figure 5).

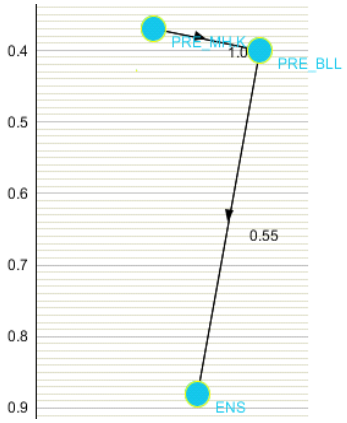


Figure 5: L'usage de plante médicinale Kininim-potsy (feuilles d'Eucalyptus globulus Labill) et son utilisation en fumigation pour riposter la pandémie en liaison avec la crainte de répercussion sur le cursus universitaire. Risque d'erreur a = 0,025

Items $\mathrm{PRE_{MH}}.\mathrm{K}$: prévention avec l'eucalyptus, $\mathrm{PRE_{BLL}}$: prévention avec fumigation et ENS : trouble sur le domaine d'enseignement. Avec un risque d'erreur 0.025, les résultats font apparaître deux règles significatives et pour les trois items. Il s'agit d'abord d'une implication significative $\mathrm{M_{GK}}(\mathrm{PRE_{MH}}.\mathrm{K} \rightarrow \mathrm{PRE_{BLL}}) = 1$ confirmant bien que tous les utilisateurs d'eucalyptus l'utilisent en fumigation, puis une deuxième implication significative plus originale. Les étudiants qui ont opté pour la prévention avec l'eucalyptus craignait non seulement pour leur sante mais aussi les troubles que le Covid-19 pouvait entrainer sur leur scolarité. Lorsque nous avons augmenté le risque d'erreur 0.05, aucune nouvelle règle n'ait apparue. Il est très intéressant de noter que comme pour les utilisateurs des produits officiels de prévention, les étudiants qui ont pris le «remède officiel» étaient soucieux de leur devenir et craignaient pour leurs études. Ceci confirme, si ceci avait besoin d'être confirmé, la maturité de ces jeunes confinés dans des conditions matérielles très difficiles, et pratiquement abandonnés à eux-mêmes même si l'État leur a apporté des aides alimentaires et matérielles comme du riz, des grains secs, de l'huile, des masques sanitaires et des désinfectants, etc. (Plan Multisectoriel D'urgence Madagascar, 2020). Le passage du risque d'erreur α de 0,025 à 0,5 a permis de mettre une autre relation très intéressante comme nous allons le voir maintenant.

3.2.4 Prévention avec de l'Eucalyptus dans une situation matérielle précaire

La figure 6 présente le même fragment de la figure 2 aux risques d'erreur α de 0,025 (figure 6 a) et 0,05 (figure 5 b). Il apparaît nettement une nouvelle liaison entre le milieu sanitaire des étudiants et l'usage d'Eucalyptus dans la prévention du covid-19.

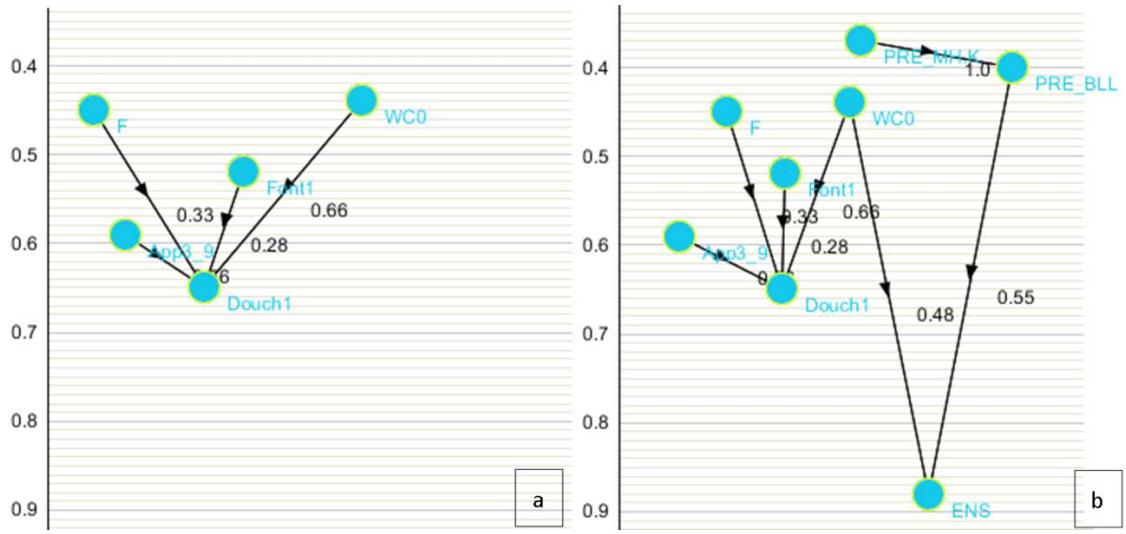


Figure 6 : Représentation graphique des conditions précaires chez les étudiants universitaires et usage de fumigations d'eucalyptus. Encerclée: la relation qui relie les modalités de prévention et les conditions matérielles de logement des étudiants

Items: F : sexe féminin, $App3_9$: 3 à 9 étudiants dans un appartement, WC : aucune toilette, $Douch1$: avoir une douche et $Font1$: une borne fontaine à disposition $\mathrm{PRE_{MH}}$. K: prévention avec l'eucalyptus, BRE_{BLL} : prévention avec fumigation et ENS : trouble sur le domaine d'enseignement. Dans cette figure, il apparaît : a) quatre règles significatives sur cinq items, b) une nouvelle règle significative additionnelle quand le risque d'erreur augmente. Les étudiants mettent en relation leurs difficiles conditions de vie dans la cité universitaire et l'impact que ceci peut avoir sur la maladie qu'ils combattent avec des fumigations d'eucalyptus. Manifestement ils veulent pouvoir continuer d'étudier quelque soient les difficultés matérielles ou sanitaires. Il est certain que cette relation n'est pas très forte (0,48) et elle n'est apparue que quand le risque d'erreur a augmenté. Mais on voit que les conditions matérielles sont sous- jacentes au problème du Covid-19 quand il s'agit pour les étudiants de penser à la poursuite de leurs études Il est incontestable que les étudiants d'Antisaranana sont trop nombreux dans des chambres trop petites ou cohabitent de 3 à 9 personnes et où tous doivent se partager un seul WC, une seule douche, une seule borne fontaine en extérieur. Rançon de la démocratisation de l'enseignement dont on ne peut que se féliciter mais où malheureusement les conditions matérielles ne suivent pas. Maturité de tous ces jeunes qui conscients de tous ces problèmes essaient de ne pas se décourager et tentent contre vents et marées, en particulier contre la pandémie du Covid-19 de poursuivre le mieux possible leurs études.

4 Conclusion et perspectives

Cet article présente les ripostes pratiquées par les étudiants de l'Université d'Antsiranana lors du premier confinement lié à la pandémie du Covid-19. Elles sont de deux types : la prévention officielle (CVO plus vitamine C) et la prévention traditionnelle à base de plantes dont l'Eucalyptus. La qualité des produits utilisés fait qu'effectivement

les étudiants confinés disposaient théoriquement de moyens permettant la désinfection de leur milieu, mais les conditions sanitaires auxquelles ils sont confrontés font que c'était bien difficile pour eux pour que ces techniques soient efficaces. Cette étude nous a permis aussi de constater que les étudiants confinés étaient dans la crainte que cette pandémie ne perturbe la poursuite de leurs études et cette crainte était probablement plus forte que la crainte de maladie elle-même. Il est de notoriété publique que même si l'immense majorité des étudiants ont pris ces traitements préventifs, plusieurs ont été atteints mais ils ne l'ont point déclaré de peur que ceci ne nuise à leurs études. Malheureusement nous n'avons pas eu accès aux données des étudiants testés positifs ou malades lors et après notre étude Faute d'enquête systématique et d'autorisations administratives, il nous a été impossible de connaître le nombre des étudiants qui ont attrapé le Covid-19. En fait, les précautions avancées qu'elles soient officielles ou traditionnelles peuvent-elles être efficaces quand la majorité des étudiants vivent dans des conditions précaires et où la promiscuité est de règle? Dès le départ de cette étude, n'était-il point utopique d'imaginer que des pratiques barrières puissent être efficaces? Probablement oui, mais il nous est apparu intéressant de suivre ces étudiants confinés et pratiquement abandonnés de tous qui plutôt que d'accepter ou refuser en masse un traitement préventif officiel, se sont tournés vers des plantes traditionnelles bien connues dans leur vie quotidienne.

Ceci montre combien est complexe ce type d'études où plusieurs facteurs interagissent et il faut toujours travailler en «système». Si les études statistiques traditionnelles sur les données brutes apportent déjà de bons résultats seules les analyses de type ASI permettent d'établir des relations avec une bonne probabilité de véracité. Le plus dur est de concevoir les questionnaires qui doivent éviter les questions redondantes ou pas assez ciblées qui en définitive ne permettent pas d'établir de relations entre les différents points étudiés. Il est certain que lors du renouvellement de ce type d'enquête il nous faudra retravailler profondément notre questionnaire. Il serait, par exemple, très intéressant de savoir pourquoi les étudiants étaient confiants dans les gestes de protection à partir de plantes naturelles, ce qui nous a fait défaut.

Madagascar dans l'ensemble de son territoire détient un grand potentiel et une grande connaissance des plantes médicinales, ainsi que leurs usages. Toutefois, les transmissions, les connaissances, les sagesses doivent être assurées. En outre, ces plantes risquent aussi une disparition à cause de multiples raisons comme la surexploitation, la déforestation ou les feux de brousse. Les malgaches doivent se réveiller et valoriser ces ressources naturelles Il faudrait envisager une formation spécifique sur les plantes médicinales malgaches puis développer et vulgariser la filière (plantes médicinales, aromatiques et/ou cosmétiques). De plus, les besoins en plantes médicinales augmentent de plus en plus de par le monde et surtout exporter des produits phytopharmaceutiques est une opportunité économique à saisir. Les malgaches doivent regarder la réalité en face, comme le mentionnait déjà l'Ordre national des pharmaciens en 2011, en effet, depuis quelques années, le marché des thérapies dites naturelles progresse et révèle un intérêt de plus en plus fort des patients à l'égard de la médication à base de plantes. Madagascar doit aussi travailler sur les contraintes normatives de ces produits. De même des travaux de recherches doivent être poursuivis par des chercheurs et autres experts.

Références

[1] Alghamdi H.A. (2021) A need to combat COVID-19; herbal disinfection techniques, formulations and preparations of human health friendly hand sanitizers. *Saudi Journal of Biological Sciences* ,28, 3943-3947

[2] Andriambelo M. (2014). *L'introduction des TIC dans l'enseignement supérieur : cas de la faculté DEGS*. Mémoire. Université d'Antananarivo. Madagascar.

[3] Chaachouay N., Douira D., Zidane L. (2021) COVID-19, prévention et traitement à base de plantes médicinales sur les marchés à base de plantes de la préfecture de Salé, nord-ouest du Maroc. *Revue européenne de médecine intégrative* 42 (2021)101285

[4] Dandouau A. (1922). *Ody et fanafody* (Charmes et remèdes) : pharmacopée Sakalave et Tsimihety). Editeur Paris.

[5] Debray, M, *et al*, (1975). *Contribution à l'inventaire des plantes médicinales de Madagascar.* Travaux et documents ORSTOM, n° 8, 1 Vol. 150 pp., Paris, 1971. Etudes Médicales- Mars 1975- N°1.

[6] Doré G. (2012). *Etude pour l'amélioration des services offerts aux étudiants de l'enseignement supérieur malgache.* Rapport final. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[7] Donat C. A. (2005) *Les conditions de développement d'une université, cas de l'Université de Toamasina*. Mémoire de maîtrise. Madagascar.

[8] Debuigne G., Couplan F., Folliard T. (2019) *Petit Larousse des plantes qui guérissent* Editeur Larousse, 1032 pages

[9] Feyaerts A.F. et Luyten W. (2020).Vitamin C as prophylaxis and adjunctive medical treatment for COVID-19. *Nutrition*79-80 , 1-4

[10] Fleurentin J., Wenige B. (2018) Un tour du monde des plantes qui soignent - Afrique, Amériques, Chine, Outremer, Europe. Editeur : Ouest France 240 pages ISBN: 9782737378997

[11] Giraud-Robert, A.M. (2005). Intérêt de l'aromathérapie dans la prise en charge des hépatites virales. *Phytothérapie* 3**,** 235–247

[12] Kras, E. (2020). Riposte de première ligne en santé communautaire contre le covid-19. *Santé Sud. Groupe SOS*.

[13] Lassak E.V. et McCarthy T. (1983) *Plantes médicinales australiennes* Editeur Methuen Australia, 240 p.,

[14] Limonier, Anne-Sophie, (2018). *La Phytothérapie de demain : les plantes médicinales au cœur de la pharmacie.* Thèse de doctorat. Faculté de Pharmacie Aix-Marseille Université. France.

[15] Musdja M.Y. (2021) Potential bangle (Zingiber montanum J.König) rhizome extract as a supplement to prevent and reduce symptoms of Covid-19 *Saudi Journal of Biological Sciences* 28(2021)2245-2243

[16] Ndiaye C. F. (2020). *Appel d'Urgence pour Covid 19- Madagascar*. juin-août 2020. Nations-Unies

[17] Ndoye. M. (2004). L'enseignement supérieur en Afrique : problématique des réformes.

[18] Ordre National des Pharmaciens, (2011). Plantes médicinales et médicaments à base de plantes. *Le Journal de l'Ordre National des Pharmaciens*, 04-juin 2011.

[19] Odukoya J., Odukoya J-O Ndinteh DT, (2021). Mesures élémentaires et évaluation des risques pour la santé des plantes médicinales d'Afrique subsaharienne utilisées pour le traitement des maladies cardiovasculaires et des facteurs de risque connexes

Journal of Trace Elements in Medicine and Biology 65(2021).

[20] Pauling L. et Roth B. (1988). *Abusez des vitamines* Edition Tchou 306 pages.

[21] Ralahady B. B. et Totohasina A. (2019a) *Experimental study of the valid rules according to the measure* M_{GK} in IJCST vol.3 N°.1.

[22] Ralahady B. B. et Totohasina A. (2019b) *ASI-MGK: Implicative Statistical Analysis tool based on* M_{GK} in IJCST vol.3 N°.1

[23] Ralahady, B. B., Totohasina, A. et Simon, J. (2019c). *Apports intéressants de l'indice asi-mgk et des courbes d'apprentissages en didactique de l'informatique*. Analyse Statistique Implicative, 10:304

[24] Razanatsoavina, C., Ralahady, B. B. et Laberch, J. C. (2019). *Influence d'une formation universitaire en agronomie sur le ressenti des traditions par les tudiants en premier cycle universitaire : cas du district de Mandritsara à Madagascar.* Analyse Statistique Implicative, 10:304.

[25] Ramiandrasoa D. (2005). in Ixe Symposium international d'aromathérapie et plantes médicinales, Grasse 2007. *Phytothérapie* 6**,** 47-55 (2008)

Annexe : Questionnaire

1. Numéro du questionnaire ……. date …….
2. Sexe? ☐Masculin ☐Féminin
3. Quel âge avez-vous? ……..
4. Vous êtes combien dans votre appartement /logement universitaire ?
   ☐3 à 9 ☐9 à 15 ☐15 à 20 ☐20 à plus
5. Dans quel quartier au campus universitaire résidez-vous?
   ☐Quartier Bloc ☐Quartier PV ☐Quartier PJ ☐Quartier Belle rose
6. Combien de borne fontaine/douche/WC dispose-t-il votre appartement ?
   Borne fontaine …….. Douche ……. WC……
7. Avez-vous déjà senti affecter par lecovid19 ?☐Oui ☐Non
8. Quels modes de prévention/traitement pratiquez-vous pour riposter contre le covid19 ?
9. Efficace? ☐ Oui ☐ Non
10. S'il y aura une campagne de vaccination, prêterez-vous à le faire?
    ☐ Oui ☐ Non ☐Aucune information
11. Si non, pourquoi?
    - ☐ Il a des effets secondaires
    - ☐ Il n'est pas rassuré
    - ☐ Il n'est pas efficace
    - ☐ Aucune information
12. Avez-vous senti des troubles causés par le covid-19 ?
    ☐ Oui ☐ Non ☐Aucune information
13. Si oui, dans quelle circonstance a-t-elle eu lieu ?
    - ☐ En relation amicale
    - ☐ En enseignement
    - ☐ En religion
    - ☐ En sport
    - ☐ Aux revenus financiers
    - ☐ Aucune information
14. Quel impact de Covid-19 avez-vous touché directement ?

UNE TECHNIQUE D'EXTRACTION DES RÈGLES D'ASSOCIATION QUANTITATIVES BASÉE SUR LE COEFFICIENT DE VARIATION

Josoa Michel Tovohery[1], André Totohasina[2], Daniel Rajaonasy Feno[3]

A METHOD BASED ON THE COEFFICIENT OF VARIATION FOR MINING QUANTITATIVE ASSOCIATION RULES

RÉSUMÉ

Notre papier propose d'extraire des règles d'association quantitatives de type $(X = \overline{X}) \Rightarrow (Y = \overline{Y})$, à partir d'une grande quantité de données, où $\overline{X}$ et $\overline{Y}$ sont respectivement les moyennes arithmétiques des variables quantitatives X et Y. Pour ce faire, nous présentons une technique basée sur l'usage des mesures purement statistiques où le coefficient de variation de Sharma joue le rôle de « Support » dans le cas qualitatif, et le rapport de corrélation comme mesures d'intérêt des règles d'association quantitatives. Sans faire de partitionnement des valeurs en plusieurs intervalles et en ne considérant que les valeurs brutes, notre méthode minimise ainsi la perte d'information.

Mots-clés : *Coefficient de variation, rapport de corrélation, règles d'association quantitatives.*

ABSTRACT

Our paper proposes to extract quantitative association rules of type $(X = \overline{X}) \Rightarrow (Y = \overline{Y})$, from a large amount of data, where $\overline{X}$ and $\overline{Y}$ are respectively the arithmetic means of quantitative variables X and Y, respectively. To do this, we present a technique based on the use of purely statistical measures where the Sharma coefficient of variation plays the role of "Support" in the qualitative case, and the correlation ratio as the measures of interest of the quantitative association rules. Without partitioning the values into several intervals and by considering only the gross values, our method minimizes the loss of information.

Keywords : *Coefficient of variation, correlation ratio, Quantitative association rule.*

1 Introduction

La fouille des règles d'association qualitative est un concept introduit par Agrawal *et al.* (1993). Ces chercheurs pionniers ont proposé d'étudier une base de données contenant des transactions d'un grand supermarché américain, afin d'aider le gérant de ce dernier à étudier les comportements de ses clients et à ranger la présentation des articles sur les étagères du supermarché pour avoir le maximum de profit en un temps record.

Le problème de la fouille des règles d'association se résume comme suit : soit D l'ensemble de toutes les transactions effectuées (ou tous les paniers des clients). Une transaction $t \in D$ est un ensemble d'articles (en anglais « itemset ») achetés par un client. Soient A l'ensemble de tous les articles dans le supermarché et X une partie de A : X est appelé un itemset, ou un ensemble d'articles, ou encore un motif. Une transaction $t \in D$ contient X, si t contient tous les articles dans X. Selon la définition donnée par Agrawal et ses collaborateurs, on appelle règle d'association une implication partielle de type $X \Rightarrow Y$,

[1] Ecole Doctorale Thématique « Science, Culture, Société et Développement » de l'Université de TOAMASINA, josoamicheltovohery@gmail.com

[2] ENSET- Université d'Antsiranana, andre.totohasina@gmail.com

[3] Faculté de Droit, d'Économie, de Gestion, et de Mathématiques, Informatique et Applications Université de TOAMASINA, fenodaniel2@yahoo.fr

où X et Y sont deux itemsets, tels que $X \cap Y = \emptyset$. Le Support d'un motif X est le nombre de transactions qui contiennent tous les articles dans X divisé par le nombre total de transactions dans D. Le support « s » de la règle $X \Rightarrow Y$ est le nombre des transactions dans D qui contiennent à la fois X et Y divisé par le nombre total de transactions dans D : on écrit $s = Supp(X \Rightarrow Y)$. La Confiance « c » de la règle $X \Rightarrow Y$ est le support de la règle $X \Rightarrow Y$ divisé par le support de X : en fait, c'est la probabilité conditionnelle sachant X de Y. On dit alors : « si X, alors Y », avec une confiance de $c \times 100\%$. Le Support et la Confiance sont souvent exprimés en pourcentage et concernent essentiellement l'analyse d'information contenue dans un contexte binaire ou qualitatif.

L'extraction des règles d'association quantitatives a été introduite pour la première fois par Gregory Piatetsky-Shapiro en 1991 selon Srikant et Agrawal (1996). L'auteur se propose implicitement d'extraire dans un grand volume de données une règle d'association quantitative de deux types : Tout d'abord, si $X = a$ alors $Y = b$, qui est notée comme $(X = a) \Rightarrow (Y = b)$, où a et b sont respectivement des valeurs particulières des attributs quantitatives X et Y. On peut prendre, par exemple, la règle fictive $(\text{Âge} = 30) \Rightarrow (\text{Nombre d}'\text{enfants} = 2)$. Ensuite, G. Piatetsky-Shapiro a aussi proposé d'extraire des règles de type $(X \in [a, b]) \Rightarrow (Y \in [c, d])$ pour les attributs dont les valeurs sont très nombreuses et éparpillées dans un intervalle (type continu). Le travail de Sujatha et Naveen (2011) illustre un exemple des travaux de Piatetsky-Shapiro pour le cas $(X \in [a, b]) \Rightarrow (Y \in [c, d])$, sans combiner les intervalles adjacents.

Plus tard, cette dernière approche a été reprise par Srikant et Agrawal (1996). Ils ont élaboré une démarche très sophistiquée afin de combiner les intervalles adjacents pour résoudre les deux problèmes que rencontrent la méthode de G. Piatetsky-Shapiro : problème de MinSupp et le problème de MinConf. C'est-à-dire, plus on a trop d'intervalles, plus les intervalles contiennent moins des valeurs et la plupart des règles ont des supports très faibles. Ensuite, plus on a peu d'intervalles, plus la confiance des règles est faible. Cette approche est combinée avec leur fameux algorithme « Apriori » découvert en 1994 dans Agrawal et Srikant (1994).

Aumann et Lindell (1999) présentent deux nouveaux types de règles d'association quantitatives qui sont plutôt basés sur une description d'un sous-ensemble de la population. Ces auteurs sont les premiers à utiliser les mesures de la statistique descriptive comme moyenne, variance et médiane dans le processus d'extraction des règles d'association. Ils ont proposé d'extraire dans un grand volume de données des règles de type $\text{Sous} - \text{ensemble de la population X} \Rightarrow \text{valeur moyenne de Y}$ en deux manières : premièrement, ils définissent une règle de type $X\,(\text{ensemble d}'\text{attributs qualitatifs}) \Rightarrow (Y = \overline{Y})$, par exemple : (Non-fumeur ET buveur du vin) ⇒ (Espérance de Vie = 85). Deuxièmement, ils définissent une règle de type $(X \in [a, b]) \Rightarrow (Y = \overline{Y})$. Citons, par exemple, les deux règles suivantes qu'ils ont trouvés : $(\text{Education} \in [2\,;13]\ \text{ans}) \Rightarrow$ (Revenu moyen $= 7.32$ \$ par heure), tandis que $(\text{Education} \in [14\,;18]\ \text{ans}) \Rightarrow$ (Revenu moyen $= 11.64$ \$ par heure). La notion de règle maximale a été inventée afin de trouver l'intervalle $[a\,;b]$ maximisant la moyenne $\overline{Y}$ de l'attribut quantitatif Y. Geoffrey I. Webb a essayé d'améliorer ce dernier type de règle proposé par Aumann et Lindell (1999) dans Webb (2001) en utilisant son algorithme OPUS et en introduisant une autre mesure de sélection des k règles plus intéressantes.

En tenant compte de ces quatre travaux pionniers, nous proposons d'extraire dans un grand volume de données quantitatives des règles d'association de type $(X = \overline{X}) \Rightarrow (Y = \overline{Y})$, qu'on interprète comme suit : « En général, si les clients achètent une quantité $\overline{X}$ de X, alors ils achèteraient aussi une quantité $\overline{Y}$ de Y ». On a une information générale sur

les comportements des clients, car une valeur moyenne $\overline{X}$ d'une variable quantitative X s'interprète comme une valeur générale représentant X, si X est homogène. Donc, afin de sélectionner des règle de tel type, il est évident qu'on doit se poser d'abord une contrainte de dispersion des valeurs de tous les attributs quantitatifs étudiés. Pour ce faire, nous proposons d'utiliser la mesure de dispersion normalisée et symétrique de Sharma *et al.*(2011), notée CVS, et on laisse à l'utilisateur de spécifier son seuil MaxCVS afin d'élaguer les attributs quantitatifs dont les valeurs sont assez dispersées (donc non homogènes). Puis, nous utilisons le rapport de corrélation, afin de quantifier la dépendance fonctionnelle et statistique entre deux attributs quantitatifs à la place des mesures de qualité des règles d'association.

Le reste de cet article est organisé comme suit : la section 2 présentera le détail de notre problème, notre technique d'extraction des règles d'association quantitative, les justifications de notre approche et les méthodologies des expérimentations effectuées ; la section 3 partagera les résultats trouvés lors des traitements des données réelles effectués ; la section 4 discutera les avantages, les limites de notre approche et les liaisons de ce présent travail avec les autres résultats déjà publiés dans la littérature ; la section 5 termine notre exposé par une conclusion.

2 Méthodes

Dans cette section, nous présentons notre approche et les données que nous allons traiter à titre d'expérimentation. Dans cette sous-section, nous allons présenter notre problème, notre approche lors de la résolution et sa justification.

2.1 Position de problème

Une fois qu'un client vienne dans un supermarché, il achète certains articles avec leurs unités de mesures respectives, que ceux – ci soient des nombres ou masses ou longueurs ou volumes, etc. Afin de bien gérer le calcul de bénéfice et le nombre de stock restant, il est important de mémoriser les transactions effectuées. Dans ce cas, on est obligé de stocker des nombres réels. À titre indicatif, nous voyons dans le Tableau 1 un exemple pratique de cet enregistrement (Tableau 1).

N° de la transaction	Huiles (litres)	Sucres (kg)	Savons (unité)	Tomates (unité)	Oeufs (unité)
1	1	0	0	3	4
2	0.5	0.5	0	5	0
3	0	1	3	0	1
4	1.5	1	2	2	1
5	2	2	0	3	3
6	0	2	0	3	2
7	1	1.5	1	0	0
8	0.75	0	2	1	3
9	1	0	3	0	2
10	1	0	1	3	2

Tableau 1 – Exemple de tableau d'enregistrement des transactions (Données fictives)

Afin d'avoir un maximum de profit, le gérant du supermarché s'intéresse, à la fois, à la découverte des articles les plus achetés par les clients (problème de la recherche d'itemsets fréquents), à la consommation moyenne des clients pour chaque article (problème de la statistique descriptive), à l'homogénéité des consommations des clients, à la connaissance des groupes d'articles que les clients achètent souvent ensemble afin de mieux ranger

l'emplacement des articles sur les étagères, à la connaissance de tous les articles courant un risque lorsqu'on cesse de vendre un article X, et surtout à la connaissance de la quantité moyenne u des articles Y vendus par un client sachant qu'il a déjà acheté une quantité de v unités de l'article X, etc.

Dans notre étude, nous proposons d'extraire des règles de type $(X = \overline{\mathrm{X}}) \Rightarrow (\mathrm{Y} = \overline{\mathrm{Y}})$ afin de répondre à ces besoins. Par exemple, supposons qu'on ait la règle d'association R1 = $(\mathrm{Tomate} = 3\ \mathrm{unités}) \Rightarrow (\mathrm{œufs} = 2\ \mathrm{unités})$. On peut dire directement que, en générale, si les clients achètent 3 Tomates, alors ils achèteraient 2 œufs. Ainsi, on a tout de suite les consommations moyennes des articles présents dans une règle valide. D'une part selon notre approche, si X et Y apparaissent dans une règle, alors ce sont tous des articles que les clients consomment souvent. De plus, un motif ne serait pas sélectionné si ces valeurs ne satisfait pas à la contrainte de dispersion MaxCVS, donc chaque motif d'une règle valide possède des valeurs homogènes selon la spécification de l'utilisateur. Ensuite, cette règle nous informe qu'il faut mettre l'étagère contenant les « Tomates » à côté de celui des « œufs », car il devient aussi plus probable que le client qui achète des tomates, achètera aussi une certaine quantité des œufs. Enfin, nous connaissons à partir de cette règle que si l'on arrête de vendre des « tomates », alors la vente « d'œufs » risquerait un problème, car les clients ont tendance d'acheter ailleurs, là où il y a tout pour économiser le temps d'achat et les dépenses en déplacement. D'où l'intérêt d'extraire des règles d'association quantitatives de type $(X = \overline{\mathrm{X}}) \Rightarrow (\mathrm{Y} = \overline{\mathrm{Y}})$.

2.2 Notre approche

Dans cette section, nous allons donner la présentation formelle de notre problème et les étapes algorithmiques de résolution. De plus, nous allons nous servir du Tableau 1 pour présenter un exemple pratique.

Soit D un tableau des données numériques, tel que les lignes représentent les achats que fait un client, et les colonnes représentent les articles vendus par le supermarché. Le Tableau 2 illustre la forme du tableau D (Tableau 2).

N° de la transaction	Article 1 : X_1	Article 2 : X_2	⋯	Article j : X_j	⋯	Article m : X_m
t_1	x_{11}	x_{12}	⋯	x_{1j}	⋯	x_{1m}
t_2	x_{21}	x_{22}	⋯	x_{2j}	⋯	x_{2m}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
t_i	x_{i1}	x_{i2}	⋯	x_{ij}	⋯	x_{im}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
t_n	x_{n1}	x_{n2}	⋯	x_{nj}	⋯	x_{nm}

Tableau 2 – Le tableau des données numériques ***D***

La valeur x_{ij} représente ici la quantité d'article X_j achetée lors de la transaction t_i. Les étapes suivantes montrent comment faire pour traiter ce tableau numérique D afin de découvrir les comportements généraux des clients du supermarché propriétaire de ces données.

2.2.1 Construction d'une matrice intermédiaire

C'est une construction d'une matrice booléenne afin d'élaguer dans notre étude les articles qui n'intéressent pas trop les clients, c'est-à-dire les articles qui se vendent plutôt rarement, car un tel article intéresserait seulement un groupe particulier des clients, mais non pas la généralité. Pour ce faire, on utilise la fonction booléenne suivante :

$$\forall\, x_{ij} \in D,\ Bool(x_{ij}) = \begin{cases} 0, si\ x_{ij} = 0\ ; \\ 1, sinon. \end{cases} \qquad (1)$$

Après cette transformation intermédiaire, on a le tableau D', tel que si $x_{ij} \in D'$, alors $x_{ij} = 0$ ou $x_{ij} = 1$ (Tableau 3).

N° de la transaction	Huiles (litres)	Sucres (kg)	Savons (unité)	Tomates (kg)	Oeufs (unité)
1	1	0	0	1	1
2	1	1	0	1	0
3	0	1	1	0	1
4	1	1	1	1	1
5	1	1	0	1	1
6	0	1	0	1	1
7	1	1	1	0	0
8	1	0	1	1	1
9	1	0	1	0	1
10	1	0	1	1	1

Tableau 3 – Tableau booléen obtenu à partir du Tableau 1

2.2.2 Usage de la matrice intermédiaire *D'*

Nous faisons usage de la matrice intermédiaire D' pour le calcul de $support$ de X_j , $j = \overline{1, m}$. À partir de D', on a :

$$Support\ (X_j) = \frac{\sum_{i=1}^{n} bool(x_{ij})}{n}. \qquad (2)$$

Le Tableau 4 illustre les supports des articles présentés dans le Tableau 1.

Article	Huiles	Sucres	Savons	Tomates	Œufs
$Support$	0.8	0.6	0.6	0.7	0.8

Tableau 4 – Support des articles figurant dans le Tableau 1

2.2.3 Premier élagage :

Nous procédons à l'élagage des articles vendus rarement ($itemests$ non fréquents) à partir de la contrainte de $MinSup$. Ici, on garde seulement les articles (1-itemset) dont la fréquence de vente est plus grande que le paramètre $MinSup$ fixé par l'utilisateur. Donc, l'article X_j est retenu si $Support(X_j) > MinSup$.

Soit $DF1$ l'ensemble des articles X_j dont la fréquence de vente $Support(X)$ est plus grand que $MinSup$. Exemple, pour $MinSup = 0.6$, les articles « Sucres » et « Savons » sont élagués. Nous revenons à D mais seulement avec les éléments de $DF1$. Nous procédons au calcul des coefficients de variation de Sharma CVS des $1 - itemsets\ X_j$ fréquents. Soient $\forall\, X_j \in DF1,\ X_j' = \{x_{ij} \in X_j \,|x_{ij} \neq 0\}$ et $n_j = Card(X_j')$. On note par x_{ij}' les éléments de X_j'. Le coefficient de variation de Sharma *et al.* (2011) d'une variable quantitative X_j est :

$$CVS(X_j) = \begin{cases} \dfrac{\sigma(X_j')}{\sqrt{[Max(X_j') - \overline{X_j'}]\,[\overline{X_j'} - Min(X_j')]}}, si\ Max(X_j') \neq \overline{X_j'}\ et\ Min(X_j') \neq \overline{X_j'}; \\ 0, sinon. \end{cases} \qquad (3)$$

avec $\overline{X_j'} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}'$ et $\sigma(X_j') = \sqrt{\frac{1}{n_j}\sum_{i=1}^{n_j}(x_{ij}' - \overline{X_j'})^2}$.

On rappelle que, dans notre approche, le calcul de $CVS(X_j)$ et la moyenne $\overline{X_j'}$ est appliqué seulement sur les valeurs x_{ij} non nulles de X_j. Le Tableau 5 montre les CVS des articles satisfaisants le $MinSupp = 0.6$ (Tableau 5).

Items fréquents	Huiles	Tomates	Œufs
CVS	0.5887635	0.5638694	0.6546537

Tableau 5 – CVS des articles les plus consommés par les clients.

2.2.4 Deuxième élagage

Nous procédons à deuxième élagage des $itemsets$ fréquents dont les valeurs sont très dispersées en utilisant le paramètre $MaxCVS$. L'article X_j est retenu si $CVS(X_j)$ est plus grand que $MaxCVS$. Soit $DF2$ l'ensemble des articles fréquents et dont les consommations des clients sont homogènes. Exemple : pour $MaxCVS = 0.7$, on a : $DF2 = \{Huiles ; Tomates ; Oeufs\}$.

2.2.5 Retour dans *D* et troisième élagage.

Calcul des rapports de corrélation $\eta_X(Y)$ et $\eta_Y(X)$ pour tout X,$Y \in DF2$, tels que $X \cap Y = \emptyset$. En même temps, retenir la règle : $X \Rightarrow Y$ si $\eta_X(Y) > \eta_Y(X)$. Si $\eta_X(Y) = \eta_Y(X)$, alors on a une implication de type $X \Leftrightarrow Y$. Nous allons approfondir cette étape avec une théorie formelle et un exemple de calcul détaillé lors de la justification de ce choix de mesure.

Remarque : Lors du calcul de $\eta_Y(X)$ et $\eta_X(Y)$, on utilise toutes les valeurs de X et Y, sans exception (contrairement au calcul de CVS à l'étape 4).

2.2.6 Quatrième élagage

On peut faire deux types d'élagage selon le critère choisi par l'utilisateur :

(a) Elagage des règles $X \Rightarrow Y$ dont le rapport corrélation $\eta_X(Y) < MinRapCor$, avec $MinRapCor$ est le seuil minimum de rapport de corrélation (voir page 11 et 12).
(b) Retenir les r tops règles d'association intéressantes, c'est-à-dire les r premières règles possédant le rapport de corrélation le plus élevé.

On procède en suite à l'affichage du résultat en tenant compte de la forme $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$, et Fin.

La Figure 1 illustre les règles d'association de type $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$ obtenues par les données du Tableau 1 pour $MinSupp = 0.6$, $MaxCVS = 0.7$ et $MinRapCor = 0.4$.

```
> Resultat
                 Premisse Implication          Consequence Rapport.Correlation
1    Tomates : moyenne = 2          => Oeufs : moyenne = 1.8          0.6955128
2 Huiles : moyenne = 0.875          => Tomates : moyenne = 2          0.4807692
3 Huiles : moyenne = 0.875          => Oeufs : moyenne = 1.8          0.4551282
>
```

Figure 1 – Résultat du traitement des données du Tableau 1 pour $MinSupp = 0.6$, $MaxCVS = 0.7$ et $MinRapCor = 0.4$.

2.3 Justifications de notre approche

Tout d'abord, nous avons décidé de traiter les valeurs réelles car la transformation booléenne fait perdre des informations, alors qu'en statistique on s'intéresse plutôt au maximum d'information.

Quantité d'information — La quantité d'information I contenue dans le tableau de contingence de X et Y est égale à la somme de la variance de X et de la variance de Y (trace de la matrice de variance – covariance) (Totohasina (2008)). On a :

$$I = V(X) + V(Y) \qquad (4)$$

En effet, pour deux variables aléatoires réelles X et Y prenant respectivement des valeurs positives ou nulles dans $X(\Omega) = \{x_1, \dots, x_n\}$ et $Y(\Omega) = \{y_1, \dots, y_m\}$; $n_{i\,.}$ et $n_{.j}$ étant les effectifs marginaux respectifs, on a la quantité d'information :

$$I_{v.réelle} = \frac{1}{N}\left(\sum_{i=1}^{n} n_{i\,.}.x_i^2 + \sum_{i=1}^{m} n_{.j}.y_j^2\right) - \frac{1}{N^2}\left[\left(\sum_{i=1}^{n} n_{i\,.}.x_i\right)^2 + \left(\sum_{i=1}^{m} n_{.j}.y_j\right)^2\right] \qquad (5)$$

où $N = \sum_i^n n_{i\,.} = \sum_j^m n_{.j}$.

Toutefois, si X et Y sont binaires, alors la quantité d'information se réduit à :

$$I_{v.binaire} = \frac{n_X + n_Y}{N} - \frac{n_X^2 + n_Y^2}{N^2} \qquad (6)$$

Cependant, on a : $n_X \leq \sum_{i=1}^{n} n_{i\,.}.{x_i}^2$ et $n_X^2 \leq \left(\sum_{i=1}^{n} n_{i\,.}.x_i\right)^2$. Donc, on a : $\frac{n_X - n_X^2}{N} \leq \frac{1}{N}\sum_{i=1}^{n} n_{i\,.}.{x_i}^2 - \left(\frac{1}{N}\sum_{i=1}^{n} n_{i\,.}.x_i\right)^2$. C'est-à-dire, $V(X_{binaire}) \leq V(X_{réelle})$. D'où, on a : $I_{v.binaire} \leq I_{v.réelle}$.

Lemme — On perd généralement de l'information en réduisant deux variables effectivement à valeurs réelles positives ou nulles, de moyenne marginales strictement supérieures à 1, en deux variables binaires (Totohasina (2008)).

Ensuite, on a choisi le coefficient de variation à la place du support, car si l'on considère le cas binaire où le support $Supp(X) = f_X = \frac{n_X}{n}$, alors le coefficient de variation de Pearson est : $CV(X) = \frac{\sqrt{f_X(1-f_X)}}{f_X}$. On a donc la relation suivante :

$$CV(X) < \alpha \Leftrightarrow f_X > \frac{1}{1+\alpha^2} \qquad (7)$$

pour tout nombre réel positif α.

Par exemple, si $CV(X) < 1$, alors $f_X > 0.5$. Ainsi, la contrainte de maximum de coefficient de variation représente la contrainte de $MinSupp$ dans le cas binaire.

Par la suite, nous avons choisi de remplacer le coefficient de variation de Pearson par le coefficient de variation de Sharma, car ce dernier est la meilleure mesure de dispersion (Sharma *et al.* (2011)). En effet, à l'instar des travaux de Lallich et Teytaud (2004) et les travaux de Grissa (2013) sur les critères d'appréciation des mesures de qualité des règles d'association, nous proposons d'énumérer d'abord les critères que les chercheurs souhaitent avoir sur une mesure de dispersion, puis nous démontrons que la mesure CVS est la meilleure mesure de dispersion de la littérature.

2.3.1 Quelques critères d'appréciation d'une mesure de dispersion :

Mesure possédant un sens concret — Selon Lallich et Teytaud (2004), il est important de travailler avec une mesure possédant un sens parlant pour l'utilisateur. On a : pour toute

mesure de dispersion M, $C_1(M) = 1$ si la mesure M possède un sens concret ; sinon $C_1(M) = 0$.

Mesure dépourvue d'unité — Afin que la comparaison des dispersions des deux variables aléatoire X et Y par la mesure M ait du sens, il faut que la valeur fournie par M soit dépourvue d'unité. Ainsi, on adopte la règle suivante : $C_2(M) = 1$, si $M(X)$ n'admet pas d'unité ; sinon, $C_2(M) = 0$ (Sørensen (2002)).

Mesure normalisée — Une mesure de dispersion est dite normalisée si sa valeur est comprise entre 0 et 1. Totohasina (2008) a affirmé que la normalisation des mesures des règles d'association consiste à ramener les valeurs d'une mesure de qualité sur l'intervalle $[-1\ ;\ 1]$. Cependant, toutes les mesures de dispersion existantes dans la littérature sont des mesures positives. Ainsi, la normalisation d'une mesure de dispersion consiste à ramener la valeur de cette dernière sur l'intervalle $[0\ ;\ 1]$. Par conséquent, on a : pour toute mesure M, $C_3(M) = 1$ si pour toute variable aléatoire réelle X, $0 \leq M(X) \leq 1$; sinon, $C_3(M) = 0$.

Mesure donnant une valeur nulle pour une variable aléatoire à valeurs constantes — Une mesure de dispersion doit être capable de détecter au moins le cas d'une parfaite stabilité. Par conséquent, $C_4(M) = 1$, si pour toute variable aléatoire réelle constante X, $M(X) = 0$; sinon $C_4(M) = 0$. C'est-à-dire, on souhaite que $M(X) = 0$, si pour tout échantillon de n valeurs de X, on a : $x_1 = x_2 = \cdots = x_n$.

Mesure stable par une translation linéaire — Une mesure de dispersion doit évaluer par une même valeur la variable X et la variable $Y = X + a$, où a est une constante réelle. Dès lors, on a : $C_5(M) = 1$, si $\forall\ a \in R, M(X + a) = M(X)$; sinon, $C_5(M) = 0$ Sørensen (2002).

Mesure stable par une multiplication avec un scalaire — Dès fois, en démographie, on est obligé de changer les unités des valeurs obtenues. Citons par exemple, la conversion des unités des monnaies en dollar. Cependant, on veut que la conversion des unités n'affecte pas la dispersion des valeurs. Ainsi, on veut que la mesure de dispersion soit stable par une multiplication par un scalaire : $C_6(M) = 1, si\ \forall\ k \in R^*, M(kX) = M(X)$; sinon $C_6(X) = 0$ (Sørensen (2002)).

Mesure retournant deux valeurs identiques pour deux séries à valeurs symétriques — Selon Sharma *et al.* (2011), une mesure de dispersion ne doit pas favoriser les séries à faibles valeurs ou celles qui ont des valeurs plus élevées. Ainsi, pour évaluer ce critère, on teste la mesure avec plusieurs couples de séries à valeurs symétriques. Prenons, par exemple, le cas des deux séries proposées par Sharma suivantes : $S_1 = \{1\ ;3\ ;11\}$ et $S_2 = \{1\ ;9\ ;11\}$. Ces deux séries ont des valeurs symétriques par rapport à $x = 6$. Une bonne mesure de dispersion doit renvoie une même valeur pour chacune de ces séries. Par conséquent, on a : $C_7(M) = 1$, si M renvoie deux valeurs identiques pour deux séries à valeurs symétriques ; sinon, $C_7(M) = 0$.

Mesure pouvant être utilisée avec des valeurs négatives et positives — Une telle mesure peut être utilisée pour évaluer la variation de toutes les variables aléatoires réelles. En effet, bien qu'une mesure soit jugée performante, le fait qu'elle pose de restriction sur le signe des valeurs engendre un grand problème sur son usage. Ainsi, on souhaite travailler avec une mesure de dispersion définie pour tous les signes des valeurs. On a donc : $C_8(M) = 1$, si la mesure M est définie pour des valeurs négatives et positives ; sinon, $C_8(M) = 0$.

Actuellement, on compte déjà onze mesures de dispersion. Le tableau 6 suivant illustre leurs appellations avec leurs expressions.

N°	Mesures de dispersion	Expressions
1	Variance	$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$
2	Ecart-type	$S(X) = \sqrt{V(X)}$
3	Étendu ou Empan	$Empan(X) = \max_{1\leq i\leq n} x_i - \min_{1\leq i\leq n} x_i$
4	Ecart moyen	$E.M(X) = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \bar{X}\rvert$
5	Ecart médian	$E.med(X) = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - Med(X)\rvert$
6	Ecart géométrique $E.géom$	$\log[E.géom(X)] = \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \log[G(X)])^2$ avec $G(X) = \sqrt[n]{\prod_{i=1}^{n} x_i}$
7	Intervalle interquartile	$IQ(X) = Q_3 - Q_1$ où Q_1 et Q_3 sont respectivement les quartiles 25% et 75%.
8	Différence moyenne	$d(X) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert x_i - x_j\rvert$
9	Coefficient de variation relative	$Q_r(X) = \frac{Q_3 - Q_1}{Med(X)}$ (1) ou $Q_r(X) = \frac{Q_3 - Q_1}{Q_3 + Q_1}$ (2)
10	Coefficient de variation de variation (de Pearson)	$CV(X) = \frac{S(X)}{\bar{X}}$
11	Coefficient de variation de Sharma	$CVS(X) = \begin{cases} \frac{s(X)}{\sqrt{[Max(X) - \bar{X}\,][\bar{X} - Min(X)]}} \\ si\ Max(X) \neq \bar{X} et\ Min(X) \neq \bar{X}\ ; \\ 0, \quad sinon. \end{cases}$

Tableau 6 – Liste des mesures de dispersion existantes dans la littérature

Le tableau 7 suivant montre les propriétés des onze mesures de dispersion recensées dans la littérature. Nous avons exclu le critère 1, car cela est jugé trop subjectif.

	C_2	C_3	C_4	C_5	C_6	C_7	C_8	Total
$V(X)$	0	0	1	1	0	1	1	4
$S(X)$	0	0	1	1	0	1	1	4
$Empan(X)$	0	0	1	1	0	1	1	4
$E.M(X)$	0	0	1	1	0	1	1	4
$E.med(X)$	0	0	1	1	0	1	1	4
$E.géom(X)$	0	0	1	0	1	0	0	2
$IQ(X)$	0	0	1	1	0	1	1	4
$d(X)$	0	0	1	1	0	1	1	4
$Q_r(X)$ (1)	1	0	1	0	1	0	1	4
$Q_r(X)$ (2)	1	1	0	1	1	0	1	5
$CV(X)$	1	0	1	0	1	0	0	3
$CVS(X)$	1	1	1	1	1	1	1	7

Tableau 7 – évaluation des 11 mesures de dispersion existantes dans la littérature.

Ce tableau montre que le coefficient de variation de Sharma est la meilleure mesure de dispersion, car il est le seul vérifiant tous les critères.

Remarques — La mesure $CVS(X)$ prend exactement ses valeurs entre $[0\,;1]$. $CVS(X)$ prend une valeur nulle pour une variable aléatoire constante X. Elle prend une valeur égale à 1 pour une série contenant seulement deux valeurs différentes x et y, mais de même effectif. Prenons, par exemple, le cas de la série $X = \{1\,;\,1\,;\,1\,;\,5\,;\,5\}$. Dans ce cas, on a : $CVS(X) = 1$. La deuxième expression du coefficient de variation relative prend une valeur égale à 1, si et seulement si $Q_1 \neq Q_3$ et $Q_1 = 0$.

Enfin, nous allons énumérer les belles propriétés de la mesure rapport de corrélation afin de justifier le choix de notre mesure d'intérêt d'une règle quantitative valide.

Rapport de corrélation — On considère deux variables aléatoires X et Y. Soient $X = \{x_1\,;\,x_2\,;\ldots\,;\,x_r\}$ et $Y = \{y_1\,;\,y_2\,;\ldots\,;\,y_s\}$. Ainsi, on a le tableau de contingence des effectifs illustrés par le tableau 8 suivant :

Y / X	y_1	y_2	…	y_j	…	y_s	TOTAL
x_1	$n_{1,1}$	$n_{1,2}$	…	$n_{1,j}$	…	$n_{1,s}$	$n_{1,.}$
x_2	$n_{2,1}$	$n_{2,2}$	…	$n_{2,j}$	…	$n_{2,s}$	$n_{2,.}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
x_i	$n_{i,1}$	$n_{i,2}$	…	$n_{i,j}$	…	$n_{i,s}$	$n_{i,.}$
⋮	⋮	⋮	…	⋮	…	⋮	⋮
x_r	$n_{r,1}$	$n_{r,2}$	…	$n_{r,j}$	…	$n_{r,s}$	$n_{r,.}$
TOTAL	$n_{.,1}$	$n_{.,2}$	…	$n_{.,j}$	…	$n_{.,s}$	N

Tableau 8 – Tableau de contingence d'effectifs de X et Y

(a) On définit les moyennes marginales de X et Y par :

$$\mu(X) = \bar{X} = \frac{1}{N}\sum_{i=1}^{r} n_{i,.}\, x_i \quad et \quad \mu(Y) = \bar{Y} = \frac{1}{N}\sum_{j=1}^{s} n_{.,j}\, y_j. \qquad (8)$$

(b) On définit la variance marginale de X et Y par :

$$\sigma^2(X) = \frac{1}{N}\sum_{i=1}^{r} n_{i,.}\,(x_i - \bar{X})^2 \quad et \quad \sigma^2(Y) = \frac{1}{N}\sum_{i=1}^{s} n_{.,j}\left(y_j - \bar{Y}\right)^2. \qquad (9)$$

(c) On définit la moyenne conditionnelle sachant $(X = x_i)$ de Y par :

$$\mu_{(X=x_i)}(Y) = \sum_{j=1}^{s} \frac{n_{i,j}}{n_{i,.}} y_j. \qquad (10)$$

(d) On définit la variance conditionnelle sachant $(X = x_i)$ de Y par :

$$\sigma^2{}_{(X=x_i)}(Y) = \sum_{j=1}^{s} \frac{n_{i,j}}{n_{i,.}}\left(y_j - \mu_{(X=x_i)}(Y)\right)^2. \qquad (11)$$

Remarques — $\mu_X(Y)$ et $\sigma^2{}_X(Y)$ sont des variables aléatoires sur $X(\Omega) = \{x_1\,;\ldots\,;\,x_r\}$. On a :

$$\mu_{X(\Omega)}(Y) = \left\{\mu_{(X=x_1)}(Y)\,;\ldots;\,\mu_{(X=x_r)}(Y)\right\} \text{ et } \sigma^2{}_{X(\Omega)}(Y) = \left\{\sigma^2{}_{(X=x_1)}(Y)\,;\ldots\,;\,\sigma^2{}_{(X=x_r)}(Y)\right\}$$

De plus, l'application $\mu_X(Y) : (X = x_k) \mapsto \mu_{(X=x_k)}(Y) = h(x_k)$ est la régression de Y en X. On estime Y par : $Y \simeq h(X)$. Dans ce cas, h n'est pas nécessairement linéaire.

(e) **Théorème de la variance totale —** Soient X et Y deux variables aléatoires réelles, telles que $\sigma^2(X)$ et $\sigma^2(Y)$ existent. On a alors:
$\sigma^2(Y) = \mu\left(\sigma^2{}_X(Y)\right) + \sigma^2\left(\mu_X(Y)\right)$.

Remarques — Si $\sigma^2(Y)$ est non nul, alors on a :

$$\frac{\mu\left(\sigma^2{}_X(Y)\right)}{\sigma^2(Y)} + \frac{\sigma^2(\mu_X(Y))}{\sigma^2(Y)} = 1. \qquad (12)$$

(f) On définit le rapport de corrélation de Y en X par le nombre réel positif ou nul $\eta_X(Y)$ défini par son carré :

$$\eta^2{}_X(Y) = \frac{\sigma^2(\mu_X(Y))}{\sigma^2(Y)}. \qquad (13)$$

$\eta^2{}_X(Y) \times 100\%$ est le pourcentage de la variance totale expliquée par l'éventuelle régression de Y en X.

(g) On définit le coefficient de liberté de Y par rapport à X par le nombre réel positif ou nul $L_X(Y)$ défini par son carré :

$$L^2{}_X(Y) = \frac{\mu\left(\sigma^2{}_X(Y)\right)}{\sigma^2(Y)}. \qquad (14)$$

$L^2{}_X(Y)$ exprime ainsi la proportion de la variance marginale de Y non expliquée par l'éventuelle régression de Y en X.

Propositions — Voici quelques propriétés évidentes reprises dans les travaux de Totohasina (2008) pages 80-81 :

1. $0 \leq \eta_X(Y) \leq 1$ et $0 \leq L_X(Y) \leq 1$ (mesure normalisée).

2. En générale, on a : $\eta_X(Y) \neq \eta_Y(X)$ et $L_X(Y) \neq L_Y(X)$. C'est-à-dire, ces deux mesures ne sont pas symétriques dans les éventuelles dépendances ou indépendances.

3. $\eta_X(Y) = 1$ $(ou\ L_X(Y) = 0) \Leftrightarrow$ (Il existe une dépendance fonctionnelle continue de Y relativement à X). Dans ce cas, on dit que Y est totalement dépendante de X. Cette relation de dépendance n'est pas nécessairement linéaire.

4. $L_X(Y) = 1$ $(ou\ \eta_X(Y) = 0)$ n'entraîne pas nécessairement que Y soit indépendante de X. On dit simplement que Y est totalement libre de X. C'est-à-dire, la connaissance de X n'apporte aucune information sur Y.

5. En général, le rapport de corrélation de Y en X est toujours supérieur ou égal à la valeur absolue du coefficient de corrélation linéaire de Y et X : $\eta_X(Y) \geq |corrélation(X,Y)|$.

6. Si $\eta_X(Y) = |corrélation(X,Y)|$ ou $\eta_X(Y) \approx |corrélation(X,Y)|$, alors il y a régression linéaire ou quasi-linéaire de Y en X.

7. S'il existe deux réels $a\ et\ b$, tels que $Y = aX + b$ et a non nul, alors $\eta_X(Y) = |corrélation(X,Y)| = 1$.

Selon ces propriétés, le rapport de corrélation, qui n'est pas symétrique, est meilleur que le coefficient de corrélation, car il détecte non seulement la dépendance linéaire, mais toute autre dépendance fonctionnelle possible. D'où le choix de la mesure rapport de

corrélation lors de l'extraction des règles d'association quantitatives (sans transformation en valeur booléenne).

Le Tableau 9 montre les valeurs des $\eta_{(X=x_i)}(Y)$ et $\eta_{(Y=y_i)}(X)$ pour $X =$ »$Tomates$» et $Y =$ »$Oeufs$».

$X = Tomates$ \ $Y = Oeuf$	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$y_4 = 3$	$y_5 = 4$	Total	$\eta_{(X=x_i)}(Y)$
$x_1 = 0$	1	1	1	0	0	3	1
$x_2 = 1$	0	0	0	1	0	1	3
$x_3 = 2$	0	1	0	0	0	1	1
$x_4 = 3$	0	0	2	1	1	4	2.75
$x_5 = 5$	1	0	0	0	0	1	0
Total	2	2	3	2	1	10	-
$\eta_{(Y=y_i)}(X)$	2.5	1	2	2	3	-	-

Tableau 9 – Valeur des $\eta_{(X=x_i)}(Y)$ et $\eta_{(Y=y_i)}(X)$ pour X = »Tomates» et Y = »œufs».

D'après la ligne de $(X = 0)$ du Tableau 8, on a :

$$\eta_{(X=0)}(Y) = \frac{(1 \times 0) + (1 \times 1) + (1 \times 2) + (0 \times 3) + (0 \times 4)}{1 + 1 + 1 + 0 + 0} = \frac{3}{3} = 1$$

D'après la colonne de $(Y = 2)$ du Tableau 8, on a :

$$\eta_{(Y=2)}(X) = \frac{(1 \times 0) + (0 \times 1) + (0 \times 2) + (2 \times 3) + (0 \times 5)}{1 + 0 + 0 + 2 + 0} = \frac{6}{3} = 2$$

Calcul de $\mu\left(\eta_{(X=x_i)}(Y)\right)$:

$$\mu\left(\eta_{(X=x_i)}(Y)\right) = \frac{3 \times \eta_{(X=0)}(Y) + 1 \times \eta_{(X=1)}(Y) + 1 \times \eta_{(X=2)}(Y) + 4 \times \eta_{(X=3)}(Y) + 1 \times \eta_{(X=5)}(Y)}{3 + 1 + 1 + 4 + 1}$$

On a : $\mu\left(\eta_{(X=x_i)}(Y)\right) = 1.8$, qui est vraiment égale à la consommation moyenne des œufs, car $\bar{Y} = 1.8$. Les pondération 1, 3 et 4 sont les effectifs marginaux des x_i.

Calcul de la variance $\sigma^2\left(\eta_{(X=x_i)}(Y)\right)$:

$$\sigma^2\left(\eta_{(X=x_i)}(Y)\right) = \frac{1}{N}\sum_{i=1}^{r} n_{i,.}\eta^2{}_{(X=x_i)}(Y) - \bar{Y}^2. \qquad (15)$$

Donc, on a : $\sigma^2\left(\eta_{(X=x_i)}(Y)\right) = \frac{(3\times1^2)+(1\times3^2)+(1\times1^2)+(4\times2.75^2)+(1\times0^2)}{10} - 1.8^2 = 1.085.$
Cependant, la variance de Y est : $\sigma^2(Y) = 1.56$ (calculer à partir de toutes les valeurs de Y sans exception).

Ainsi, on a le rapport de corrélation :

$$\eta_X(Y) = \frac{\sigma^2\left(\eta_{(X=x_i)}(Y)\right)}{\sigma^2(Y)} = \frac{1.085}{1.56} = 0.6955128$$

De la même façon, on a : $\eta_Y(X) = 0.1346154$.

Comme $\eta_Y(X) < \eta_X(Y)$, alors la règle $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$ est éliminée à l'étape 6 pour éviter la redondance des règles. De plus, $\eta_X(Y) > MinRapCor = 0.4$, alors la règle $\bar{X} \Rightarrow \bar{Y}$ est retenue. Ainsi, nous affichons au résultat la règle :

$(Tomate = 2) \Rightarrow (Oeufs = 1.8)$ $avec$ $\eta_X(Y) = 0.695$, valeur très élevée.

Interprétation — En générale, si les clients achètent 3 tomates, alors ils achèteront 2 œufs.

3 Matériels

Dans cet article, nous avons utilisé le logiciel R version 4.0.4 pour la programmation et les traitements de nos jeux de données. Ensuite, nous avons expérimenté notre méthode avec les notes des 2351 élèves série A2 de la région DIANA – Madagascar, candidats au baccalauréat de l'année 2014. Notre jeu de données contient la moyenne générale et les notes des matières suivantes :

- Langue : Malagasy, Français ;
- Langue vivante au choix : Anglais ou Allemand ou Russe ou Espagnol ;
- Histoire Géographie ;
- Philosophie ;
- Matières scientifiques : Mathématiques générales, Physique – chimie et Science de la vie et de la terre ;
- Education Physique et Sportive (E.P.S).

On rappelle que, selon le système éducatif de Madagascar de l'année 2014, les élèves candidats au baccalauréat sont répartis en 04 séries :

- Série A1 : pour les élèves purement littéraires (mathématiques coefficient 01, les deux autres matières SVT et Physique chimie sont facultatives) ;
- Série A2 : pour les élèves littéraires mais qui s'intéressent aux mathématiques (mathématiques coefficient 03, les deux autres matières SVT et Physique chimie sont facultatives) ;
- Série C : pour les élèves purement scientifiques, forts en mathématiques et en physique chimie.
- Série D : pour les élèves scientifiques qui maîtrisent le plus la matière SVT, et qui maîtrise peu les mathématiques et la physique – chimie.

4 Résultats

La Figure 2 illustre un résultat pour $MinSupp = 0.99$, $MaxCVS = 0.6$ et $MinRapCor = 0.2$. Tout d'abord, nous voyons que les neufs premières règles disposent comme prémisse la moyenne générale et comme conséquences les neuf matières de l'examen. Ces neufs règles forment une description générale du résultat de l'examen du baccalauréat session 2014 : en générale, les élèves ont eu une moyenne générale de $\frac{8.67}{20}$. Si l'élève a obtenu une moyenne générale de $\frac{8.67}{20}$, alors il aurait une note de 5 /20 en mathématiques, 9/20 en Histoire géographie, 11/20 en Malagasy, 10/20 en Philosophie, 6/20 en physique – chimie, 7/20 en Français, 5/20 en SVT, 5/20 en Langue vivante (que ce soit Anglais ou Espagnol ou Allemand ou Russe) et 11/20 en EPS. De plus, bien que ces élèves soient littéraires, on remarque également que leurs notes dans les matières littéraires sont faibles, à l'exception de la langue malagasy qui est la langue maternelle. Cela montre que l'élève n'est pas fort dans sa spécialité.

```
> Resultat
                        Premisse Implication                  Consequence Rapport.Correlation
1     Moyenne.G : moyenne = 8.67          =>  Mathematiques : moyenne = 5.4              0.68
2     Moyenne.G : moyenne = 8.67          =>  Histoire.geo : moyenne = 9.04              0.67
3     Moyenne.G : moyenne = 8.67          =>     Malagasy : moyenne = 11.48              0.57
4     Moyenne.G : moyenne = 8.67          =>   Philosophie : moyenne = 9.71              0.57
5     Moyenne.G : moyenne = 8.67          =>     Physique.Chi : moyenne = 6              0.53
6     Moyenne.G : moyenne = 8.67          =>      Francais : moyenne = 7.39              0.50
7     Moyenne.G : moyenne = 8.67          =>          SVT : moyenne = 4.59               0.47
8     Moyenne.G : moyenne = 8.67          => Langue.Vivante : moyenne = 4.98             0.40
9     Moyenne.G : moyenne = 8.67          =>          EPS : moyenne = 11.27              0.32
10 Mathematiques : moyenne = 5.4          =>     Physique.Chi : moyenne = 6              0.31
11 Mathematiques : moyenne = 5.4          =>          SVT : moyenne = 4.59               0.23
```

Figure 2 –. Résultat pour MinSup = 0.99, MaxCVS = 0.6 et MinRapCor = 0.2

Ensuite, on note que la meilleure règle valide est la suivante : (Moyenne générale=8.67)⇒(Mathématiques=5.4). Cette règle nous incite à penser que la réussite des élèves série A2 dépend de leurs notes des mathématiques. Cela est évident, car la série A2 de Madagascar est faite pour les élèves littéraires qui s'intéressent aux mathématiques (coefficient 03). C'est pourquoi la faiblesse de note de mathématiques entraîne un échec pour ces candidats de la série A2.

Les deux dernières règles montrent l'existence d'une liaison fonctionnelle entre les notes des matières scientifiques. Ce résultat peut être interprété comme suit : en général, si l'élève série A2 de la région DIANA de Madagascar a une note égale à 05/20 en mathématiques, alors il aurait une note de 05/20 en Science de la Vie et de la Terre, et une note de 06/20 en Physique Chimie. Ce niveau démontre clairement que les élèves de série A (série littéraire) ne maîtrisent pas les matières scientifiques. On peut aussi admettre que, en tant que littéraire, ces élèves négligent les matières scientifiques.

Ces informations pourraient aider les décideurs en éducation de Madagascar de décider :

- d'améliorer la qualité d'enseignement des matières littéraires ;
- de diminuer les coefficients des matières scientifiques dans la série A, car ces matières n'intéressent pas les élèves de la série A ;
- d'éliminer peut – être la série A2, car leurs notes des mathématiques sont très faibles : 05/20. Toutefois, la série A2 est faite pour les élèves série A qui s'intéressent beaucoup aux mathématiques (coefficient 03 selon le système de l'année 2014).

D'où, la justification du nouveau programme scolaire du Plan Sectoriel de l'Education (PSE) actuel pour les séries littéraires. Les décideurs ont décidé d'éliminer la série A2, en fixant seulement la série L. En série L, les matières littéraires sont toutes dotées de coefficients 04 et 05, tandis que les matières scientifiques sont toutes de coefficient 01. Ainsi, les élèves littéraires peuvent se spécialiser dans leurs domaines.

5 Discussion

Dans cette section, nous allons discuter et comparer théoriquement les résultats de notre méthode avec les résultats obtenus par d'autres théories existantes dans la littérature.

5.1 Comparaison avec la théorie de Aumann et Lindell

Selon la théorie de Aumann et Lindell (1999), une règle $(X \in [a\,;b]) \Rightarrow Y(M_Y = \overline{Y})$ est dite irréductible si pour toute constante $c \in [a\,;b]$, on a : $\overline{Y}_{X\in[a,c]} > \overline{Y}$ et $\overline{Y}_{X\in[c,b]} > \overline{Y}$, où $\overline{Y}$ est la moyenne arithmétique de Y ; $\overline{Y}_{X\in[a,c]}$ et $\overline{Y}_{X\in[c,b]}$ correspondent respectivement à la moyenne arithmétique de Y sur les individus vérifiant $X \in [a, c]$ et $X \in [c, b]$. Ensuite, une règle $(X \in [a\,;b]) \Rightarrow Y(M_Y = \overline{Y})$ est dite maximale si l'on ne peut plus élargir l'intervalle $[a, b]$ de la prémisse pour que cette règle soit irréductible. Une règle maximale est la règle irréductible qui contient l'intervalle la plus longue $[a, b]$. Alors, le fait qu'une règle $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$ soit validée avec un seuil $MaxCVS$ faible (inférieur à 0.3) et un seuil $MinRapCor$ élevé (supérieur à 0.7) ne nous permet pas de conclure que la règle $(X \in [\min(X)\,, \max(X)]) \Rightarrow (Y = \bar{Y})$ est quasi-maximale selon la théorie de Aumann et Lindell (1999).

En effet, si une règle $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$ est valide avec un seuil $MinRapCor$ élevé, alors il existe une relation de dépendance continue, non constante et non nécessairement linéaire f entre les deux variables aléatoires X et Y, telle que $Y = f(X)$. Alors, il suffit simplement de considérer deux variables aléatoires X et Y linéairement et positivement dépendantes : $Y = aX + b$, avec $a > 0$. Dans ce cas, on peut trouver une constante $c \in [\min(X)\,, \max(X)]$ et $c \leq \bar{X}$, telle que $\bar{Y}_{[\min(X),c]} \leq \bar{Y}$. Ainsi, la règle $(X \in [\min(X)\,, \max(X)]) \Rightarrow (Y = \bar{Y})$ n'est pas maximale selon la théorie de Aumann et Lindell (1999).

5.2 Comparaison selon la théorie de Agrawal et al.

Tout d'abord, selon Chafaï et Zitt (2016), le théorème de Bienaymé – Tchebycheff affirme que pour une variable aléatoire X ayant une moyenne E(X) et une variance ($X \in L^2$), on a :

$$\forall\, r > 0, P(|X - E(X)| \geq r) \leq \frac{\sigma^2(X)}{r^2}. \qquad (16)$$

C'est une majoration de la probabilité pour que l'écart entre les valeurs de X et sa valeur moyenne $E(X)$ soit plus grand que r. Ainsi, plus $\sigma^2(X)$ est faible, plus $E(X)$ représente vraiment une valeur typique de X. Dans ce cas, on dit que la moyenne $E(X)$ est significative, car les valeurs de X sont homogènes ou regroupées autour de la valeur moyenne $E(X)$. Par conséquent, si $\sigma^2(X)$ est faible, alors il existe un intervalle $[a\,;b]$ contenant $E(X)$, et un petit nombre réel $\varepsilon \in [0\,;1]$, tel que $P(X \notin [a\,;b]) < \varepsilon$ et $[a\,;b] \subset [Min(X)\,;Max(X)]$. En passant à l'événement contraire, il existe un intervalle $[a\,;b]$ contenant $E(X)$, et un petit nombre réel $\varepsilon \in [0\,;1]$, tel que $P(X \in [a\,;b]) \geq 1 - \varepsilon$ et $[a\,;b] \subset [Min(X)\,;Max(X)]$.

Toutefois, on a : $CVS(X) = a \times \sigma(X)$, avec $a = 0$, si $Max(X) = \bar{X}$ ou $Min(X) = \bar{X}$; $a = \frac{1}{\sqrt{[Max(X)-\bar{X}\,]\,[\,\bar{X}-Min(X)]}}$, si $Max(X) \neq \bar{X}$ et $Min(X) \neq \bar{X}$. Comme $a \geq 0$, alors si $\sigma(X)$ est faible, alors $CVS(X)$ l'est aussi. Par conséquent, si $CVS(X)$ est faible, alors il existe un intervalle $[a\,;b]$ contenant $E(X)$, et un petit nombre réel $\varepsilon \in [0\,;1]$, tel que $P(X \in [a\,;b]) \geq 1 - \varepsilon$ et $[a\,;b] \subset [Min(X)\,;Max(X)]$. Cependant, le fait que $CVS(X)$ soit faible nous permet de faire l'approximation $[a\,;b] \approx [Min(X)\,;Max(X)]$. Ensuite, la règle $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$ peut être considérée comme $(X \in [\min(X)\,, \max(X)]) \Rightarrow (Y \in [\min(Y)\,, \max(Y)])$, car les moyennes $\bar{X}$ et $\bar{Y}$ représentent respectivement les valeurs générales de X et Y dans les intervalles $[\min(X)\,, \max(X)]$ et $[\min(Y)\,, \max(Y)]$. Par conséquent, si CVS(X) et CVS(Y) sont

faibles, alors les deux intervalles [min(X) , max(X)]) et [min(Y) , max(Y)]) peuvent être considérés respectivement comme les « meilleures partitions »pour les attributs quantitatifs X et Y. Ainsi, si CVS(X) et CVS(Y) sont faibles, alors on peut avoir tout de suite les « meilleures partitions » nécessaires pour engendrer une règle quantitative comme dans le travail de Srikant et Agrawal (1996). Le seul problème qui ne nous permet pas d'avoir une même partition que celle de Srikant et Agrawal (1996) est la contrainte de MaxSup qu'ils ont intégré dans leur algorithme afin d'éviter le problème de MinConf. Plus nous utilisons un MaxSupp élevé et un MaxCVS faible, plus on retrouve des partitions similaires. Cependant, les règles obtenues ne seront pas toujours les mêmes, car la liaison entre les mesures confiance et le rapport de corrélation est floue. Notre approche minimise la perte d'informations, car on ne fait point de partitionnement. Toutefois, les résultats sont limités à des règles comportant seulement $1-$ itemset dans la prémisse et dans la conséquence. Ainsi, il reste encore à trouver une autre technique permettant de réappliquer cette méthode afin d'extraire des règles avec $k-$ itemsets.

6 Conclusion

Nous avons introduit une nouvelle technique d'extraction des règles d'association quantitative utilisant la mesure de dispersion « coefficient de variation de Sharma CVS » et la mesure de dépendance fonctionnelle « rapport de corrélation ». En général, le traitement se fait en trois étapes : (1) sélection des items fréquents satisfaisant la contrainte $MinSupp$ fixé par l'utilisateur, en passant par une matrice booléenne intermédiaire ; (2) sélection des items fréquents satisfaisant la contrainte $MaxCVS$ spécifié par l'utilisateur, afin de travailler seulement avec les items possédant des valeurs homogènes ; (3) formation des règles quantitatives et sélection des règles satisfaisant la contrainte minimum de rapport de corrélation donnée par l'utilisateur, notée $MinRapCor$. Le résultat obtenu est un ensemble des règles d'association quantitatives de type $(X = \bar{X}) \Rightarrow (Y = \bar{Y})$, telle que $\bar{X}$ et $\bar{Y}$ sont les moyennes des valeurs des attributs quantitatifs X et Y. Une telle règle est interprétée comme un comportement général du phénomène étudié. Les résultats des expérimentations avec les données réelles de l'éducation de la région nord de Madagascar présentés dans ce papier justifient les reformes en éducation effectuées par les décideurs de Madagascar sur le Plan Sectoriel de l'Education dans la section littéraire. Ainsi, ce papier offre un outil d'aide à la prise de décision scientifique à tous les gérants, économistes et hauts responsables d'un pays. L'extension de notre technique à des règles d'association dont la prémisse et le conséquent sont de $k-itemsets$ est la suite naturelle de notre travail. En perspective, nous pensons que la justification de l'emploi du *rapport de corrélation* pourrait se faire aussi à partir de la considération de l'intensité d'implication de Gras.

7 Remerciements

Nous remercions Monsieur le Chef du service du baccalauréat de l'Université d'Antsiranana pour son accord à l'utilisation des données nécessaires à la réalisation de cet article.

Références

[1] Agrawal R., Imielinski T. et Swami A. (1993), *Mining association rules between sets of items in large databases*, Proceedings of the ACM SIGMOD international conference on Management of data. 207–216.

[2] Agrawal R. et Srikant R. (1994), *Fast algorithms for mining association rules*. Proceedings of the 20th International Conference on Very Large Data Bases, 487 – 499.

[3] Aumann Y. et Lindell Y. (1999). *A statistical theory for quantitative association rules*. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. August, 261–270.

[4] Chafaï D. et Zitt P. A. (2016), Probabilités – Préparation à l'agrégation interne, CreatSpace Independent Publishing, https://doi.org/978-1537566542.

[5] Sharma R., Shandil R.G. et Kapoor G. (2011), A note on Karl Pearson's coefficient of dispersion. Himachal Pradesh University Journal, https://pdfs.semanticscholar.org/229c/6e64c38e0b39773b03b82f458d8779e1f638.pdf or https://studylib.net/doc/8762538/a-note-on-karl-pearson-s-coefficient-of-dispersion

[6] Sørensen B. J. (2002), *The Use and Misuse of the Coefficient of Variation in Organizational Demography Research,* Sociological Methods & Research, 30(4), 475–491.

[7] Srikant R. et Agrawal R. (1996), *Mining quantitative association rules in large relational databases*, Proceedings of the 1996 ACM SIGMOD international conference on Management of data, 1–12.

[8] Sujatha D. et Naveen C. H. (2011), *Quantitative Association Rule Mining on Weighted Transactional Data.* International Journal of Information and Education Technology, Vol. 1, No. 3.

[9] Reh W. and Scheffier B. (1996), Significance Tests and Confidence Intervals for Coefficients of Variation. The statistical software newsletter, 449 – 452.

[10] Totohasina A. (2008), *Contribution à l'étude des mesures de la qualité des règles d'association : normalisation sous cinq contraintes et cas de MGK : propriétés, bases composites des règles et extension en vue d'applications en statistique et en sciences physiques*. Thèse de HDR spécialité Mathématiques et informatique, Université de Madagascar.

[11] Webb G.I. (2001), *Discovering associations with numeric variables*, Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 383–388.

UNE ÉTUDE COMPARATIVE DES PROPRIÉTÉS DES MESURES MGK ET ODD-RATIO

Bruce Masonova Solozafy Bemena[1], André Totohasina[2], Daniel Rajaonasy Feno[3]

A COMPARATIVE STUDY OF THE PROPERTIES OF MGK AND ODD-RATIO MEASUREMENTS

RÉSUMÉ

En épidémiologie, la règle d'association est utilisée pour déterminer les facteurs à l'origine des maladies : l'analyse statistique implicative s'avère ainsi un outil nécessaire en épidémiologie aussi. Les épidémiologistes ont choisi plus souvent ladite mesure Odds-ratio dans leurs études de la quantification de lien implicatif entre une exposition et la maladie. Afin d'obtenir des bons résultats, nous devons être certains que la mesure Odds-ratio est vraiment la plus pertinente mesure existante. Aussi, s'avère-t-il nécessaire d'étudier les propriétés mathématiques de Odd-ratio. Ce papier propose une étude comparative sur les propriétés mathématiques des mesures Odds-ratio, M_{GK} et la mesure normalisée de Odd-Ratio. Nous avons choisi la mesure MGK parce que la littérature la considère comme une bonne mesure d'extraction des règles d'association implicative selon ses propriétés mathématiques.

Mots-clés : *Épidémiologie, règle d'association, normalisation, mesures de qualité implicative, probabilité conditionnelle.*

ABSTRACT

In epidemiology, the association rule is used to determine the causal factors of diseases. Epidemiologists have always chosen the Odds Ratio measure in their studies of the quantification of the link between an exposure and disease. Because they believe it is the most relevant measure. In order to get good results, we need to be sure that the Odds-ratio measure is really the most relevant measure available. To do this, we need to check its mathematical properties. This paper proposes a comparative study on the mathematical properties of the Odds-ratio, MGK and the normalized Odd-Ratio measure. We have chosen the M_{GK} measure because it has been attested in the literature as a good measure for extracting implicative association rules according to its mathematical properties.

Keywords : *Epidemiology, association rule, standardization, implicative quality measures, probability conditional.*

1 Introduction

A titre indicatif, en épidémiologie, pour chercher les facteurs étiologiques (causes) des maladies, l'étude s'intéresse à l'association entre une exposition E (ou facteur de risque) et une maladie M. Pour permettre de quantifier un tel lien, nous nous intéressons à l'étude de la mesure d'association. Parallèlement à cela, nous avons constaté que plusieurs travaux de thèses et de publications comme M. Gail et al. (2010), Antony (2002), Rao et Miller (2008) ont utilisé souvent la mesure Odd-Ratio dans leurs études.

[1] Lycée Mixte de Sambava B.P.90 - Sambava Madagascar brucebemena@gmail.com

[2] Université d'Antsiranana B.P.0 - Antsiranana Madagascar andre.totohasina@gmail.com

[3] Université de Toamasina B.P.591 - Toamasina Madagascar fenodaniel2@yahoo.fr

Plusieurs épidémiologistes affirment aussi dans leurs travaux Held (2010), Bouyer (2003) que la mesure Odd-Ratio est très efficace dans l'étude de cas témoins. Par ailleurs, la littérature atteste que la mesure M_{GK} est considérée comme une bonne mesure d'extraction des règles d'association implicative selon ses propriétés mathématiques étudiées dans (Buchanan et H. Shortliffe, 1984), (Guillaume, 2000), (Totohasina et al. (2004)), (X. Wu et al. (2004)), (Totohasina et al. (2005)), (Rakotomalala et al. (2019)), (Rakotomalala (2019)), (Ralahady et Totohasina (2019)). Or, une étude comparative entre les mesures M_{GK} et Odd-Ratio est déjà faite en 2017 dans nos propres travaux de Master (Masonova, 2017) : en fait, Odd-Ratio a une faille face à une situation intuitive, en l'occurrence, elle ne prend pas une valeur fixe à l'implication logique. Ce qui provoque la difficulté de définir un seuil minimum et d'interpréter les règles d'association ainsi extraites à partir de cette dernière. Dans la foulée de l'étude effectuée par (Armand et al., 2016b), (Armand et al., 2016a), nous proposons d'approfondir l'étude comparative de M_{GK} et la mesure normalisée de Odd-Ratio selon les 22 propriétés proposées par (Grissa, 2013) Dans cet article, nous prendrons comme référence le travail de Grissa (2013). Elle cite, dans ses travaux de thèse les 22 propriétés pour évaluer la performance des mesures de qualité, dont nous allons revoir et évaluer sur ces trois mesures M_{GK}, Odd-Ratio et Odd-Ratio normalisée pour rassurer les épidémiologistes sur le choix de mesure à utiliser.

Dans ce qui suit, notre travail se divise en cinq sections. La section 2 rappelle quelques notations et définitions utiles. La section 3 explique le lien entre la mesure de qualité des règles d'association et l'épidémiologie. La section 4 montre des comportements de ces trois mesures aux situations de référence. La section 5 propose une étude des comportements avec les 22 propriétés de Grissa (2013). La section 6 pose une conclusion et nous parlons quelques perspectives.

2 Notations et Définitions

Nous présentons dans cette section quelques notations et définitions autour de la fouille de données et la mesure des règles d'association. Soient A un ensemble de m articles ou attributs dits aussi variables, ε un ensemble de n transactions ou entités définies sur l'ensemble d'attributs. Un sous ensemble X de A est appelé motif. Selon B. Ganter et R. Wille (1999), en Analyse Formelle de Concepts, un contexte formel est un triplet K = (ε, A, R), où ε et A sont des ensembles finis, et R est une relation binaire de ε vers A.

Définition .1. Considérons un contexte binaire K (ε, A, R), où ε et A sont des ensembles finis. Un tel contexte sera appelé un contexte de la fouille de données.

Soient K = (ε , A, R) un contexte de la fouille de données, X et Y deux motifs de K.

- Pour un motif X, son dual X' est défini par $X' = \{e \in \varepsilon | \forall x \in X, x\}$.
- On définit la probabilité P sur l'espace probabilisable discret fini (e,$P(e)$) par : pour tout événement E *de* e : $P(E) = \frac{|E|}{|e|}$, où |A| désigne la cardinalité de l'ensemble A.
- On dit que :
- X et Y sont incompatibles, si $X' \cap Y' = \emptyset$ soit $P(X' \cap Y') = 0$.
- X favorise Y, si $P(Y'|X') > P(Y')$, $P(Y'|X')$ désigne la probabilité conditionnelle de Y' *sachant* X'.
- X et Y sont statistiquement indépendants, si $P(Y'|X') = P(Y')$.

- X défavorise Y, si $P(Y'|X') < P(Y')$.
- X implique logiquement Y, si $X' \subseteq Y'$, i.e. $P(Y'|X') = 1$

Définition .2. Selon Feno (2007), une règle d'association est une implication de la forme $X \rightarrow Y$ exprimant le fait que les attributs dans X tendent à apparaître avec ceux dans Y.

Définition .3. Selon Feno (2007) et Totohasina (2008), une mesure de qualité ou mesure d'intérêt des règles est une fonction μ de l'ensemble des règles d'association à valeurs dans R, telle que pour toute règle d'association $X \rightarrow Y$, la valeur $\mu(X \rightarrow Y)$ dépend exclusivement des quatre paramètres n, P(X'), P(Y') et P(X'∩Y'), où P désigne la probabilité discrète uniforme sur l'espace probabilisable (e,P(e)) et n désigne la cardinalité de e (| e |=n).

Définition .4. Soit $X \rightarrow Y$ une règle d'association. Selon Totohasina (2003), une mesure de qualité μ est dite normalisée si elle vérifie les cinq conditions ci-dessous :

a) $\mu(X \rightarrow Y) = -1$, si P(Y'|X') = 0 i.e. X et Y sont incompatibles ;

b) $-1 < \mu(X \rightarrow Y) < 0$, si 0≠P(Y'|X') < P(Y') i.e. X et Y sont négativement dépendants (en répulsion partielle) ;

c) $\mu(X \rightarrow Y) = 0$, si P(Y'|X') = P(Y') (i.e. X et Y sont indépendants) ;

d) $0<\mu(X \rightarrow Y)<1$, si 1≠P(Y'|X') > P(Y') i.e. si X favorise Y ou X et Y s'attirent partiellement ;

e) $\mu(X \rightarrow Y) = 1$, si P(Y'|X') = 1 (ou encore si X implique totalement Y).

Définition .5. Soit X et Y deux motifs d'un contexte de la fouille de données. Guillaume (2000), Feno (2007) et Totohasina (2008) définissent la mesure MGK par

$$M_{GK}(X \rightarrow Y) = \begin{cases} \dfrac{P(Y'|X') - P(Y')}{1 - P(Y')} \; si\ X\ favorise\ Y \\ \dfrac{P(Y'|X') - P(Y')}{P(Y')} \; si\ X\ défavorise\ Y \end{cases} \quad (1)$$

Définition .6. Soit X et Y deux motifs d'un contexte de la fouille de données. P. N. Tan et al. (2002) définit la mesure Odd-Ratio (OR) par

$$OR(X \rightarrow Y) = \frac{P(X' \cap Y')P(\overline{X'} \cap \overline{Y'})}{P(\overline{X'} \cap Y')P(X' \cap \overline{Y'})} \quad (2)$$

A la page 117, les travaux d'Armand (2019) montrent que :

$$OR(X \rightarrow Y) = \frac{P\left(\frac{Y'}{X'}\right)\left(1-P(X')-P(Y')+P\left(\frac{Y'}{X'}\right)P(X')\right)}{\left(P(Y')-P\left(\frac{Y'}{X'}\right)P(X')\right)\left(1-P\left(\frac{Y'}{X'}\right)\right)}. \quad (3)$$

Théorème .1. Soit X et Y deux motifs d'un contexte de la fouille de données. Armand (2019) sur la page 117 démontre que la mesure Odd-Ratio homographie normalisée (ORhn) par

$$OR_{hn}(X \to Y)$$

$$= \begin{cases} \dfrac{P\left(\dfrac{Y'}{X'}\right) - P(Y')}{P\left(\dfrac{Y'}{X'}\right)\left(1 - P(X') - P(Y') + P\left(\dfrac{Y'}{X'}\right)P(X')\right)}, & si\ X\ favorise\ Y; \\ \dfrac{P\left(\dfrac{Y'}{X'}\right) - P(Y')}{\left(1 - P\left(\dfrac{Y'}{X'}\right)\right)\left(P(Y') - P\left(\dfrac{Y'}{X'}\right)P(X')\right)}, & si\ X\ défavorise\ Y. \end{cases} \quad (4)$$

Soit, concernant sa composante favorisante, il vient que :

$$OR_{hn}(favorisante)(X \to Y) = \frac{(1 - P(Y'))M_{GK}(X \to Y)}{(1 - P(Y'/X'))(P(Y') - P(Y'/X')P(X'))}.$$

3 Lien entre la mesure de qualité des règles d'association et l'épidémiologie

L'épidémiologie est une des disciplines qui concourent de façon importante aux progrès des connaissances en santé environnementale et en santé publique. Son atout majeur est d'étudier les relations entre environnement et santé à un niveau global. Cette globalisation est faite selon deux dimensions : En premier lieu, les résultats des études épidémiologiques portent sur des groupes de sujets, définis par exemple par leur exposition à telle ou telle condition environnementale. Le second niveau de globalisation se situe au sein de l'individu lui-même. L'épidémiologie ne cherche pas à étudier ni à définir les mécanismes d'action des expositions sur l'organisme humain, mais, elle mesure leur effet (intégratif) par la survenue de pathologie ou, de façon plus générale, des événements de santé (Bouyer (2003)).

Comme, l'épidémiologie est une discipline plutôt récente. Son champ d'intérêt s'accroît au fil des années, et sa méthodologie est encore en pleine évolution. L'épidémiologie est classiquement définie comme l'étude de la distribution des maladies et de leurs déterminants dans les populations humaines (Rothman, 1998). Cependant, son champ s'est rapidement étendu pour couvrir l'étiologie de l'ensemble des problèmes de santé ainsi que leur contrôle (Last, 1983). La définition de l'épidémiologie montre qu'il faut s'intéresser à deux types de mesures. D'une part, celle qui permettent de caractériser la distribution des maladies ; il s'agit des mesures de risque et d'incidence. D'autre part, celles qui permettent de quantifier le lien entre une exposition et la maladie (étude de cas-témoins) ; il s'agit des mesures des règles d'association (Bouyer (2003)). Une étude de cas-témoins examine le degré d'association entre l'exposition à un agent potentiellement nocif et la prévalence d'une maladie. A cet effet, on identifie d'abord un certain nombre de personnes (les cas) atteintes d'une maladie (par ex. le cancer pulmonaire) (Held, 2010).

4 Étude des comportements des trois mesures aux situations de référence

Nous présentons dans cette section les principales propriétés des mesures MGK, Odd-Ratio et Odd-Ratio homographie normalisé. Donc, nous considérons un contexte de la fouille de données K = (ε, *A, R*) et les motifs *X* et *Y* de K.

Les résultats des propositions suivantes résultent de la définition de M_{GK}, *OR*, OR_{hn}.

Proposition 1. (Situations de référence) Selon Feno (2007), pour tout motifs X et Y , on a :

- *X et Y sont incompatibles, si et seulement si* $M_{GK}(X \rightarrow Y) = -1$ *;*
- *X défavorise Y, si et seulement si* $-1 < M_{GK}(X \rightarrow Y) < 0$ *;*
- *X et Y sont indépendants, si et seulement si* $M_{GK}(X \rightarrow Y) = 0$ *;*
- *X favorise Y, si et seulement si* $0 < M_{GK}(X \rightarrow Y) < 1$ *;*
- *X implique logiquement Y, si et seulement si* $M_{GK}(X \rightarrow Y) = 1$.

Les propriétés ci-dessus expriment le fait que M_{GK} prend ses valeurs sur l'intervalle [-1, 1] tout en reflétant les situations de référence.

Proposition 2. (Situations de référence) Selon Feno (2007), pour tout motifs X et Y, on a :

- *X et Y sont incompatibles, si et seulement si* $OR(X \rightarrow Y) = 0$ *;*
- *X défavorise Y, si et seulement si* $0 < OR(X \rightarrow Y) < 1$ *;*
- *X et Y sont indépendants, si et seulement si* $OR(X \rightarrow Y) = 1$ *;*
- *X favorise Y, si et seulement si* $1 < OR(X \rightarrow Y) < +1$ *;*
- *X implique logiquement Y, si et seulement si* $OR(X \rightarrow Y) = +1$.

Les propriétés ci-dessus expriment le fait que Odd-Ratio(*OR*) prend ses valeurs sur l'intervalle [0, +1] ; cependant elle n'est pas normalisée. Aussi, il s'avère plutôt déterminer sa normalisée et dès lors, considérer celle-ci pour donner du sens à une étude comparative. Les propriétés ci-dessous expriment le fait que Odd-Ratio homographie normalisée (OR_{hn}) prend ses valeurs sur l'intervalle [-1, 1] tout en reflétant les situations de référence.

Proposition 3. ***(Situations de référence)*** *Selon Armand (2019), pour tout motifs X et Y, on a :*

- *X et Y sont incompatibles, si et seulement si* $ORhn(X \rightarrow Y) = -1$ *;*
- *X défavorise Y, si et seulement si* $-1 < ORhn(X \rightarrow Y) < 0$ *;*
- *X et Y sont indépendants, si et seulement si* $ORhn(X \rightarrow Y) = 0$ *;*

- *X favorise Y, si et seulement si* $0 < ORhn(X \rightarrow Y) < 1$ *;*
- *X implique logiquement Y, si et seulement si* $ORhn(X \rightarrow Y) = 1$.

5 Étude des comportements avec les 22 propriétés de Grissa

Nous continuons notre étude des comportements des mesures MGK, OR et OR_{hn} en utilisant les 22 propriétés dans les travaux de Grissa (2013). Les 22 propriétés sont : P_1 : Intelligibilité ou compréhensibilité de la mesure, P_2 : Facilité à fixer un seuil d'acceptation de la règle, P_3 : Mesure non symétrique, P_4 : Mesure non symétrique au sens de la négation de la conclusion, P_5 : Mesure évaluant de la même façon $(X \rightarrow Y)$ et $\overline{X} \rightarrow \overline{Y}$ dans le cas de l'implication logique, P_6 : Mesure croissante en fonction du nombre d'exemples, P_7 : Mesure croissante en fonction de la taille de l'ensemble d'apprentissage, P_8 : Mesure décroissante en fonction de la taille du conséquent ou de la taille de la prémisse, P_9 : Mesure admettant une valeur fixe dans le cas de l'indépendance, P_{10} : Mesure admettant une valeur fixe dans le cas de l'implication logique, P_{11} : Mesure admettant une valeur fixe dans le cas de l'équilibre, P_{12} : Mesure admettant des valeurs identifiables en cas d'attraction entre X et Y, P_{13} : Mesure admettant des valeurs identifiables en cas de répulsion entre X et Y, P_{14} : Mesure capable de tolérer les premiers contre-exemples, P_{15} : Mesure invariante en cas de dilatation de certains effectifs, P_{16} : Mesure capable de différencier entre les règles $X \rightarrow Y$ et $\overline{X} \rightarrow Y$ selon une relation d'opposition, P_{17} : Mesure capable de différencier entre les règles $X \rightarrow Y$ et $X \rightarrow \overline{Y}$selon une relation d'opposition , P_{18} : Mesure évaluant de la même façon les règles $X \rightarrow Y$ et $\overline{X} \rightarrow \overline{Y}$, P_{19} : Mesure possédant une taille la prémisse aléatoire, P_{20} : Mesure statistique, P_{21} : Mesure discriminante et P_{22} : Mesure robuste.

5.1 Évaluation des mesures MGK, *OR* et OR_{hn} selon les propriétés

Grissa (2013) a affirmé dans ses travaux que parmi les 22 propriétés proposées ci-dessus, seulement 19 sont étudiées. Comme, les propriétés P_1 : *Intelligibilité ou compréhensibilité de la mesure*, P_2 : *Facilité à fixer un seuil d'acceptation de la règle* et P_{22} : *Mesure robuste* n'ont pas été retenue dans cette étude, on doit seulement évaluer les mesures étudiées par les 19 propriétés proposées : P_3, P_4, P_5,…, P_{21}. En effet, nous pensons que les deux premières sont subjectives et fonction de la connaissance qu'à l'utilisateur en statistique et la troisième propriété est difficile à étudier puisqu'elle nécessite des outils de calcul très poussés afin d'éviter les erreurs de calcul. Pour le comportement des trois mesures MGK, *OR* et OR_{hn}, nous étudions dans le tableau (1) suivant, le nombre 1 signifie que la mesure en question vérifie la propriété concernée *Pi* tel que $3 \leq i \leq 21$. Et le nombre 0 signifie que la mesure en question ne valide pas la propriété concernée *Pi.* Ce travail va aboutir la construction d'une matrice, que nous présentons dans le tableau (1). Cette étude représente le résultat des comportements des trois mesures M_{GK}, *OR* et OR_{hn} selon leurs propriétés. La démonstration qui vérifie notre résultat dans le tableau (1), se trouve dans l'annexe. Posons : Mesures = M et propriétés = P, nous obtenons :

P / M	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Total
M_{GK}	1	1	1	1	1	0	1	1	0	1	1	1	0	0	0	0	0	1	1	12
OR	0	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	1	1	11
OR_{Im}	1	1	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0	1	1	11

Tableau 9 :Matrice décrivant les trois mesures selon les 19 propriétés

En étudiant les propriétés mathématiques de ces trois mesures, nous avons obtenu les théorèmes ci-dessous. Selon Lallich et Teytaud. (2004), nous définissons une mesure qui admet une valeur fixe à l'indépendance comme ci-dessous.

Définition .7. Soit m une mesure de qualité des règles d'association. Une valeur a est la valeur de la mesure m dans le cas de l'indépendance.

- Si m n'admet pas une valeur fixe dans le cas de l'indépendance et P désigne la probabilité discrète uniforme sur l'espace probabilisable (e, P(e)) i.e. si $\forall a \in R, \exists (X \to Y)$, tel que $P_{X'}(Y') = P(Y')$ et $m(X \to Y) \neq$ a, alors la propriété $P_9(m) = 0$;
- Si m admet une valeur fixe dans le cas de l'indépendance i.e. si $\exists a \in R, \forall (X \to Y)$, tel que $P_{X'}(Y') = P(Y')$ et $m(X \to Y) =$ a, alors la propriété $P_9(m) = 1$.

En étudiant la propriété *P*9, nous obtenons le théorème suivant. Soient $\bar{R} = R \cup \infty$, la règle d'association $X \to Y$ et la mesure probabiliste de qualité μ tel que μ_{imp} est la valeur de $\mu\ (X \to Y)$ à l'incompatibilité, μ_{ind} celle de $\mu\ (X \to Y)$ à l'indépendance et μ_{inc} la valeur de $\mu\ (X \to Y)$ à l'implication logique, où $(\mu_{\text{imp}}, \mu_{\text{ind}}, \mu_{\text{imp}}) \in \bar{R}^3$.

Théorème .2. (a) Une Mesure probabiliste de qualité μ admet une valeur fixe à l'indépendance si, et seulement si, $\mu_{ind} \neq +\infty$ est vérifiée à la situation de référence tel que si $(\mu_{imp}, \mu_{ind}\ \mu_{inc}) \in \bar{R}^3$. (b) Toutes les mesures affines normalisables et homographies normalisables ont une valeur constante μ_{ind} à l'indépendance, où $\mu_{ind} \in R$.

Preuve : (a) Suivant la définition (8) et la justification de ce théorème dans la démonstration de la propriété P_9 de l'annexe. D'où le théorème énoncé. (b) Suivant la définition (4) des mesures normalisées et la justification de ce théorème dans la démonstration de la propriété P_9 de l'annexe. D'où le théorème énoncé.

Selon Lallich et Teytaud. (2004) et Lenca et al. (2003), nous définissons une mesure qui admet une valeur fixe dans le cas de l'implication logique comme ci-dessous.

Définition .9. Soit m une mesure de qualité des règles d'association. Une valeur b est la valeur de la mesure m dans le cas de l'implication logique.

- Si m n'admet pas une valeur fixe dans le cas de l'implication logique et P désigne la probabilité discrète uniforme sur l'espace probabilisable (e, P(e)) i.e. si $\forall b \in R, \exists (X \to Y)$, tel que $P_{X'}(Y') = 1$ et $m(X \to Y) \neq b$, alors la propriété $P_{10}(m) = 0$;
- Si m admet une valeur fixe dans le cas de l'implication logique i.e. si si $\exists b \in R,\ \forall (X \to Y)$, tel que $P_{X'}(Y') = 1 \Rightarrow\ m(X \to Y) = b$, alors la propriété $P_{10}(m) = 1$.

En étudiant la propriété P_{10}, nous obtenons le théorème suivant.

Théorème .3. Toutes les mesures affines normalisables et homographies normalisables ont une valeur constante μ_{imp} à l'implication logique, où $\mu_{imp} \in R$.

Preuve : Suivant la définition (4) des mesures normalisées qui vérifient directement la propriété P_{10} et la justification de ce théorème dans la démonstration de la propriété P_{10} de l'annexe. D'où le théorème énoncé ci-après.

Selon J. Blanchard et al. (2005), nous définissons une valeur fixe *c* dans le cas de l'équilibre comme ci-dessous.

Définition .9. Soit m une mesure de qualité des règles d'association. Une valeur fixe c'est un point de référence qui est l'équilibre lorsqu'une règle possède autant d'exemples que de contre-exemples.

- Si m n'admet pas une valeur fixe dans le cas de l'équilibre et P désigne la probabilité discrète uniforme sur l'espace probabilisable (e, P(e)) i.e. si $\forall c \in R, \exists (\mathrm{X} \to \mathrm{Y})$, tel que $P_{X'}(Y') = \frac{P(X')}{2}$ et $m(\mathrm{X} \to \mathrm{Y}) \neq c$, alors la propriété $P_{11}(m) = 0$;
- Si m admet une valeur fixe dans le cas de l'équilibre i.e. si $\exists c \in R, \forall (\mathrm{X} \to \mathrm{Y})$,, tel que $P_{X'}(Y') = \frac{P(X')}{2} => m(\mathrm{X} \to \mathrm{Y}) = c$, alors la propriété $P_{11}(m) = 1$.

Théorème .4. Toutes les mesures affines normalisables et homographies normalisables n'admettent pas d'une valeur constante dans le cas de l'équilibre.

Preuve : Suivant la définition des mesures normalisés (4) et la justification dans la démonstration de la propriété P_{11} de l'annexe. D'où le théorème énoncé ci-après.

Selon Lallich et Teytaud. (2004), nous définissons une mesure discriminante comme ci-dessous.

Définition .10. Une mesure de qualité m est dite discriminante si elle est capable de distinguer les règles intéressantes lorsque la taille de l'ensemble d'apprentissage n croit.

- Si $\exists \eta \in \mathrm{N}^*, \quad \forall n > \eta, \forall X_1 \to Y_1, \forall X_2 \to Y_2$ tel que $\left(P_{X'_1}(Y_1') > P(Y_1') \; et \; P_{X'_2}(Y_2') > P(Y_2')\right)$ => $\mathrm{m}(X_1 \to Y_1,) \approx \mathrm{m}(X_2 \to Y_2)$ où avec et P désigne la probabilité discrète uniforme sur l'espace probabilisable (e, P(e)), alors la propriété $P_{21}(m) = 0$ (non discriminante) ;
- Si $\forall \eta \in \mathrm{N}^*, \quad \exists n > \eta, \exists X_1 \to Y_1, \exists X_2 \to Y_2$ tel que $\left(P_{X'_1}(Y_1') > P(Y_1') \; et \; P_{X'_2}(Y_2') > P(Y_2')\right)$ => $\mathrm{m}(X_1 \to Y_1,) \neq \mathrm{m}(X_2 \to Y_2)$, alors la propriété $P_{21}(m) = 1$ (discriminante).

En étudiant la propriété P_{21}, nous obtenons le théorème suivant.

Soient $\bar{R} = R \cup \infty$, la règle d'association $X \to Y$ et la mesure probabiliste de qualité μ tel que μ_{imp} est la valeur de $\mu\ (X \to Y)$ à l'incompatibilité, μ_{ind} celle de $\mu\ (X \to Y)$ à l'indépendance et μ_{inc} la valeur de $\mu\ (X \to Y)$ à l'implication logique, où $(\mu_{\mathrm{imp}}, \mu_{\mathrm{ind}}, \mu_{\mathrm{imp}}) \in \bar{R}^3$.

Théorème .5. (a) Une Mesure probabiliste de qualité μ est une mesure discriminante si, et seulement si, $\mu_{imp} \neq \mu_{ind} \neq \mu_{inc}$ est vérifiée à la situation de référence, où $(\mu_{\text{imp}}, \mu_{\text{ind}}, \mu_{\text{imp}}) \in \bar{R}^3$. (b) Toutes les mesures affines normalisables et homographies normalisables sont des mesures discriminantes.

Preuve (a) : Suivant la définition (11) d'une mesure discriminante et la justification dans la démonstration de la propriété P_{21} de l'annexe. D'où le théorème.

Preuve (b) : La justification de ce théorème se trouve dans la démonstration de la propriété P_{21} de l'annexe.

5.2 Récapitulation

D'après l'analyse effectuée ci-dessus, nous affirmons que les mesures M_{GK}, *OR* et OR_{hn} sont toutes des bonnes mesures de qualité des règles d'association. En effet, elles ont validé environ la moitié de ces propriétés. D'après le tableau 1 ci-dessus que la mesure M_{GK} en a validé 12 sur 19 propriétés ; puis la mesure Odd-Ratio (OR) en a validé 11 sur 19 propriétés et la mesure homographie normalisée de Odd-Ratio (OR_{hn}) en a validé 11 aussi sur 19 propriétés. En conséquence, par rapport à ce résultat, la meilleure mesure parmi les trois proposées est la mesure M_{GK}.

6 Conclusion et perspective

Ce papier a résumé les différentes propriétés des mesures M_{GK}, *OR* et OR_{hn}. L'étude comparative de ces propriétés nous permet de découvrir la mesure la plus pertinente parmi les trois mesures. Par rapport aux résultats que nous avons obtenus aux études comparatives des comportements des trois mesures M_{GK}, *OR* et OR_{hn} dans ce papier, il n'y a pas de doute que la mesure M_{GK} est la plus pertinente entre les 03 mesures proposées. Ces résultats ont bien répondu aux incertitudes de penser que la mesure Odd-Ratio avait les meilleures propriétés mathématiques par rapport aux autres mesures. En épidémiologie, pour avoir des résultats plus fiables, précises et plus faciles à interpréter sur l'étude de la quantification de lien d'implication statistique entre une exposition et la maladie, nous conseillons aux analystes et aux épidémiologistes de choisir la mesure M_{GK} dans leurs travaux. Comme perspective, nous proposons naturellement l'utilisation de cette mesure M_{GK} dans des études épidémiologiques.

Références

[33] Antony, S. (2002). *Basic Statistics and Epidemiology*. in publication data, British Library Catotoguing.

[34] Armand (2019). *Exploration des propriétés des homographies : application à la normalisation des mesures de qualité des règles d'association*. Ph. D. thesis, Éditions Universitaires européennes.

[35] Armand, Feno D.R., Totohaisina, A. (2016a). Nouvelle vision unificatrice des mesures d'intérêt : une normalisation par homographie. *AAFD SFC Sciences des données, défis mathématiques et algorithmiques*,287–292.

[36] Armand, Feno D.R., Totohaisina, A. (2016b). Vers une théorie de fonction de normalisation des mesures d'intérêt. *AAFD SFC Sciences des données, défis mathématiques et algorithmiques*, 281–286.

[37] Ganter B., Wille, R. (1999). Formal concept analysis. *Mathematical foundations, Springer Verlag, Berlin*, .

[38] Bouyer, J. (2003). *Epidémiologie environnement et santé publique*. Acton vale paris, Tec et Doc.

[39] Buchanan, B. G. et E. H. Shortliffe (1984). Rule-based expert systems. *Library of Congress cataloging in Publication Data*, 32–41.

[40] Feno, D. R. (2007). *Mesure de qualité des règles d'association : normalisation et caractérisation des règles d'association des bases*. Ph. D. thesis, Université de La Réunion spécialité : Mathématiques Informatique.

[41] Grissa, D. (2013). *Etude comportementale des mesures d'intérêt d'extraction de connaissances*. Ph. D. thesis, Université Blaise Pascal-Clermont-Ferrand II ; Université de Tunis-El Manar (Tunisie).

[42] Guillaume, S. (2000). *Traitement des données volumineuses. Mesures et algorithmes d'extraction des règles d'association et règles ordinales*. Ph. D. thesis, Université de Lyon.

[43] Held, U. (2010). *Qu'est-ce que l'Odds ratio et à quoi sert-il ?* biostatistique suisse, Université du Spita.

[44] Last, J. (1983). *A dictionar of epidemiology*. Oxford Medical Publications.

[45] Rothman K. (1998). *Causal inference*. Epidemiology Ressources, Chesnut Hill.

[46] Blanchard J., Guillet F., Briand H., Gras, R. (2005). Ipee : Indice probabiliste d'écart á l'équilibre pour l'évaluation de la qualité des règles. *In Atelier Qualité des Données et des Connaissances*, 26–34.

[47] Lallich, S. et O. Teytaud. (2004). Evaluation et validation de mesures d'intérêt des règles d'association. *RNTI–E–1, spécial*, 193–217.

[48] Lenca, P., Meyer P., P. Vaillantl, et S. Lallich (2003). *Critères d'évaluation des mesures de qualité en ecd*. Revue des Nouvelles Technologies de l'Information, Mesures de Qualité pour la Fouille de Données, 123–134.

[49] Masonova, S. B. B. (2017). *Etudes mathématiques des modèles de quelques maladies infectieuses*. Ph. D. thesis, ENSET-Université d'Antsiranana-Madagascar.

[50] Gail M., Krickeberg K., Samet J.M., Tsiatis A., Wong W. (2010). *Statistics for biology and health*. pp. 85–209.

[51] Tan P. N., Kumar V., Srivastava J. (2002). Selecting the right interestingness measure for association patterns. *In Proc. of the 8th ACM Intl. Conf. on Knowledge Discovery and Data Mining, KDD 02*, 32–41.

[52] Rakotomalala, H. F. (2019). *Classification Hiérarchique Implicative et Cohésitive selon la mesure M_{GK}-Application en didactique de l'informatique*. Didactique des mathématiques et de l'informatique, Université d'Antananarivo.

[53] Rakotomalala, H. F., Diatta J., Totohasina (2019). Une mesure de cohésion basée sur la mesure de qualité des règles d'association M_{GK}. *SFC Sciences des données, Actes des* 24; *mes Rencontres de la Société Francophone de Classification*, 21–24.

[54] Ralahady, B. B., Totohasina A. (2019). Experimental study of the valid rules according to the measure M_{GK}. *International Journal of Computer Science and Technology (IJCST)*.

[55] Rao, C. R. et J. P. Miller (2008). *Epidemiology and Medical Statistics*. Amsterdam the netherlands, Library of Congress Cataloging in Publication Data.

[56] Totohasina, A. (2003). Normalisation de mesures probabilistes de la qualité des règles in : Proc. sfds'03. *XXXVième Journées de statistiques*, 985–988.

[57] Totohasina, A. (2008). *Contribution à l'étude des mesures de qualité des règles d'associations : normalisation sous cinq contraintes et cas de M_{GK} : propriétés, bases composites des règles et extension en vue d'applications en statistique et en sciences physiques.* Ph. D. thesis, Université, d'Antsiranana, Madagascar : Mathématiques Informatique.

Annexe

Pour consulter le contenu du fichier en annexe, contacter l'un auteur de ce chapitre.

COMPARISON OF TWO CLUSTER ANALYSIS OF VARIABLES: STATISTICAL IMPLICATIVE ANALYSIS VS CLUSTOFVAR

Germán Nieto[1], Pablo Gregori[2]

COMPARAISON DE DEUX ANALYSES PAR CLUSTERS DE VARIABLES : ANALYSE IMPLICATIVE STATISTIQUE VS CLUSTOFVAR

ABSTRACT

The main goal of this paper is the analysis of the relations among the questions of an official statistics survey on living conditions conducted in Spain in the year 2019. We have used two clustering methods, ClustOfVar and Statistical Implicative Analysis, in order to get their different points of view. The second goal of the contribution is to show existing tools for the choice of the number of clusters and the comparison of dendrograms obtained by different methods.

Keywords: *Cluster Analysis; ClustOfVar; Living Conditions Survey; Segmentation of variables; Similarity tree; Statistical Implicative Analysis*

RÉSUMÉ

L'objectif principal de cet article est l'analyse des relations entre les questions d'une enquête statistique officielle menée en Espagne en 2019 sur les conditions de vie. Nous avons utilisé deux méthodes de clustering, ClustOfVar et Analyse Statistique Implicative, afin d'obtenir leurs différents points de vue. Le deuxième objectif de la contribution est de montrer des outils existants pour le choix du nombre de clusters et la comparaison des dendrogrammes obtenus par différentes méthodes.

Mots-clés : *Clustering; ClustOfVar; Enquête sur les conditions de vie; Segmentation des variables; Arbre de similarité; Analyse statistique implicative.*

1 Introduction

Cluster analysis is a widely used data mining technique to separate a whole set of individuals into a number of groups, so that individuals within the same group are highly similar to each other, while individuals into different groups are rather distinct. Those groups are called clusters. The comparison among the individuals is done through their values in several variables, by using a distance (or a similarity), among possibly many choices.

It can also be addressed to make groups of variables: just transpose the original dataset, the role of individuals and variables is swapped, and the variables can be grouped using the same technique. In this paper it is undertake this approach, in order to discover proximity relations among variables in an official statistics survey about living conditions conducted in Spain. Then, it will deal with distances (or similarities) among variables.

1 Universitat Jaume I, Av. Vicent Sos Baynat, s/n 12071 Castelló de la Plana, Espanya, gnietomerino@gmail.com; al401111@uji.es

2 Departamento de Matemáticas. Instituto de Matemáticas y Aplicaciones de Castellón. Universitat Jaume I, Av. Vicent Sos Baynat, s/n 12071 Castelló de la Plana, Espanya, gregori@uji.es

A type of clustering algorithm is the ascending hierarchical clustering. It builds a hierarchy starting with an initial level, where each variable is considered as a cluster in its own. Then at each new level, the couple of most similar clusters (within the previous level) makes up a new cluster, while the rest of clusters of the previous level remain the same in the new level. In the final level, the only two clusters, present at the previous level, are merged in a single cluster containing all the variables. When the user is interested in getting the partition of variables in clusters, she has to choose first the number of clusters, and then get the clusters from the level of the hierarchy containing that number of clusters.

While the distance (or similarity) among variables is used in the first level, the distance (or similarity) among clusters of variables, for the subsequent levels, can be defined in several ways, called methods. Widely used methods are Single Linkage Clustering, Complete Linkage Clustering and Weighted Pair Group Centroid Method.

Several specific methods have been developed for the clustering of numerical variables such as the VARCLUS function of the SAS computer application, Likelihood Linkage Analysis, CLV, Diametrical Clustering and Specific R package: ClustVarLV. (Vigneau, Chen, and Qannari, 2015)

However, concerning qualitative variables or mixtures of quantitative and qualitative variables, far fewer methods have been proposed (Chavent et al., 2012). A recent technique called ClustOfVar (Chavent et al., 2012) is being successfully used (compared to more complex interpretation techniques, such as factor analysis) in various fields, such as Sociology, Learning Analytics, Environmental Sciences, etc., to relate variables in specific studies (Brida et al., 2014). The R package ClustOfVar (Chavent et al., 2012) was specifically developed for this purpose. The ClustOfVar method is specifically devoted to variable clustering, regardless the type of variables, numeric or categorical. This approach uses the PCAMIX method (Kiers, 1991) for doing a principal component analysis with numeric and non numeric data.

The Statistical Implicative Analysis (SIA) (Gras et al., 2008) encompasses a cluster analysis, an association rule analysis, and a combination of both that can deal with both categorical, ordinal, and interval multivariate data. Its particular point of view is based on measuring the similarity and the association among variables by contrasting the observed data to the expected data under the assumption of statistical independence. These methods have been implemented and extended since its foundation in the MS Windows proprietary software CHIC for Classification Hiérarchique Implicative et Cohésitive (ARDM, 2021; Couturier, 2008). Recently, most of the algorithms are freely available through the R package rchic (Couturier 2008).

The Living Conditions Survey (LCS) is an annual survey aimed at households, whose predecessor was the European Union Household Panel (EUHP), which was carried out during the 1994-2001 period (INE, 2019).

The LCS has been carried out since 2004. Based on harmonised criteria for all European Union countries, its main objective is to have a reference source on statistics comparing income distribution and social exclusion within Europe. The carrying out of the LCS provides the European Commission with a first class statistical instrument for poverty and inequality studies, a follow up of social cohesion within the relevant territory, the study of population needs and the impact of social and economic policies on households and people, as well as the design of new policies. The survey is carried out

during Spring of each year, usually between March and June. Its main objective is to provide information regarding the income, level and composition of poverty and social exclusion in Spain, and to enable making comparisons with other European Union countries. For this purpose, the incomes, which are referred to the natural year previous to carrying out the survey, are collected. Also, many other questions about living conditions, which are referred to the time of the survey, are collected. A sample of 15,000 dwellings is taken for the LCS. These dwellings are distributed in 2,000 Census sections across the whole national territory. The data collection method is a personal interview of the members of each household included in the sample. In order to complete missing data or to correct errors, the interviewer can complement the interviews with telephone calls.

The basic LCS questionnaires are:

- Household file (basic data on the household members)
- Household questionnaire (information on the dwelling, equipment, economic situation, and the household income)
- Individual questionnaire (data on activity, personal income, health, education)

2 Methodology

Our main interest is to make inquiries on the relation among the questions included in an important official survey. Since those questions involve quantitative and qualitative answers, we are going to use both ClustOfVar and the similarity tree of SIA, because both of these methods can handle mixed data. We shall describe in Sections 2.1 and 2.2 the basics of ClustOfVar and the similarity tree of SIA. The resulting information, shown in Section 3, shall be interpreted in terms of the meaning of the questions of the survey in Section 4.

In addition, it will be showed how to compare dendrograms using an existing tool that is implemented in the package dendextend of the R Statistical Software (Section 2.3).

The starting point is a mixed qualitative / quantitative multivariate dataset made of n observations or individuals in p variables denoted as $X_1, X_2, \dots, X_p$.

Every ascending hierarchical cluster method follows the same steps:

[29] At the initial level of the hierarchy, every variable is considered to be a cluster (or class) on its own.

[30] Then, the “distance” or “similarity” among every couple of clusters is checked.

[31] The couple of “less distant” (or “more similar”) clusters are merged to form a new cluster for the next level of the hierarchy.

[32] The next level of the hierarchy is made of the new cluster plus all the clusters of the preceding level, except the two merged ones.

[33] At the final level, all the variables belong to the same and only cluster.

It is showed, in the following sections, how each method computes the distance and similarity among clusters.

2.1 ClustOfVar

ClustOfVar is based on PCAMIX (Chavent et al. 2012), a principal component analysis method for a mixture of qualitative and quantitative variables. It includes the ordinary PCA and MCA as special cases.

First, the homogeneity $H(C)$ of a cluster of variables $C = \{X_i\}_i$ is computed by applying the PCAMIX to the dataset of variables in C, then taking the first principal component, say Y_1, and finally computing the value:

$$H(C) := \sum_{\substack{X_j \in C \\ X_j \text{numerical}}} r^2\left(X_j, Y_1\right) + \sum_{\substack{X_j \in C \\ X_j \text{categorical}}} \eta^2\left(X_j, Y_1\right),$$

where r stands for the squared Pearson correlation coefficient and η stands for squared correlation ratio (Shaldehi, 2013; Elashoff and Dunbar, 1972, p. 36). We recall that the squared correlation ratio stands for the ratio of the "sum of squares within groups" over the "total sum of squares", considering the numeric variable and the groups given by the categories of the categorical variable. It is a coefficient of nonlinear association.

The distance (dissimilarity indeed) among two clusters of variables C_1 and C_2 is defined in terms of the homogeneity of the clusters, as $d(C_1, C_2) = H(C_1) + H(C_2) - H(C_1 \cup C_2)$.

The agglomerative process successively merges the two clusters with the smallest dissimilarity d for the next level, and let the other clusters invariant.

At the final step, a single cluster gathers all the variables and the process ends.

The R package ClustOfVar (Chavent et al. 2012) includes the function hclustvar() which builds the hierarchy of the variables, and the function plot.hclustvar() which represents the dendrogram of this hierarchy. Finally the function cutreevar() cuts this dendrogram and gives one of the nested partitions according to the selected number of clusters.

2.2 Similarity trees of Statistical Implicative Analysis

In contrast with ClustOfVar, the Statistical Implicative Analysis measures the similarity between variables and classes.

The similarity between two variables X_1 and X_2 measures how surprisingly large is the observed number of coincidences between the two variables compared to a situation of statistical independence. It was defined initially for binary data. In that case, if n_1 denotes the number of observed 1's in variable X_1, n_2 denotes the number of observed 1's in variable X_2 and n_{12} denotes the number of observed coincident 1's (present in both variables X_1 and X_2), then

$$s(X_1, X_2) = Pr(N_{12} \leq n_{12}),$$

where N_{12} represents the random variable of the number of coincident 1's in a couple of independent binary random variables, such as X_1 and X_2 (Lerman, 1970). Depending on the sampling scheme, it follows either a binomial or a Poisson law. This definition has been properly extended to all the types of variables (Gras et al., 2008).

The similarity matrix contains the similarity index for all the couples of variables.

The similarity index between two clusters of variables C_1 and C_2, each one containing r_1 and r_2 variables, respectively, is computed as (see Lerman, 1970):

$$s(C_1, C_2) = \left(\max\{s(X_i, X_j) : X_i \in C_1, X_j \in C_2\}\right)^{r_1 \times r_2}$$

It is important to mention that the similarity index is not discriminant with very large samples, and a modification, called entropic version, was introduced in (Gras et al., 2008). It is used this version because our sample is clearly large.

At present, there are only two computer software programs for this. One of them is CHIC (ARDM, 2021; Couturier, 2008), with copyright license, running under MS Windows. It is the reference software for SIA practitionners. It was created in 1984 (Gras, 1979) and it has been maintained since then, by Saddo Ag Almouloud, Harrisson Ratsimba-Rajohn and Raphael Couturier, incorporating all the novelties in the SIA theory. This software implements, indeed, the three types of analysis included under the name of SIA, namely, the similarity tree (variable clustering), the implicative graph (association rules) and the cohesion tree (rule clustering). As an example, Elia and Gagatsis (2008) uses the two first analysis of CHIC, together with confirmatory factor analysis, in order to compare and get more insights on data regarding students' understanding of the mathematical concept of function.

The CHIC user interface is data analysis oriented, based in windows and menus, where the user chooses the required options. However, it has not a programming language, then it cannot be used for simulation studies or large scale analysis.

The output of any analysis performed by CHIC is given in three windows, that we explain for the case of the similarity tree:

- A first window shows aesthetic tables, including the original data, some frequency tables, correlations, etc.
- A second window offers plain text, describing the computation of the similarity index between all the couples of variables (the similarity matrix). Next, every level of the hierarchy of the similarity tree, with the new formed cluster and the similarity index between the merged clusters.
- Finally, a third graphics windows shows the dendrogram.

2.2.1 SIA similarity analisys conducted in R

Recently, CHIC has been *partially* ported to R in a package called rchic (Couturier, 2021a, 2021b). With respect to the similarity analysis, rchic produces a graphics windows (the plot of the dendrogram), and it stores the similarity matrix in an R variable of class matrix, so that it can be handled in an R script.

In order to be able to use some interesting R functions with the similarity tree of SIA, we need to turn it into an hclust object. This step requires the description of the tree: on the one hand, the sequence of variables (clusters) that are merged at each step of the tree; on the other hand the value of the similarity index between the two merged variables (clusters) at each step.

Another aspect of CHIC that seems to be not yet ported to rchic is the computation of the entropic version of the similarity index. It has been necessary to us in this contribution, since our dataset is large. Therefore, it is used the original CHIC, and copy "by hand" all the values of the similarity index and the structure of the tree in order to build it in the R environment, and use it with fruitful R functions.

It would be interesting encourage R-Chic developers to implement the entropic version of the similarity index, as well as the storage of the tree as an object compatible with the class hclust. In this way, the similarity trees of SIA will be available to researchers who will be able to take advantage of this choice for their investigations, for direct applications with real data as well as for simulation studies.

It is sketch the steps that we have taken to turn a SIA (or any) similarity tree into an hclust object in R:

- Label the original p variables with negative numbers from -1 to $-p$.
- The first cluster (of two variables) is labelled as 1, and each new formed clusters will take number 2, 3, and so on.
- merge: define a matrix with 2 columns and p rows. Each row contain the (two) labels of the (two) merged variables (or clusters) at that level of the tree.
- heights: define a vector of the similarity matrix between the variables (clusters) merged at each level of the tree (row of the merge matrix).
- The object as.hclust(list(merge=merge, heights=heights)) has the class hclust, and it can be an argument for functions dealing with this class in R.
- The object as.dendrogram(as.hclust(list(merge=merge, heights=heights))) has the class dendrogram, and it can be an argument for functions dealing with this class in R.

2.3 Comparison of dendrograms

One of our goals is to derive conclusions about the relation between the questions of an official statistics survey conducted in Spain. For that aim, it is used two different methods. They give two different points of view on the same reality.

Dendrograms are complex objects, and their comparison is of interest in several fields. A theory for the comparison of dendrograms was developed in Sokal and Rohlf (1962) (see Galili, 2015). The cophenetic correlation coefficient is a measure of how faithfully a dendrogram preserves the pairwise distances between the original unmodeled data points.

dendextend is an R package for creating and comparing visually appealing tree diagrams. It provides utility functions for manipulating dendrogram objects (their color, shape, and content) as well as several advanced methods for comparing trees to one another (both statistically and visually). The tanglegram function allows the visual comparison of two dendrograms, from different algorithms or experiments, by facing them one in front of the other and connecting their labels with lines. Distinct branches are marked with a dashed line. For easier and nicer plotting, dendlist concatenates the two dendrograms together, while untangle attempts to rotate trees with un-aligned labels in search for a good layout.

2.4 Determining the number of clusters

Determining the optimal number of clusters in a data set requires the user to specify the number of clusters k to be generated. A simple solution consists of inspecting the dendrogram produced using hierarchical clustering to see if it suggests a particular number of clusters. Unfortunately, this approach is subjective.

A popular method for selecting the number of clusters is based on stability of partitions from a hierarchy of variables. Moreover it is proposed to use a method based on a bootstrap approach to evaluate the stability of the partitions and determine suitable numbers of clusters. It is implemented in the R function stability.

This function evaluates the stability of partitions obtained from a hierarchy of p variables. This hierarchy is performed with hclustvar and the stability of the partitions, from 2 to $p-1$ clusters, is evaluated with a bootstrap approach. The boostrap approach is the following:

- hclustvar is applied to B boostrap samples of the n rows.
- The partitions of 2 to $p-1$ clusters obtained from the B bootstrap hierarchies are compared with the partitions from the initial hierarchy.
- The mean of the corrected Rand indices is plotted according to the number of clusters. Thus, using the general form of an index corrected for chance: $\frac{\text{Index}-\text{Expected Index}}{\text{Maximum Index}-\text{Expected Index}}$, which is bounded above by 1, and takes on the value 0 when the index equals its expected value, the corrected Rand index would have the form (assuming a maximum Rand index of 1): $\frac{\sum_{ij}\binom{n_{ij}}{2}-\frac{\sum_i\binom{n_{i\cdot}}{2}\sum_j\binom{n_{\cdot j}}{2}}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i\binom{n_{i\cdot}}{2}+\sum_j\binom{n_{\cdot j}}{2}\right]-\frac{\sum_i\binom{n_{i\cdot}}{2}\sum_j\binom{n_{\cdot j}}{2}}{\binom{n}{2}}}$ (Hubert and Arabie, 1985).
- Its graphical representation helps in the determination of a suitable numbers of clusters.

Because of the large number of bootstrap samples, the large sample size of our application, and the computational cost of the hclustvar function, the plot of the stability of partitions is computationally expensive.

2.5 Living Conditions Survey

The main goal of the Living Conditions Survey (LCS) is to provide a reference source on comparative statistics on the distribution of income and social exclusion in the European environment.

The LCS started in 2004 and the resulting microdata files (transversal and longitudinal) are generated with an annual periodicity. Microdata files comprise individual data for one given statistic, which have been filtered appropriately to achieve anonymous information so as to ensure confidentiality.

The survey is annual. Each year is denominated a survey cycle. A quarter of sample is renewed in each cycle. The survey uses the census definition given for a family dwelling, considering the census section as the primary census section sample, and the

main family dwelling as the last sample unit. Initially, the selected sample includes all private households resident in the selected main family dwellings. The population under investigation (target population) are persons who are members of private households who live in main family dwellings, as well as the households as units.

The geographical scope of the research is made up by the whole of Spain. The collection method is mainly a personal interview with the members of each of the households under study, who are considered surveyable when the interview is performed. The interviewer visits the houses and requests the information needed to complete the questionnaires. Microdata files contain personalised and conveniently anonymous information for each year of the survey.

Dataset used is formed by File H: Detailed household data file (esudbYYh) of The Living Conditions Survey (LCS) in INEBase (INE, 2021) of year 2019. It is a file that contains a record for each collaborating household. The file contains detailed information about each home, distributed in the following sections:

- Basic household data (HB variables).
- Data on housing (HH variables).
- Data about social exclusion (HS variables).
- Data about income (HY variables).
- Other complementary variables.
- Data on material deficiency (variable HD).

(INE 2021).

3 Results

It is collected data on two types of variables: seven numerical variables and twenty categorical variables.

Based on a sample of the survey data, it is filtered out and prepared the data to be used by the ClustOfVar (Chavent et al., 2017a) and rchic (Couturier, 2021b) R packages.

Our dataset subesudbscaled19 contains $n = 11435$ observations described according to their performances in $p = 27$ variables.

The names of the variables have been trimmed for convinience, trying to keep their meaning, and are listed as:

"DispIncome"	"ImptRent"	"HelpFamily"
"SocialContrib""	"TaxPropty"	"HelpDwelling"
"DelayBills"	"AffordHoliday"	"AffordEat"
"AffordExpenses""	"AffordPhone"	"AffordTV
"AffordComputer"	"AffordCar"	"MkEndMeet"
"MinMonthIncome"	"SupportExpenses"	"ShortageLight"
"PresenceNoise"	"EnvironmentProblem"	"DelinquencyProblem"
"NoRooms"	"ProblemDwelling"	"SuitableTemperature"
"FacilityShower"	"FacilityToilet"	"AllowChgFurniture"

Table 1 : List of variables

The distribution of the numeric variables is described in Figure 1. It is seen that all the variables have a distribution with a tiny fraction of outliers (bars are not appreciated, but the axis range declares it). It has been decided not to take any action on this regard.

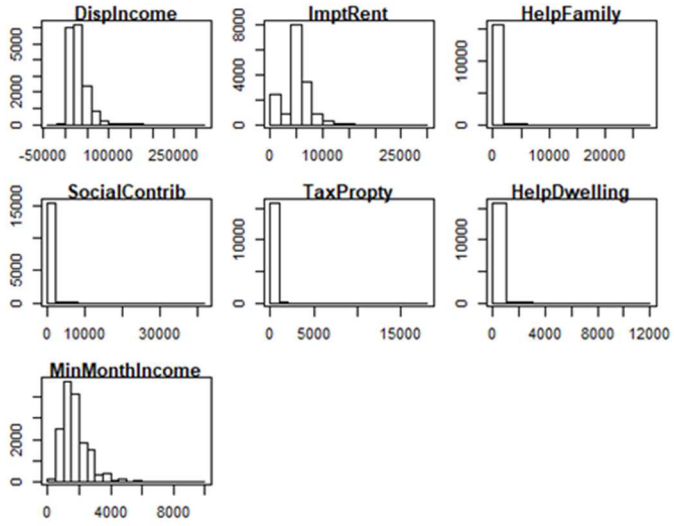


Figure 1 – Distribution of the seven numeric variables. All of them contain tiny fractions of very large values, that we have respected.

3.1 ClustOfVar method

The application of the ClustOfVar methods leads to the dendogram shown in Figure 2.

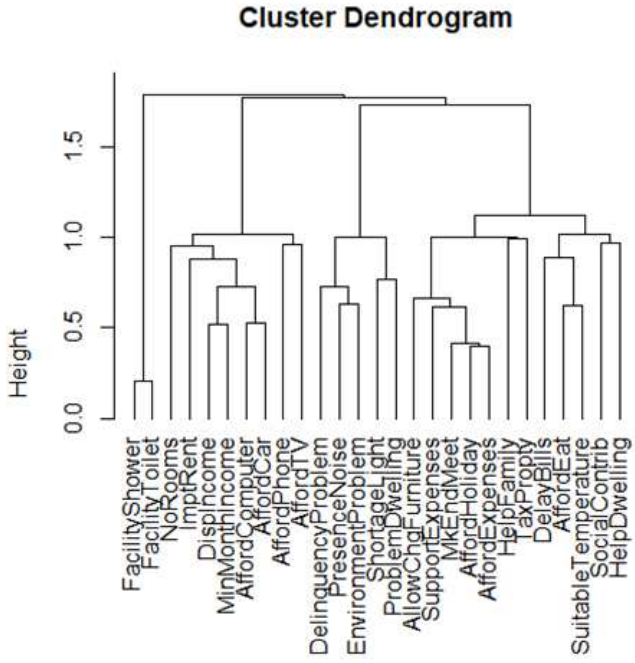


Figure 2 – Dendogram resulting from the analysis of our dataset through the ClusOfVar method.

3.2 SIA similarity tree

The analysis of our dataset with CHIC, using the entropic version and the Poisson law, produces the dendrogram. Following the steps described in Section 2.2.1, it is transfered to the R environment in order to produce a comparable plot, that we show in Figure 3. The use of the classical version of the similarity index produced a rather different tree that showed the effect of chaining. It was clearly inadequate for the large sample that it is being dealed with.

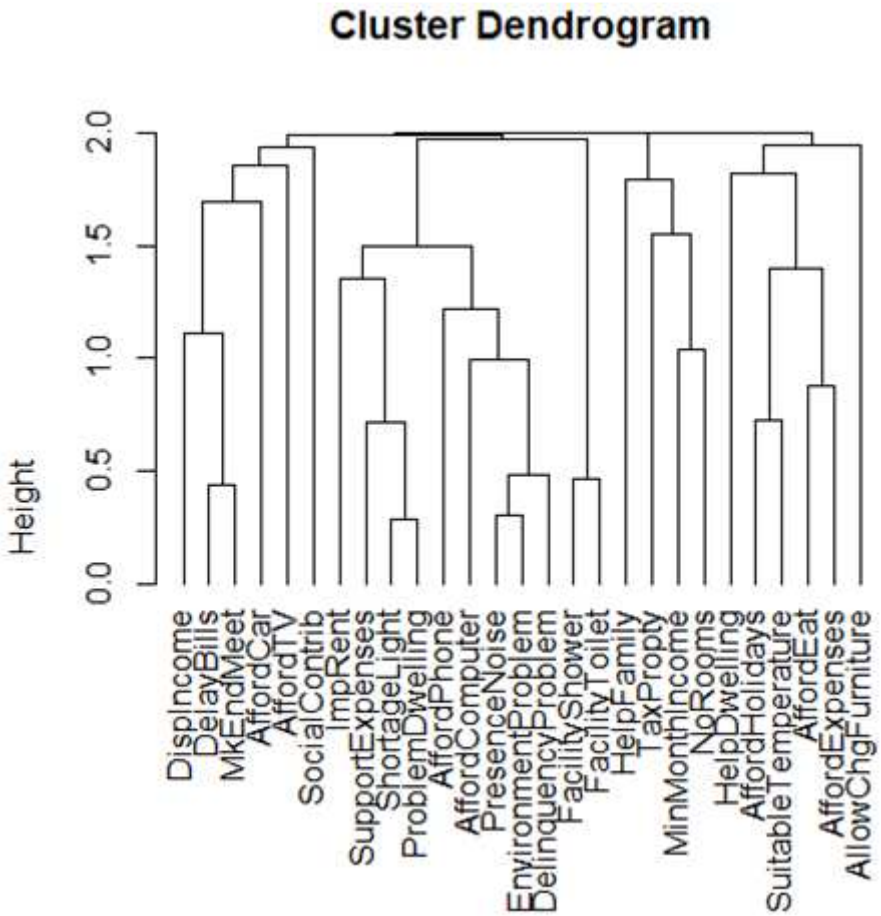


Figure 3 – Dendogram resulting from the analysis of our dataset through the SIA method.

3.3 Number of clusters

A method for estimating the number of clusters is to use a "stability" plot. We use the stability function of the ClustOfVar package. This function evaluates the stability of partitions obtained from a hierarchy of p variables. The stability of the partitions of 2 to $p-1$ clusters is evaluated with a bootstrap approach. The boostrap approch is the following: `hclustvar` is applied to B boostrap samples of the n rows. The partitions of 2 to $p-1$ clusters obtained from the B bootstrap hierarchies are compared with the partitions from the initial hierarchy. The mean of the corrected Rand indices is plotted according to the number of clusters. This graphical representation helps in the determination of a suitable numbers of clusters (see Figure 4).

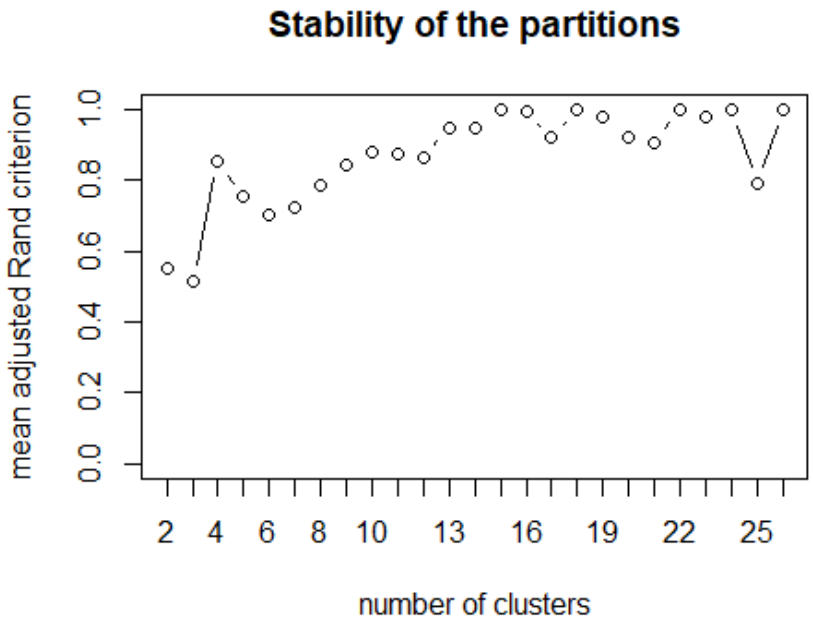


Figure 4 – Stability plot for the choice of the most suitable number of clusters in the analysis of our dataset through the ClusOfVar method.

The maximum for stability is 1.0, so the higher the peak, the better. In this case, the plot of Figure 4 suggests that a partition in K=4 four clusters leads to a stable partition of the data. A partition into a larger number of classes is not more stable, unless K>12, which is clearly not effective in terms of summary information.

3.4 Comparison ClustOfVar vs SIA clusters

It can be seen now how the variables have been classified within the four clusters for each of the two methods. First, a summary of variables per cluster through the ClustOfVar method is shown in Figure 5.

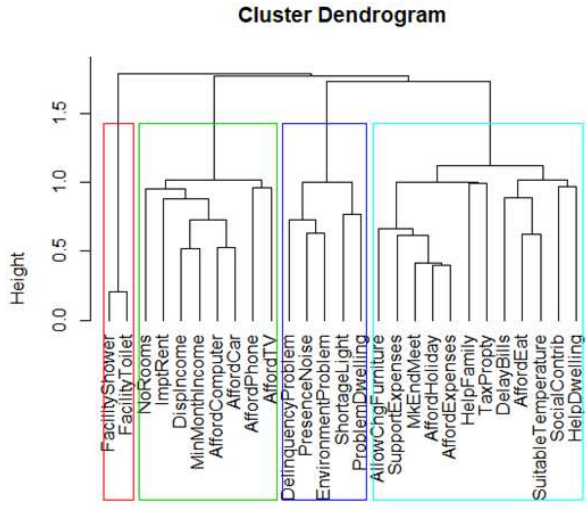


Figure 5 – Dendogram with a highlight of the selected clusters in the analysis of our dataset through the ClusOfVar method.

Next, a summary of our study variables per cluster through the SIA similarity tree is shown in Figure 6.

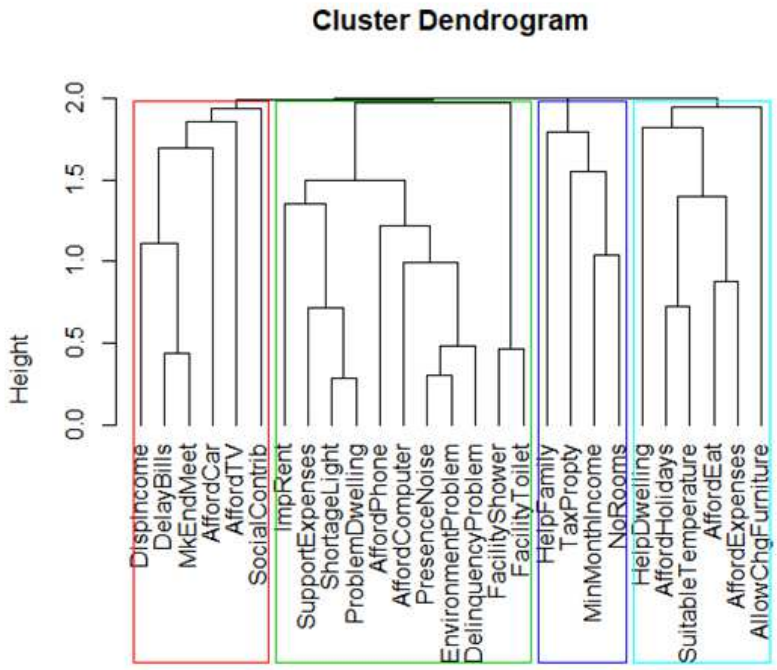


Figure 6 – Dendogram with a highlight of the selected clusters in the analysis of our dataset through the SIA method.

Finally, the tanglegram plot is used, which gives two dendrogram (with the same set of labels), one facing the other, and having their labels connected by lines (see Figure 7). tanglegram can be used for visually comparing any two methods of hierarchical clustering, but they must be encoded as `hclust` trees in the R language.

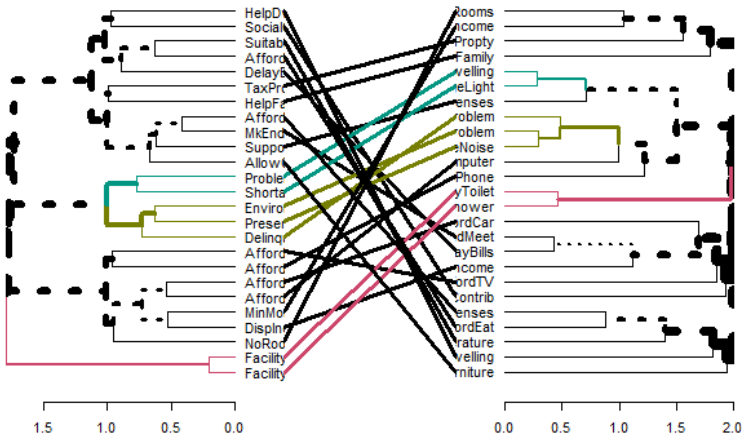


Figure 7 – Plot resulting from the tanglegram function, for the comparison of the two obtained dendograms.

Notice how the "unique" nodes (i.e., nodes which contains a combination of labels/items, which are not present in the other tree) are highlighted with dashed lines. Also notice how the connecting lines are colored to highlight two sub-trees which are

present in both dendrograms. It is colored the branches of the trees to show the two common sub-trees.

The correlation between the tree topologies can be calculated using different measures with cor.dendlist. (Sokal and Rohlf, 1962; Witten and Tibshirani, 2010; Fowlkes and Mallows, 1983; Galili, 2021). The cophenetic distance between two observations that have been clustered is defined to be the inter-group dissimilarity at which the two observations are first combined into a single cluster. This distance has many ties and restrictions. The cophenetic correlation (Sokal and Rohlf, 1962) is the correlation between two cophenetic distance matrices of two trees. In this case the result is `0.2300824`. The value can range between -1 to 1. A value near 0 means that the two trees are not statistically similar. For an exact p-value of such statistics one should invoke a permutation test. It works through permutation over the labels of one tree many times and calculating the distribution under the null hypothesis (keeping the trees topologies constant).

4 Conclusions

It can be observed that the two clustering methods have lead to apparently rather different results in the tree structures. However, it is certain that some similarities are found when grouping the variables. On the one hand, the examination of the clusters obtained with ClustOfVar indicates that Cluster 1 refers to “Housing”, since it aggregates variables “FacilityShower” and “FacilityToilet”. Cluster 2 can be attributed to “Social Exclusion”, because it includes variables such as “MinMonthIncome”, “AffordComputer”, “AffordCar”, “AffordPhone” and “AffordTV”. Cluster 3 collects variables of “SocialExclusion”, such as “DelinquencyProblem”, “PresenceNoise”, “EnvironmentProblem” and “ShortageLight”. And Cluster 4 is the group that collects the highest number of variables, and it is rather heterogeneous. It can be considered that contains variables of the kind “Social Exclusion” (such as “SupportExpenses”, “MkEndMeet”, “AffordExpenses”, “DelayBills”, “AffordEat” and “AffordHoliday”), “Income” (such as “HelpFamily”, “TaxPropty”, “SocialContrib” and “HelpDewlling”), and “Material Deficiency” (such as “AllowCghFurniture”).

On the other hand, the application of SIA has lead to a Cluster 1 with variables “DelayBills”, “MkEndMeet”, “AffordCar”, “AffordTV”, that we consider as “Social Exclusion”. Regarding to Cluster 2, it includes the variables “ShortageLight”, “ProblemDwelling”, “PresenceNoise”, “EnvironmentProblem”, “DelinquencyProblem”, “FacilityShower”, “FacilityToilet”, “AffordComputer”, and “AffordPhone”, hence we relate it to “Social Exclusion” and “Housing”. Cluster 3, with variables “HelpFamily” and “TaxPropty”, seems to be a group related with “Income”, and finally, Cluster 4, composed of variables “HelpDwelling”, “AffordHolidays”, “AffordEat”, “AffordExpenses”, “SuitableTemperature” and “AllowChfFurniture”, are related with “Social Exclusion”, “Housing” and “Material Deficiency”.

For the two clustering methods, it is found that the ClustOfVar is less flexible in the sense that there are no options for the user to choose. The SIA tree has the classical and entropic version, and the binomial and Poisson laws for modeling the number of coincidences. With regard to usability, the computation with ClustOfVar is fast, except for the quest of the number of clusters (stability plot), since the number of bootstrap

samples is large. The computation of the tree with CHIC and `rchic` is also very fast. However, the entropic version is restricted to the CHIC software, which is not compatible with R. We hope that the developers of `rchic` will implement the entropic version soon, as well as the storage of the computations of the tree (that are just printed out in the computation window of CHIC when performing an analysis) as R objects, so that the researchers can handle them efficiently in the R environment. In the meantime, the transfer of the CHIC results to R objects is a hard task, and many of the SIA procedures cannot be involved in intensive simulation studies, such as the stability plot or any Markov Chain MonteCarlo analysis.

References

[1] ARDM, Logiciel CHIC, Available at https://ardm.eu/produit/logiciel-chic. Accessed on 7 June 2021.

[2] Brida, Juan Gabriel, Vincenzo Fasone, Raffaele Scuderi, and Sandra Aguirre. 2014. "ClustOfVar and the Segmentation of Cruise Passengers from Mixed Data: Some Managerial Implications." *Knowledge-Based Systems* 70 (June). https://doi.org/10.1016/j.knosys.2014.06.016.

[3] Chavent, M., V. Kuentz-Simonet, B. Liquet, and J. Saracco. 2012. "ClustOfVar: An R Package for the Clustering of Variables." *Journal of Statistical Software* 50 (13): 1–16. https://doi.org/10.18637/jss.v050.i13.

[4] ———. 2017a. "CRAN - Package Clustofvar." 2017. https://cran.r-project.org/web/packages/ClustOfVar/index.html.

[5] ———. 2017b. "Package 'Clustofvar'." 2017. https://cran.r-project.org/web/packages/ClustOfVar/ClustOfVar.pdf.

[6] Couturier, R. 2008. "CHIC: Cohesive Hierarchical Implicative Classification." *Studies in Computational Intelligence* 127: 41–53. https://doi.org/10.1007/978-3-540-78983-3_2.

[7] Couturier, R. 2021a. "GitHub - Rcouturier/Rchic: CHIC for R." 2021. https://github.com/rcouturier/Rchic.

[8] ———. 2021b. "Rchic - | Raphael Couturier." 2021. https://members.femto-st.fr/raphael-couturier/en/rchic.

[9] Elashoff, J. D., and C. R. Dunbar. 1979. "Measures of Association". https://eric.ed.gov/?id=ED069788.

[10] Elia, I., & Gagatsis, A. 2008. A comparison between the hierarchical clustering

[11] of variables, implicative statistical analysis and confirmatory factor analysis. In. R.

[12] Gras, E. Suzuki, F. Guillet, & F. Spagnolo (Eds.), Studies in Computational Intelligence127: Statistical Implicative Analysis (pp. 131-163). Heidelberg: Springer-Verlag.

[13] https://doi.org/10.1007/978-3-540-78983-3_7

[14] Fowlkes, E. B., and C. L. Mallows. 1983. “A Method for Comparing Two Hierarchical Clusterings.” *Journal of the American Statistical Association* 78 (383): 553–69. https://doi.org/10.1080/01621459.1983.10478008.

[15] Galili, Tal. 2021. “Introduction to Dendextend.” 2021. https://cran.r-project.org/web/packages/dendextend/vignettes/dendextend.html#comparing-two-dendrograms.

[16] ———. 2015. “Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering.” *Bioinformatics (Oxford, England)* 31 (July). https://doi.org/10.1093/bioinformatics/btv428.

[17] Gras, R. 1979. “Contribution à L’étude Expérimentale et à L’analyse de Certaines Acquisitions Cognitives et de Certains Objectifs En Didactique Des Mathématiques.” *Thèse de L’Université de Rennes 1*.

[18] Gras, R., E. Suzuki, F. Guillet, and F. Spagnolo, eds. 2008. *Statistical Implicative Analysis, Theory and Applications*. Vol. 127. Studies in Computational Intelligence. Springer.

[19] Hubert, L., and P. Arabie. 1985. “Comparing Partitions.” *Journal of Classification* 2 (1): 193–218. https://doi.org/10.1007/BF01908075.

[20] INE, Instituto Nacional de Estadística. 2019. “Press Section / Living Conditions Survey (Lcs).” https://www.ine.es/en/prensa/ecv_prensa_en.htm.

[21] ———. 2021. “INE. Instituto Nacional de Estadística.” 2021. https://www.ine.es.

[22] Kiers, H. A. L. 1991. “Simple Structure in Component Analysis Techniques for Mixtures of Qualitative and Quantitative Variables.” *Psychometrika* 56 (2): 197–212. https://doi.org/10.1007/BF02294458.

[23] Lerman, I. 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. http://www.numdam.org/item?id=MSH_1970__32__5_0.

[24] Shaldehi, Ahmad Hedayatpanah. 2013. “Using Eta Correlation Ratio in Analyzing Strongly Nonlinear Relationship Between Two Variables in Practical Researches” 7 (3): 213–20. https://doi.org/10.22436/jmcs.07.03.07.

[25] Sokal, Robert, and F. Rohlf. 1962. “Sokal Rr, Rohlf Fj. The Comparison of Dendrograms by Objective Methods. Taxon 11: 33-40.” *Taxon* 11 (February): 33–40. https://doi.org/10.2307/1217208.

[26] Vigneau, E., M. Chen, and E. M. Qannari. 2015. “ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables.” *R Journal* 7 (2): 134–48. https://doi.org/10.32614/rj-2015-026.

[27] Witten, D., and R. Tibshirani. 2010. “Erratum: A Framework for Feature Selection in Clustering (Journal of the American Statistical Association (2010) 105 713-726)).” *Journal of the American Statistical Association* 105 (492): 1637. https://doi.org/10.1198/jasa.2010.tm10574.

ANALYZING MULTIPLE-REPRESENTATION FLEXIBILITY AND PROBLEM-SOLVING ABILITY THROUGH LONGITUDINAL EMPIRICAL DATA: INSIGHTS FROM IMPLICATIVE ANALYSIS

Eleni Deliyianni,[1] Iliada Elia[2]

ANALYSER LA FLEXIBILITÉ REPRÉSENTATIONNELLE MULTIPLE ET LA CAPACITÉ À RÉSOUDRE LES PROBLÈMES À L'AIDE DE DONNÉES EMPIRIQUES LONGITUDINALES : APERÇUS DE L'ANALYSE IMPLICATIVE

ABSTRACT

*T*he study aims to study developmentally the implicative relations between each multiple-representation flexibility and problem-solving ability component in fraction addition as the students move within and between primary and secondary school. The study followed longitudinal study design. A test that measured students' multiple-representation flexibility and problem-solving ability in fraction addition was administered to the same students three times, with intervals of 3 to 4 months between them. Participants were 388 primary and secondary school students (about 11-14 years old) in Cyprus: 108 students were 5th graders, 132 students were 6th graders, and 148 students were 7th graders at the first measurement. The interaction of the representation field, the function performed by each representation, and the relative sub-concept of fraction addition play a decisive role in the difficulty of multiple-representation flexibility and problem-solving tasks in all the administrations of the test for each age group. However, various differentiations are apparent concerning the task complexity in each age group and administration.

Keywords : *representations, translation, treatment, problem solving*

RÉSUMÉ

L'étude vise à étudier sur le plan du développement les relations implicatives entre chaque composante de flexibilité représentationnelle multiple et de capacité de résolution de problèmes dans l'addition de fractions à mesure que les élèves se déplacent à l'intérieur et entre l'école primaire et secondaire. L'étude a suivi la conception de l'étude longitudinale. Un test mesurant la flexibilité représentationnelle multiple et la capacité de résolution de problèmes des élèves en addition de fractions a été administré aux mêmes élèves trois fois, avec des intervalles de 3 à 4 mois entre eux. Les participants étaient 388 élèves du primaire et du secondaire (environ 11-14 ans) à Chypre : 108 élèves étaient des élèves de 5e année, 132 élèves étaient des élèves de 6e année et 148 élèves étaient des élèves de 7e année lors de la première mesure. L'interaction du champ de représentation, la fonction remplie par chaque représentation et le sous-concept relatif d'addition de fractions jouent un rôle décisif dans la difficulté de la flexibilité représentationnelle multiple et des tâches de résolution de problèmes dans toutes les administrations du test pour chaque groupe d'âge. Cependant, diverses différenciations sont apparentes concernant la complexité des tâches dans chaque groupe d'âge et administration.

Mots-clés : *représentations traduction, traitement, résolution de problèmes*

[1] Cyprus Ministry of Education, Culture, Sports and Youth, deliyianni6@hotmail.com

[2] University of Cyprus, Department of Education, P.O.Box 20537, 1678 Nicosia, elia.iliada@ucy.ac.cy

1 Introduction

There is strong support in the mathematics education community that students can grasp the meaning of mathematical concepts by experiencing multiple mathematical representations (e.g. Sierpinska, 1992; Gagatsis & Shiakalli, 2004; Duval, 2006; Gagatsis, Monoyiou, Deliyianni,, & Philippou,2010; Gagatsis, Christodoulou, & Elia, 2013; Gagatsis, Deliyianni, Elia, Panaoura, & Michael-Chrysanthou, 2016; Nicolaou, S., Gagatsis, Panaoura, Deliyianni, Elia, & Televantou, 2020).

The term "representations" is interpreted as the tools used for representing mathematical ideas such as tables, graphs, and equations (Confrey & Smith, 1991). By a translation process, we mean the psychological processes involving the moving from one mode of representations to another (Janvier, 1987). During the last two decades, the critical problem of translation between and within representations, and the importance of moving among multiple representations and connecting them are addressed in several studies (Sfard, 1992; Yerushalmy, 1997; Goldin, 1998; Reading, 1999). Most of them are based on the assumption, that students´ ability to understand mathematical concepts depends on their ability to make translations among several modes of representations. Others examine the construction and transformation of representations through the process of mathematical problem solving (Yamada, 2000; Gagatsis & Shiakalli, 2004; Theodoulou, R., Gagatsis, A., & Theodoulou, 2004; Elia, Gagatsis, & Demetriou, 2007; Gagatsis, Agathangelou, & Papakosta, 2010).

A large variety of studies concerning a multidimensional approach to the role of representations in understanding and learning mathematics and in problem solving have been produced by the team of researchers in Didactics of mathematics in the University of Cyprus. Some of the studies concern students of the primary education in Cyprus (Markou, & Gagatsis, 2003; Theodoulou, R., Gagatsis, A., & Theodoulou, 2004; Elia, Gagatsis, & Demetriou, 2007; Deliyianni, & Gagatsis, 2009; Gagatsis, Agathangelou, & Papakosta, 2010; Gagatsis, Christodoulou, & Elia, 2013; 2017; Michael – Chrysanthou, Gagatsis, & Papandreou, 2014). Some other studies concern students of the secondary education in Cyprus and in Greece (Elia., Gagatsis, & Gras, 2005; Deliyianni, & Gagatsis, 2009; Deliyianni, & Gagatsis, 2014; Gagatsis, & Deliyianni, 2015). In some of the above studies the researchers compare the use of representations between primary and secondary students (Deliyianni, & Gagatsis, 2014; Gagatsis, & Deliyianni, 2015). Finally, some studies concern university students in Cyprus (Nicolaou, Gagatsis, Deliyianni, Panaoura, Elia, & Anastasiadou, 2017; Nicolaou, Gagatsis, Deliyianni, Panaoura, & Elia, 2020). Moreover, in some studies the relations between the students' use of representations.

Most of studies on representations are characterized by the key word "translation" between representations. In fact, changing modes of representation is the threshold of mathematical comprehension for learners at each stage of curriculum (Duval, 2006; Elia, Gagatsis, & Demetriou, 2007). The ability to pass from one representation to another was also associated with success in problem solving. Different statistical methods have been used in the elaboration of the data of these studies. We strongly believe that the implicative statistical analysis is appropriate for the study of the use of representations in mathematics teaching and learning.

Implicative Statistical Analysis was founded by R. Gras's thesis doctoral thesis in relation to objectives of Didactics of Mathematics (Gras, 1979). This method has been followed by researchers in many scientific fields as a reliable and valid method thanks to Gras's extensive research in collaboration with his students and colleagues (Gras, Couturier, Blanchard, Briand, Kuntz, and Peter, 2004; Gras, Suzuki, Guillet, and Spagnolo, 2008; Gras, Regnier, and Guillet, 2009; Gras, Régnier, Marinica, and Guillet, (2013; Gras, and Couturier; 2013). Its application via the CHIC software has made it possible to produce numerous researches works on a global scale (Couturier, 2008). The above-mentioned method is particularly effective in the domain of the use of representations in the teaching of mathematics and in particular in representational flexibility (Gagatsis, Deliyianni, Elia, & Panaoura, 2010; Gagatsis, Deliyianni, Elia, & Panaoura, 2011; Christodoulou, & Gagatsis, 2014) and in the phenomenon of compartmentalization of representations (Gagatsis, Shiakalli, & Panaoura,2003; Elia., Gagatsis, & Gras, 2005; Gagatsis, Elia, & Mousoulides, 2006; Gagatsis, & Deliyianni,2015; Panaoura, Michael-Chrysanthou, Gagatsis, Elia, & Philippou, 2017).

In fact, the phenomenon of compartmentalization has been revealed in different mathematical concepts by using the Implicative Statistical Analysis. In these studies, the students have the tendency to distinguish the mathematical tasks according the starting representation of the tasks. Some studies concern the use of the representation (the geometrical model) of the number line in addition and subtraction of whole numbers, of fractions and of decimal numbers (Gagatsis, Shiakalli, & Panaoura, 2003; Gagatsis, Kyriakides, & Panaoura, 2004; Shiakalli, & Gagatsis, 2005; Michaelidou, & Gagatsis, 2005; Deliyianni, & Gagatsis, 2009; Panaoura, Gagatsis, Deliyianni, & Elia, 2010; Gagatsis, & Deliyianni, 2015). Some others concern the use of different representations in learning the concept of function (Elia., Gagatsis, & Gras, 2005; Gagatsis, Elia, & Mousoulides, 2006; Elia, Panaoura, Eracleous, & Gagatsis, 2007; Panaoura, Michael-Chrysanthou, Gagatsis, Elia, & Philippou, 2017). In all the above studies the insights from the application of the Implicative Statistical analysis were very decisive in revealing the phenomenon of compartmentalization which has negative consequences in students' learning mathematics.

For example, Gagatsis and Shiakalli (2004) indicate that translation (conversion) ability should be considered as an important factor in problem solving. They also suggest that one cannot claim that those students who are able to successfully solve a translation task will be successful in problem solving. In the same way one cannot claim that students who fail to solve problems will surely fail in solving direct translation tasks as there are many factors involved in problem solving. Translation ability can be reliably classified as one of these factors. The relationship between translation ability and problem-solving ability cannot be considered implicative since success in problem solving is not solely influenced by translation ability, but by many interacting factors (Gagatsis & Shiakalli, 2004).

Moreover, Deliyianni and Gagatsis (2014) concentrated on the cognitive processes of the students dealing with fraction addition conversions and problems in one-year period, during their transition within primary school, within secondary school, or from primary to secondary school within the scope of mathematical working space (Kuzniak, Tanguay & Elia, 2016) using implicative analysis. A compartmentalized way of thinking is evident for the students move from primary to secondary school. Lack of implications or connections among different types of conversion (i.e., with different starting

representation) of the same mathematical content is the main phenomenon of compartmentalization (Duval, 2006) and indicates that students of this age group did not construct the whole meaning of fraction addition.

Deliyianni, Gagatsis, Elia and Panaoura's (2015) study provided evidence for the tenability of a cognitive structural model, which is founded primarily on a synthesis of major theoretical approaches in the field of representations in mathematics learning, and on previous research in the learning of fractions and decimal numbers. The construct of representational flexibility in fraction number addition encompassed seven major components, which all involve inter- and intra-representation transformations: the abilities to recognize whether diagrammatic representations correspond to symbolic expressions of fraction addition with the same denominators or different denominators; the abilities to manipulate fraction additions to calculate the sum within the symbolic system of representations; and the abilities to convert fraction additions from symbolic expression to diagrammatic representation or from diagrammatic representation to symbolic expression, respectively. This structure reveals also that the representational transformation competences of recognition and conversion were affected, respectively, by the complexity of the concepts involved, i.e., the variation of the units of the quantities added, and the direction of the conversion, i.e., from or towards symbolic expression, respectively. Results also showed that two first-order factors were needed to explain the problem-solving ability in fraction addition, indicating the differential effect of the modes of representation that is diagrammatic and verbal form on problem-solving ability. Multiple-representation flexibility and the problem-solving ability were found to be major components of students' representational thinking of the mathematical idea of fraction addition. Even though these two cognitive constructs do not cover the whole spectrum of the understanding of fraction addition, Deliyianni et al (2015) claim that they reflect a wide range of fraction number addition tasks and should be both taken into consideration in teaching. Despite the distinctive characteristics of representational flexibility and problem-solving competences, there is a strong interrelation between them, which could be attributed to the representational transformations of the concept that students needed to apply in the solution processes of both types of tasks, either explicitly or implicitly (Gagatsis, Deliyianni, Elia, & Panaoura, 2017). Furthermore, evidence is provided for the strong interrelations between representational flexibility and problem solving developmentally (Deliyianni & Gagatsis, 2013). The results indicated the students' established pre-existent knowledge and the important role the initial state of these cognitive parameters plays on their advancement.

Moving a step forward, the present study purports to throw some light about the nature and the contribution of multiple representations to mathematical learning and problem-solving. In particular, the present study aims to study developmentally the implicative relations between each representational flexibility and problem-solving ability component in fraction addition (Deliyianni et al. 2015) as the students move within and between primary and secondary school.

2 Methodology

The study followed Deliyianni and Gagatsis's (2013) longitudinal study design. A test (Deliyianni et al., 2015) that measured students' multiple-representation flexibility and problem-solving ability in fraction addition was administered to the same students three

times, with intervals of 3 to 4 months between them. Participants were 388 primary and secondary school students (about 11-14 years old) in Cyprus: 108 students were 5th graders, 132 students were 6th graders, and 148 students were 7th graders at the first measurement. The first measurement took place at the end of the school year. Consequently, 108 students were 6th graders, 132 students were 7th graders, and 148 students were 8th graders at the second and third measurements. There was approximately the same number of boys and girls in each age group. Students were from mixed socio-economic backgrounds. All students had already received teaching in the fraction concept, fraction equivalence, fraction order, and same and different denominator fraction addition before the first measurement.

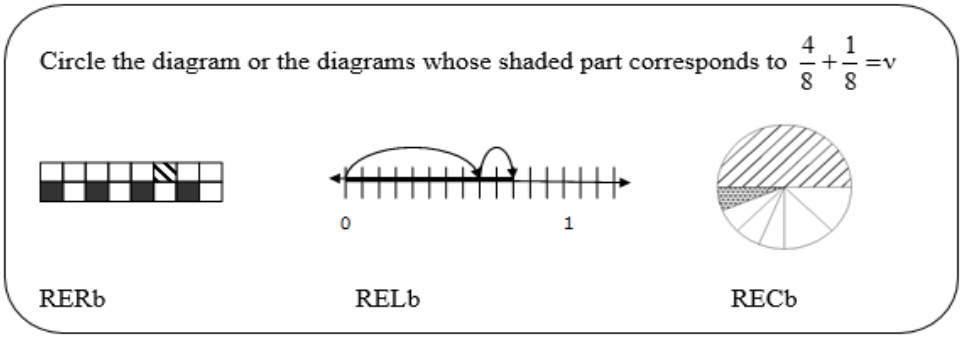
Multiple-representation flexibility tasks differ by considering three sub-dimensions: the representational alteration performed (recognition, manipulation, conversion), the involved representation field (symbolic and diagrammatic representation), and the sub-concept of fraction addition (same and different denominator). On the other hand, problem-solving tasks differ in the following two dimensions: the field of representation (verbal and diagrammatic representation) and the relevant mathematical structure.

The following tasks examine the flexibility of using multiple representations in fraction addition:

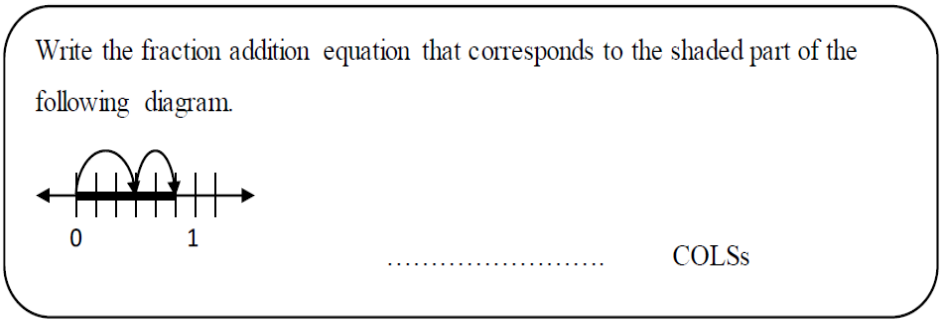
1. There are two same-denominator fraction addition tasks and two different denominator fraction addition tasks. These tasks are represented and solved symbolically, with the sum as the unknown quantity. In one same denominator fraction addition, the resulting sum is less than one (TRSa), while in the other addition of the same denominator fractions, the result is equal to one (TRSd). Regarding the two different denominator fraction additions (TRSb, TRSc), one addition allows the students to apply the equivalence of fractions and handle it as the same denominator fraction addition (TRSb). The sum is less than one in both.

2. Three tasks aim to recognize the concept of adding the same and different denominator fractions in three diagrammatic representations: the circular and rectangular surface area and the geometric model of the number line. Each of these tasks consists of three sub-parts, which are scored separately. The first recognition task concerns the addition of the same denominator fractions (RELa, RECa, RERa). Students are requested to indicate in which diagram or diagrams the shaded part corresponds to a symbolic representation of the addition of the same denominator fractions. In this task, the correct answer is the one presented in the number line. The shading in the other two diagrams acts as a distractor to indicate students' misconceptions about the part-whole relationship and the relationship between the parts of the whole. Students' perception about the part-whole relationship is studied using divisions twice the denominator, while the perception about the relationship of the whole parts is studied using unequal area surfaces. The second recognition task concerns the addition of the same denominator fractions considering the concept of equivalence (RERb, RELb, RECb). Thus, the number of subdivisions in the diagrams are multiple or divisors of the denominator. In this task, the correct answer is the one presented in the number line. The shading in the other two diagrams acts as a distractor, indicating students' misunderstanding of the part-whole relationship due to the numerator and denominator treated as two different numbers. Finally, the third recognition task concerns the addition of the different denominator fractions considering the fraction equivalence (RECc, RERb, RELc). In this task, the

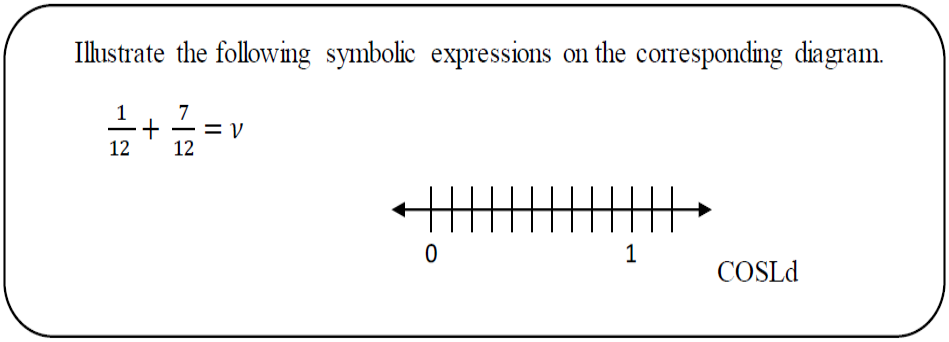
correct answer is presented on a circular and rectangular surface while the shading on the number line acts as a distractor indicating a misconception in the part-whole relationship.



3. There are four conversion tasks from diagrammatic to symbolic representation, two of which concern the addition of the same denominator fractions while the other two concern the addition of the different denominator fractions. The number line (COLSs) and the circular surface (COCSs) are the initial representations in the addition of the same denominator fractions, and the number line (COLSd) and the rectangular area surface in the addition of the different denominator fractions (CORSd).



4. There are four conversion tasks from symbolic to diagrammatic representation, two of which concern the addition of the same denominator fractions and the other two the addition of the different denominator fractions. In the addition of the same denominator fractions, the circular surface (COSCs) and the number line (COSLs) are the final representations, with subdivisions equal to the denominator. In the addition of the different denominator fractions, the number line (COSLd), and the rectangular area surface (COSLd) with the number of subdivisions, the minimum common multiple of the denominators of the two addends, are the final representations.



The unknown quantity in all conversion tasks is the sum, which is always less than or equal to one. In addition, all eight conversion tasks include two addends. It is worth mentioning that the test does not consist of diagrammatic treatment tasks because the students' ability to manipulate diagrammatic addition tasks examined through the conversion tasks in which the diagram is the final representation.

The test also includes four problems which are presented with different fields of representation as follows:

1. In the first problem-solving task, the data are presented diagrammatically on a rectangular area. The known quantity is the sum, and the unknown one is the three addends (PD).

2. The second problem is a verbal problem-solving task, and an auxiliary diagrammatic representation (PVD) is included to facilitate interpretation, according to the Ainsworth (1999) classification of the functions of multiple representations in problem-solving. The diagrammatic representation is the rectangular surface area. Verbal and diagrammatic representations present the problem's data in such a way that the innate qualities of each facilitate the interpretation of the other, regardless of issues of intimacy or experience. The known quantity is in the position of the sum and what is required is to find the four addends.

3. In the third problem-solving task, the data are presented verbally (Evarmib, 2007). The concept of addition is indirectly involved in the problem, while the unknown quantity requires knowledge of the interpretation of the fraction as a ratio (PV).

4. The fourth problem is verbal. The students are requested to justify (JV) whether the sum of two addends with a numerator a number smaller than the denominator can be greater than one.

In the data, the multivariate analysis of repeated measures was performed with the statistical package SPSS. The purpose of this analysis was to compare performance and identify differences in student performance in multiple-representation tasks and problem-solving tasks, which arise during the transition between and within an education level. In addition, statistical implicative analysis (Gras, Suzuki, Guillet & Spagnolo (2008) was applied to the data using the CHIC software program. By applying this analysis, the implicative relations between the variables that are students' responses to the tasks become apparent through the implicative diagrams that are created.

3 Results

3.1 The effect of age and transition time on the conceptual understanding of fraction addition

The multivariate analysis of repeated measurements compared student mean performance that at the first measurement attended Grade 5, Grade 6, and Grade 7 in three different periods for twelve months. Table 1 presents the mean performance of students in the three age groups in terms of the dimensions of the flexibility of using multiple representations and their ability to solve various representation problems in the three administrations of the test.

The results of the research suggest that the transition time [Pillai's F (14,372)= 7.91, p <0.001] had a significant effect on student performance. The performance of students in multiple-representation flexibility and problem-solving dimensions was significantly improved within the period of twelve months. This finding is justified considering the cognitive development and teaching that takes place over one year period in the whole sample. It is worth noting that the overall performance of students in the three age groups in the specific dimensions of fraction-addition understanding did not differ significantly [Pillai's F (14, 760) = 1.28, p = 0.213]. However, there is a significant interaction of time with the age of the students [Pillai's F (28, 746) = 5.25, p <0.001] in terms of the flexibility of using multiple representations and the ability to solve problems of adding fractions. Specifically, it was revealed that student performance in recognizing different denominator fraction addition tasks [F (4, 379) = 7.51, p <0.001], computing symbolically the same and different denominator fractions [F (4, 379) = 16.68, p <0.001], translating from diagrammatic to the symbolic representation of the addition of same and different denominator fractions [F (4, 379) = 3.41, p <0.01], solving verbal problems of adding fractions [F (4, 379) = 5.33, p <0.001] and problems with main or auxiliary diagrammatic fraction-addition representation [F (4, 379) = 10.71, p <0.001] differs depending on whether students move within primary education, from primary to secondary education and within secondary education.

Figures 1 to 7 present the change in the mean performance in each dimension of multiple-representation flexibility and problem-solving with various representations in the same-and-different- denominator-fraction-addition separately during the three administrations. The first age group refers to the students who move from Grade 5 to Grade 6, the second age group consists of the students who change education level (Grade 6 to Grade 7), and the third age group refers to the students who move from Grade 7 to Grade 8.

Figure 1 shows the mean performance of students in the recognition tasks of the same-denominator-fraction addition in a variety of diagrammatic representations of the three age groups during a year. The students of the first age group, who move within the primary educational level, steadily and gradually improve their performance in the specific period in tasks that require recognizing the same-denominator-fraction-addition in various diagrammatic representations.

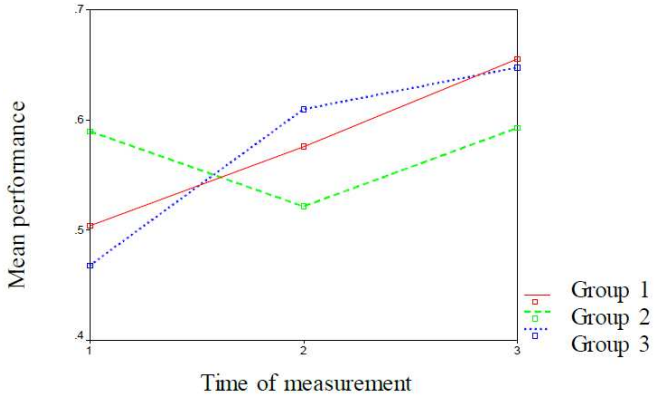


Figure 1 - Mean performance of students in the same-denominator-fraction-addition-recognition tasks per age group during the transition period

The students of the second age group in the first administration had the highest performance in the specific tasks. However, during the transition to secondary education, their performance decreases. This age group has the lowest performance in the second administration compared to the other two age groups. However, in the last administration, the performance of this group improves and reaches the same point as in the previous school year. Therefore, in the period between the first and the third administration, stagnation is apparent in the second age group's ability to recognize the same denominator fraction addition in various diagrammatic representations. The third age group initially has the lowest performance in these recognition tasks. The performance of this group improved significantly in the second administration achieving the highest performance to the other age groups in the specific recognition projects. However, this group's ability to recognize the same denominator fraction addition improves only slightly between the second and third administrations.

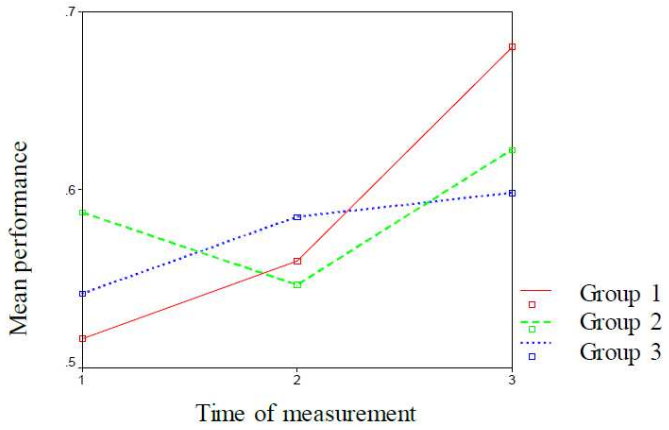


Figure 2 - Mean performance of students in the different-denominator-fraction-addition-recognition tasks per age group during the transition period

Figure 2 shows the mean performance of students in recognizing the addition of different denominator fractions of the three age groups at the three administrations. The mean performance in these tasks increases for the first age group over a period of one year. Of great importance is the improvement to recognize different denominator fraction addition in Grade 6, i.e., between the second and third administration. On the contrary, there is a decrease in the mean performance in the specific recognition tasks for the second age group during their transition from primary to secondary school. Their performance improves between the second and third administrations, i.e., within the school year of attendance in secondary school. Regarding the third age group, improvement of the performance from the first to the second administration in the recognition tasks of different denominator fraction addition is particularly evident. Nevertheless, there is stagnation in the mean performance of the students in secondary school during the school year, i.e., between the second and third administrations.

Figure 3 presents the mean performance of students in the symbolic handling of the addition of same and different denominator fractions of the three age groups during the three administrations. The first age group shows a slight decrease in performance in handling when moving to the next grade. During the school year, however, their ability to symbolically add fractions improves. On the contrary, despite the emphasis placed on symbolic representations in secondary school textbooks, second-age group performance during the transition to secondary school is reduced. Despite the slight improvement in

performance in symbolic-fraction-addition tasks for these students during the school year, their mean performance in the third administration is lower than the corresponding mean performance of the same age group during the previous school year. On the other hand, the mean performance of the third age group improves significantly during the transition to Grade 8. During the school year, however, their mean performance in symbolic-fraction-addition tasks is slightly decreased.

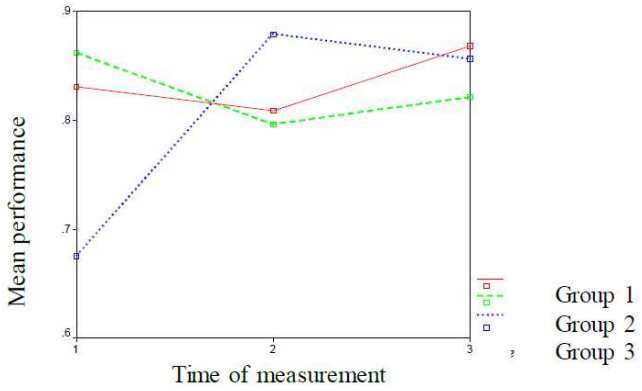


Figure 3 - Mean performance of students in the symbolic-fraction-addition-treatment tasks per age group during the transition period

In Figure 4 the mean performance of students in conversion-same-and-different-denominator-fraction-addition tasks from symbolic to diagrammatic representations of the three age groups at the three administrations is presented. According to the results of Figure 4, the linear improvement of the mean student performance in translation tasks is evident from a symbolic to a diagrammatic representation of the addition of fractions for students in the first and third age group, who study into primary and secondary education, respectively. On the contrary, students who move from Grade 6 to Grade 7 mean performance in these conversion tasks decreases in the second administration and gradually increases in the third one.

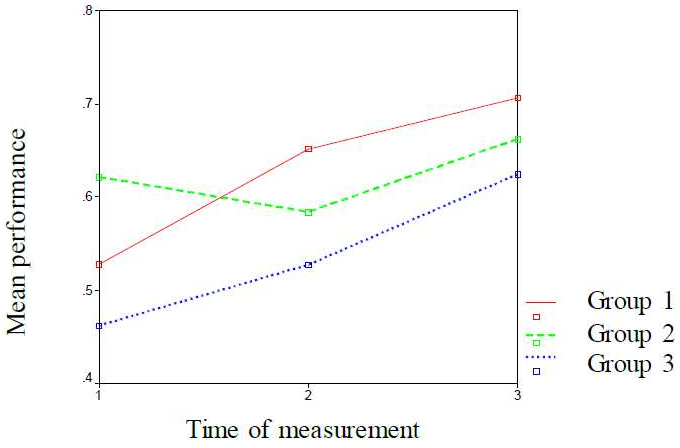


Figure 4 - Mean performance of students in the translation-same-and-different-denominator-fraction-addition tasks from diagrammatic to symbolic representations per age group during the transition period

Figure 5 presents the mean performance of students in translation-same-and-different-denominator-fraction-addition tasks from diagrammatic to symbolic representations of the three age groups at the three administrations. The mean performance of the students

of the first age group improves significantly in both the second and the third administration concerning the mean performance in the specific tasks in the first administration. It is worth noting that the mean performance in translation tasks from diagrammatic to a symbolic representation of the first age group in the second and third administration is higher than the corresponding mean performance for the other two age groups. On the other hand, the mean performance of the second and third age groups during the three administrations is approximately at the same levels.

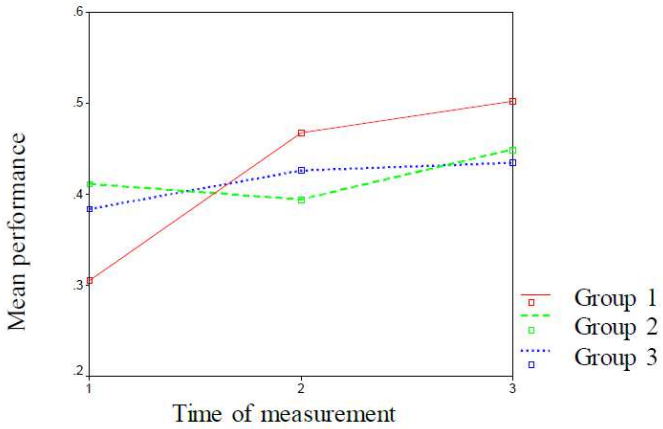


Figure 5 - Mean performance of students in the translation-same-and-different-denominator-fraction-addition tasks from diagrammatic to symbolic representations per age group during the transition period

Figure 6 shows the mean performance of students in the verbal fraction-addition-problem-solving tasks of the three age groups during a year. The first age group's performance in these tasks improves linearly during the transition within the primary education to both the second and the third administration. The performance of the third age group during the transition within secondary education in the specific tasks also shows a linear improvement. On the contrary, the mean performance of the second-age students decreases initially with their transition from primary to secondary education. In the third administration, their performance on verbal problems increases significantly. However, it is not higher compared to the first administration for the specific students.

According to Figure 7, which shows the mean performance of students in solving diagrammatic problems of adding fractions of the three age groups in the three administrations, the mean performance of students in the first age group improves over one year in the specific tasks. Students' mean performance in the second age group decreases in the second and the third administration of the test compared with their performance in the first administration. The student's performance in the specific tasks improves in the third administration compared with the second administration. During the transition of the third group of students within secondary education, the students' mean performance in the problems with main or auxiliary diagrammatic representation initially improves, and then stagnation is observed.

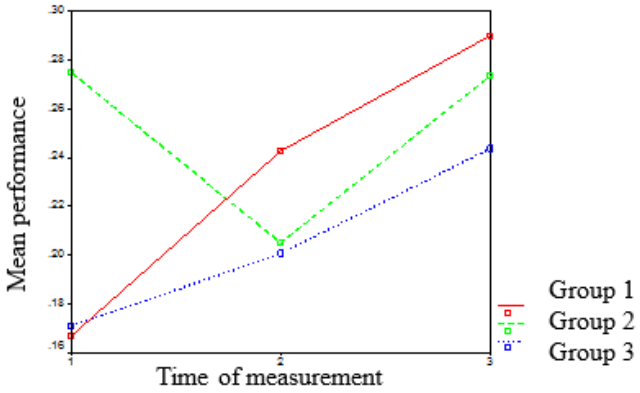


Figure 6 - Mean performance of students in the verbal fraction-addition-problem-solving tasks per age group during the transition period

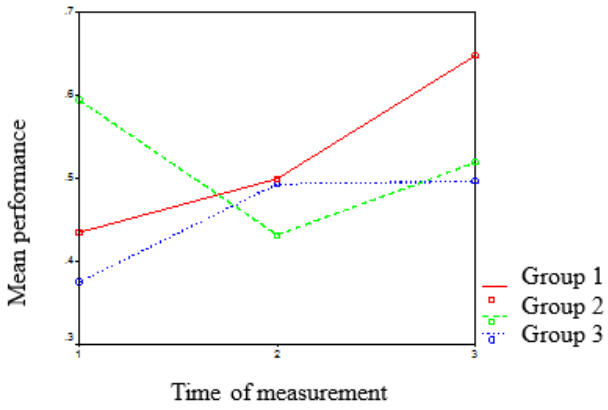


Figure 7- Mean performance of students in the verbal fraction-addition-problem-solving tasks per age group during the transition period

Based on the results in Figure 1-7, in all cognitive aspects of the performance of the second age group is high in the first phase of the study. However, with the transition to secondary education, their performance decreases in the second administration. In the third administration, their performance reaches approximately the same levels as it was a year ago during the first administration. Therefore, stagnation is evident in students' performance in the second age group. In contrast, the first and third age group of students show improved performance in almost all dimensions of the flexibility of using multiple representations and problem solving with different representation fields. In other words, during the transition within the same level, primary or secondary, the students' performance improves significantly. Concerning students' performance in the transition to secondary education, an improvement is indicated mainly from the first to the second administration of the test. The third age group exhibits stagnation in their performance between the second and the third administration of the test in recognizing the addition of fractions with different denominators, symbolic treatment of the same and different denominator fraction addition, translation from diagrammatic to a symbolic representation of the fraction addition and solving diagrammatic problems of fraction addition. Regarding the transition into primary education, the improvement in performance in the cognitive dimensions of fraction addition is almost linear. Despite the improvement in performance due to the time effect, the mean performance remains low

in all age groups in the various dimensions of understanding fraction addition. An exception is the fraction-addition-symbolic-treatment tasks.

3.2 Implicative relations of multiple-representation flexibility and problem-solving tasks with a variety of fraction addition representations

Figures 8 to 10 include the implicative diagrams of students' responses of each age group to the flexibility-of-using-multiple-representations tasks and solving verbal and diagrammatic problems of adding fractions to the three administrations of the test. This analysis allows using the implicative relationships of all the tasks to detect the degree of difficulty and possible changes during the transition in and between primary and secondary education.

Figure 8 shows the implicative relationships between the multi-representation flexibility and problem-solving tasks with different representations in fraction addition in the first, second, and third administration of the test, respectively, for students in the first age group who move from Grade 5 to Grade 6.

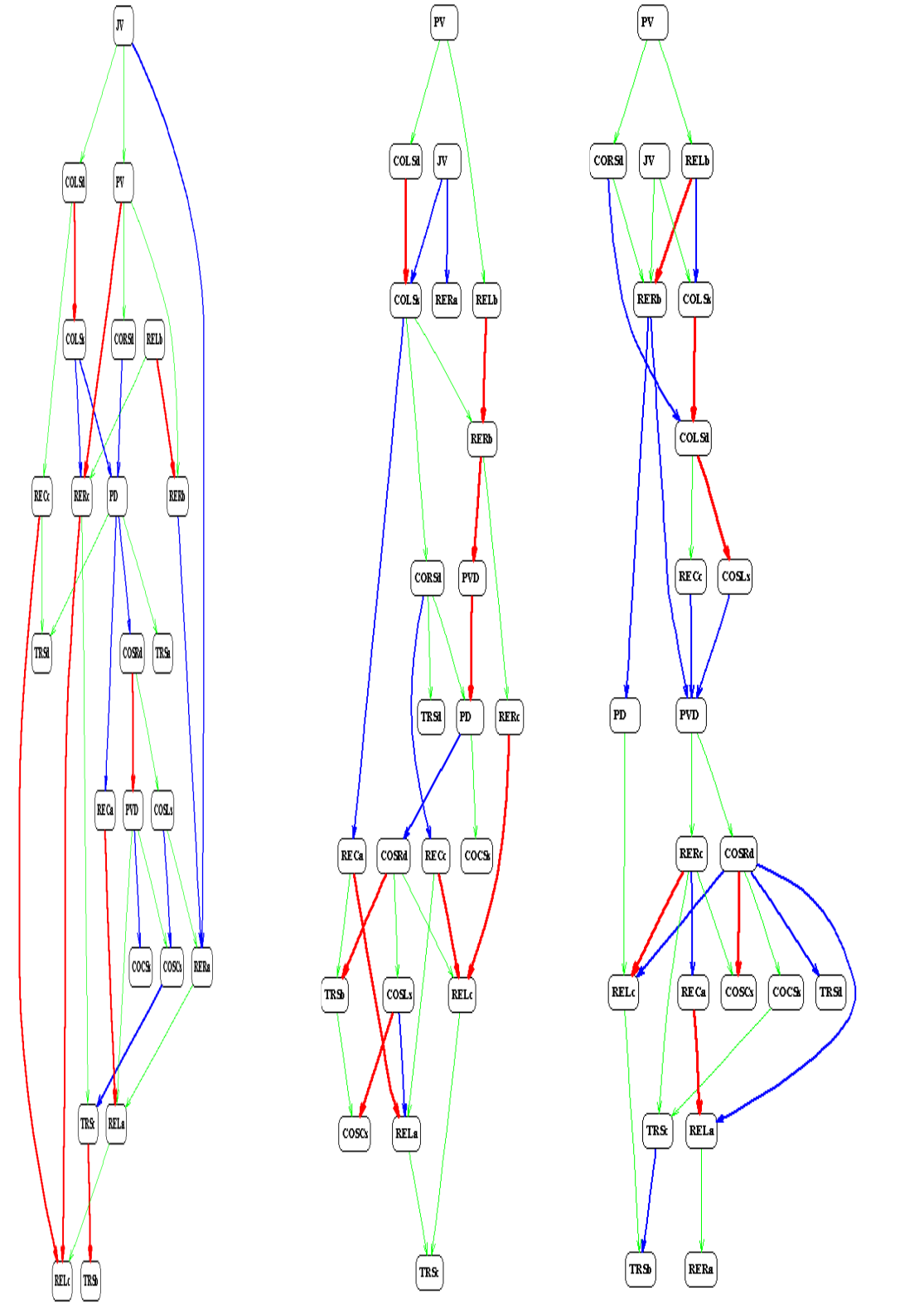
At the beginning of each implicative chain, there are the tasks that cause difficulty for the students of the specific age group. Specifically, in the first administration of the test the verbal problems, the translation tasks from the numerical line to the symbolic representation of the same and different denominator fraction addition, and the task of recognizing the fraction addition of different denominators on a rectangular area in which the number of divisions is twice the denominator were the tasks with the increased degree of difficulty. During the transition of students to 6th grade in both the second and third administration, these tasks are still the most difficult. A decisive role in the specific hierarchy of the degree of difficulty plays the interaction of both the type of task (problem-solving, translation), the involved initial representation (verbal representation, number line), and the relevant sub-concept (addition of different denominator fractions).

On the other hand, at the end of the implicative chains for each administration, easy tasks are found. In the first administration, the lowest-degree-of-difficulty tasks are the-same-and-different- denominator-symbolic treatment tasks, the recognition tasks of the same and different denominator fraction addition in the number line, and the translation task from the circular surface area to the symbolic representation of the same denominator fraction addition. In the second administration at the end of the implicative chain are the same tasks. In addition, there are the translation tasks from the symbolic representation of the same denominator fraction addition on the number line and on the circular area and the task of recognizing the same denominator fraction addition on a rectangular area. The number of tasks with a small degree of difficulty increases in the third administration. The symbolic treatment tasks of the same and different denominator fraction addition, the recognition tasks of the same denominator fraction addition in a variety of representations, the recognition tasks of the same denominator fraction addition in the number line, and the translation tasks of fraction-addition-symbolic representation to the circular surface and vice versa, are included.

In general, during the transition from 5th grade to 6th grade, the number of a-small-degree-of-difficulty multiple-representation-flexibility tasks increases. Therefore, some multi-representation flexibility tasks, which are considered in the first administration as of a-moderately-difficult tasks from the students of this age group, due to the teaching and cognitive development that takes place are easy tasks for them in the second and third

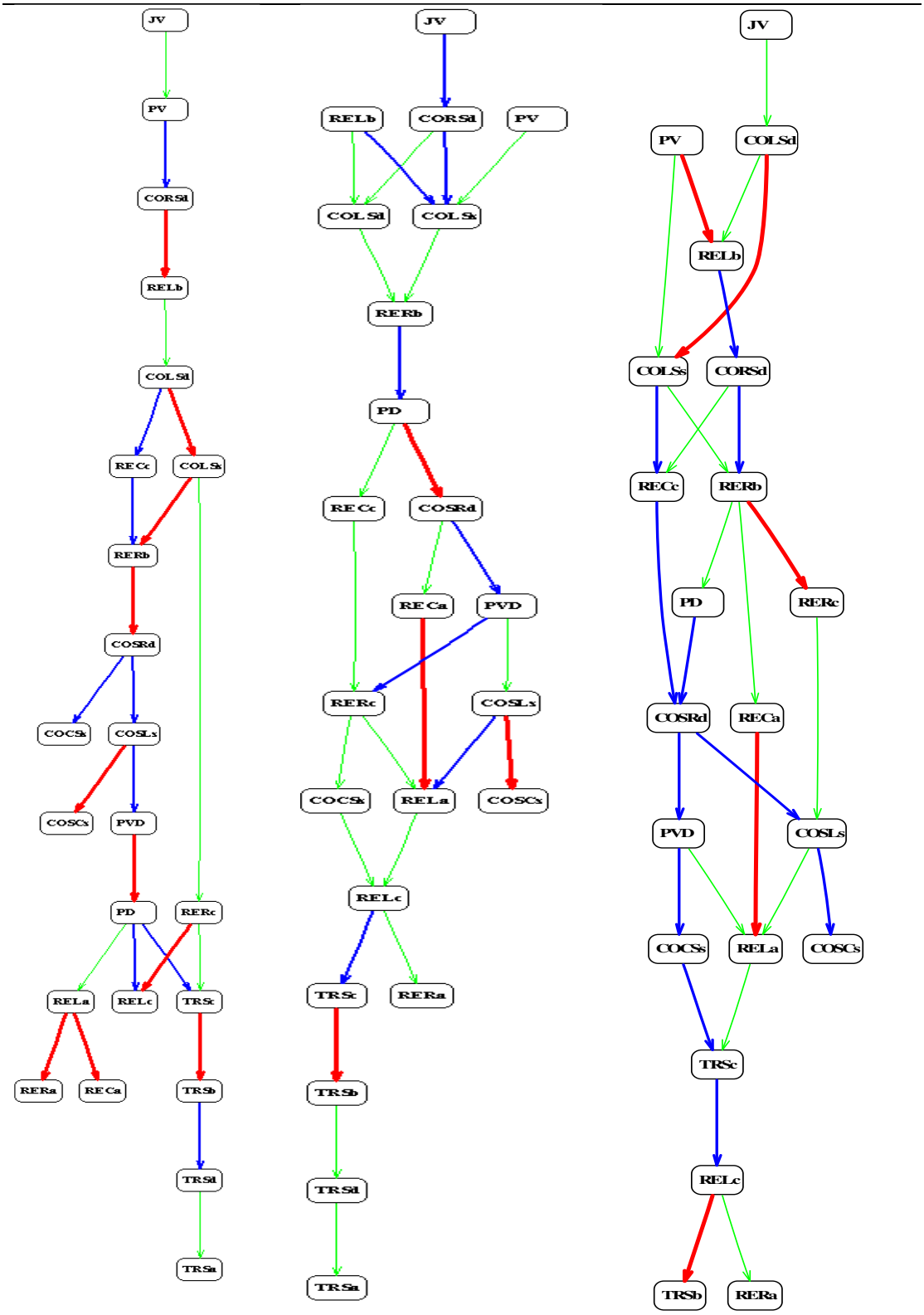
administrations. This result is in line with the improvement of the students' mean performance during one year in the various cognitive dimensions of the flexibility of using multiple representations of the fraction addition, according to the results of the multivariate analysis. However, despite the teaching and cognitive development during the three administrations, the number and type of tasks with an increased degree of difficulty do not change. This fact may explain why, despite the improvement in the performance of the multiple-representation-flexibility tasks and solving the fraction addition problems from the first to the third administration, the mean performance of the specific age group is at low levels. Exceptions are treatment tasks, which in all three administrations are considered easy tasks.

Figure 9 presents the implicative diagrams of the second age group students' answers, which move from primary to secondary school, in fraction-addition multiple-representation-flexibility tasks and problem-solving tasks in the first, second, and third administration. In the first administration, the tasks of solving a verbal problem, translation from a number line and a rectangular surface area in a symbolic representation of adding same denominator fractions and recognizing the addition of the same denominator fractions in the number line with divisions twice the denominator have a great degree of difficulty. In the second and third administrations, the same tasks have a great deal of difficulty. Students consider as demanding tasks also the translation task from number line to the symbolic representation of the same denominator fraction addition and recognition task of the same denominator fraction addition on a rectangular area with divisions twice the number of the denominator.



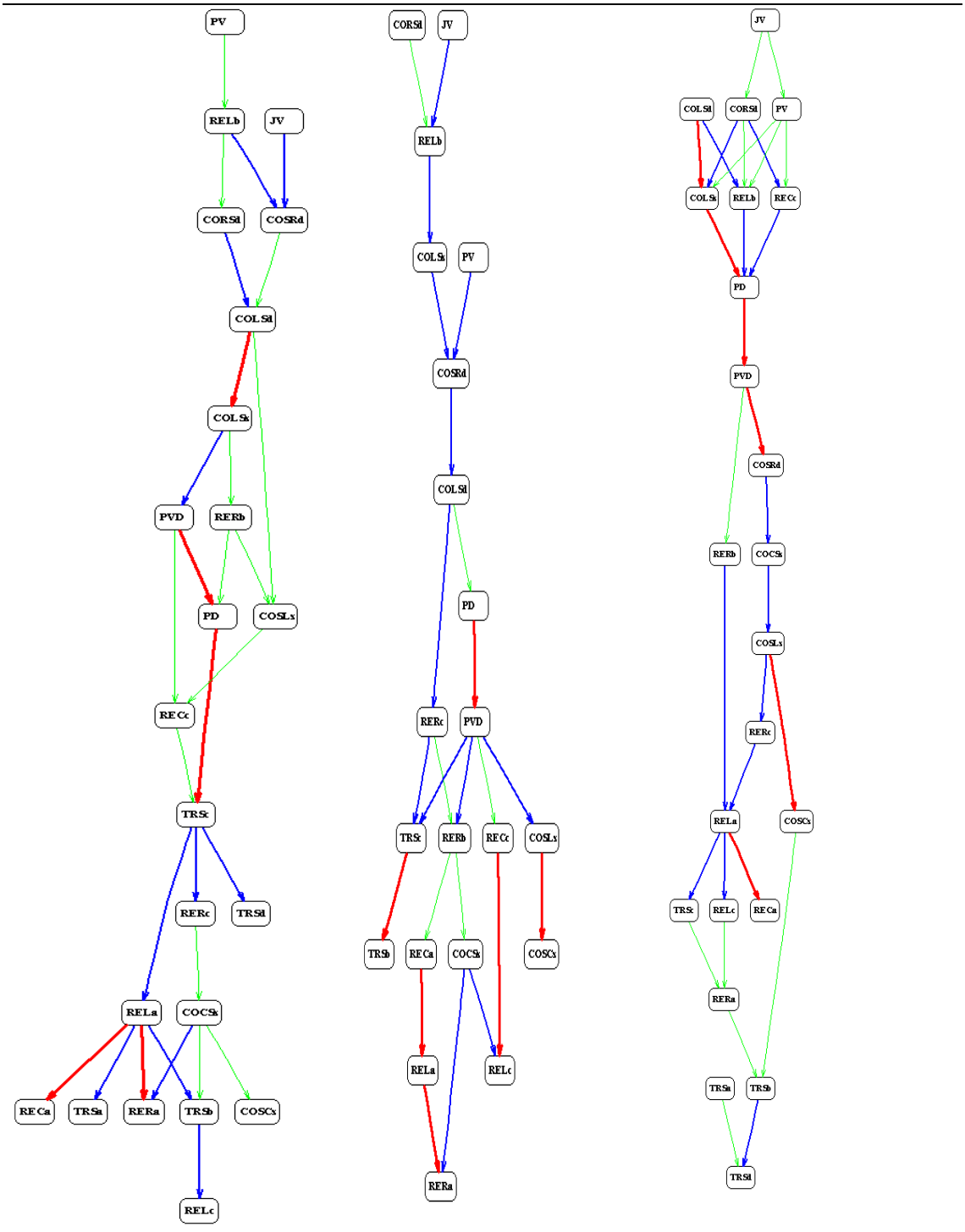
1st administration 2nd administration 3rd administration

Figure 8- Implicative chains of the first age group students' response to the tasks of the test in the three administrations



1st administration 2nd administration 3rd administration

Figure 9 - Implicative chains of the second age group students' response to the tasks of the test in the three administrations.



1st administration 2nd administration 3rd administration

Figure 10 - Implicative chains of the third age group students' response to the tasks of the test in the three administrations.

According to Figure 9, in the first administration, many of the tasks have a relatively small degree of difficulty for this age group. The specific tasks include the symbolic same- and-different-denominator-fraction-addition treatment tasks, the recognition tasks

of the same denominator fraction addition in rectangular and circular surface area and the number line, the recognition tasks of the same denominator fraction addition in the number line and rectangular surface translation tasks from the symbolic representation of the same denominator fraction addition to a circular surface area and vice versa, and the problems with main or auxiliary diagrammatic representation. In the second and third administrations, the number of tasks with a small degree of difficulty reduces. At the end of the implicative chain remains a part of the tasks that are considered as easy in the first administration that is the symbolic treatment tasks, the recognition same and different denominator fraction addition tasks on the rectangular area and the numerical line, respectively, and the translation task from the symbolic representation of adding same denominator fractions to the circular surface area. The same students, while last year they solve with great ease some recognition, translation, and problem-solving tasks during their transition to secondary education faced the same tasks as a moderate degree of difficulty. On the other hand, some tasks that these students faced as moderately demanding in the 6th grade during their transition to secondary school are demanding, and their correct solution implies success in other multiple-representation flexibility and problem-solving tasks.

In the last two administrations, there is an increase in the number of tasks that the students consider challenging or moderately complex. This result is related to the stagnation of the students' performance during their transition to secondary education in multi-representation flexibility tasks. Particularly characteristic of this performance pattern is the way they solve problems with main or auxiliary diagrammatic representation. The specific tasks in the first administration are treated by students as easy. In the second and third administrations, the diagrammatic problem and the verbal problem with auxiliary diagrammatic representation are included in the tasks with a high and moderate degree of difficulty, respectively. In all administrations, the interaction of the representation field, the function performed, and the relative sub-concept of fraction addition apparently plays an important role in determining the degree of difficulty of multi-representation flexibility tasks and solving fraction addition problems.

Figure 10 shows the implicative diagrams of the third age group students' answers, which move from primary to secondary school, in fraction-addition multiple-representation-flexibility tasks and problem-solving tasks in the first, second, and third administration. In the first administration, the verbal problems, the task of recognizing the addition of different denominator fractions in the number line with divisions twice the number of the denominator and the translation tasks from a rectangular area and the number line in a symbolic representation of adding different denominator fractions and the translation task from different-denominator fraction addition symbolic representation on a rectangular surface area are the most difficult tasks for students of the third age group. These tasks imply the students' success in others both in the second and the third administrations. It is worth mentioning the tasks in which the number line is involved have an increased degree of difficulty in the third administration as this geometric model is used to a limited extent in secondary education. On the other hand, the tasks that require finding the sum in same and different denominator fraction additions, recognize the addition of same denominator fractions in a variety of diagrammatic representations, recognize the addition of different-denominator fractions in the number line, and translation from symbolic addition of the same denominator fractions to the circular area and vice versa are the easiest in the first and third administration. The easy tasks increase

in the second administration. Some tasks are of moderate complexity for the students of the third age group in the first administration and are considered easy tasks in the second administration. However, some of these tasks are still of moderate difficulty in the third administration. Indicatively two tasks are mentioning: recognizing the addition of different denominator fractions addition on a circular surface area and recognizing the same denominator fraction addition on a rectangular surface area with subdivisions twice the number of the denominator. Nevertheless, in the third administration, the low-complexity tasks are more than the corresponding number in the first one. The fact that in the second and third administrations the number of lesser difficulty tasks increases confirms the results of the multiple analysis that indicates an increase in the mean performance of the second age group during the transition into secondary education.

4 Discussion

Fractions is one of the most important mathematical topics that students learn at the primary and secondary school levels; it is also one of the most multifaceted (Pedersen & Bjerre, 2021). The present study, considering previous studies results, aims to study developmentally the implicative relations between each representational flexibility and problem-solving ability component in fraction addition (Deliyianni et al. 2015) as the students move within and between primary and secondary school. According to the implicative analysis results, at the beginning of each implicative chain, in all the administrations of the test for each age group, there are the verbal problem-solving tasks (JV, PV). The tasks that follow reveal that the interaction of the representation field, the function performed by each representation, and the relative sub-concept of adding fractions play a decisive role in the difficulty of multiple representation flexibility and problem-solving tasks. We could assume that representational flexibility tasks are a prerequisite for success in problem-solving tasks.

However, various differentiations are apparent concerning the task complexity in each age group and administration. During the transition from primary to secondary education the number of tasks of high or moderate difficulty increases. In the case of transition within primary and secondary school, the number of tasks of moderate difficulty decreases while the number of tasks that the students deem as easy increases. The specific results are in line with the results of the multiple regression analysis, according to which the mean performance improves within the same educational level, primary or secondary. On the contrary, there is a mean performance decrease during the transition from primary to secondary education.

The fact that the tasks students in the first and third age groups find difficult still have the same degree of difficulty during the three administrations indicates that despite teaching and cognitive development, students have difficulty in understanding fraction addition in multiple representational tasks. Therefore, the results of the implicative relationships are in line with the low averages observed despite the improvement in the performance in the first and third age groups in the recognition, translation, and problem-solving problems of the addition of fractions. In addition, difficulties in understanding fraction addition are indicated by the fact that none of the age groups was able to highlight the common underlying concept of fraction addition as a whole in multiple-representation flexibility and problem-solving tasks with a variety of fields representing the addition of fractions. It is worth noting that the students encountered difficulties regarding the

geometric model of the number line in both translation and recognition tasks as in other studies (Gagatsis, Shiakalli, & Panaoura, 2003; Gagatsis, Kyriakides, & Panaoura, 2004; Shiakalli, & Gagatsis, 2005; Michaelidou, & Gagatsis, 2005; Deliyianni, & Gagatsis, 2009). Recent studies have shown the advantages of using the number line over the area models (e.g., Hamdan & Gunderson, 2017).–Therefore, Pedersen and Bjerre (2021) emphasize that visual representations remain a referent of fractions (Rau & Matthews, 2017) but acknowledge that relations between iconic or visual representations and the referent fraction can be opaque, making it important to use diverse representations when teaching fractions.

It would be interesting in the future to consider the differences that may arise in implicative relationships if intervention programs are implemented: a. in-service teacher training and b. teaching that emphasizes inter- and intra-representation transformations and problem-solving ability with various representations.

References

[12] Ainsworth, S. (1999). The functions of multiple representations. *Computer & Education*, 33, 131- 152.

[13] Christodoulou, T. & Gagatsis, A. (2014). Representational flexibility in graphs. *Proceedings of the 5th Conference of the Greek Council of Researchers in Mathematics Education (En.E.Di.M.)*. Greece: University of Western Macedonia (in Greek).

[14] Confrey, J., & Smith, E. (1991). A framework for functions: Prototypes, multiple representations and transformations. In R. G. Underhill (Ed.), *Proceedings of the 13th annual meeting of the North American Chapter of The International Group for the Psychology of Mathematics Education* (pp. 57-63). Blacksburg: Virginia Polytechnic Institute and State University

[15] Couturier, R. (2008). "CHIC: Cohesive Hierarchical Implicative Classification." *Studies in Computational Intelligence 127*: 41–53. https://doi.org/10.1007/978-3-540-78983-3_2.

[16] Deliyianni, E. & Gagatsis, A. (2009). La comprensione dell'addizione di frazioni nella scuola primaria : Il ruolo delle rappresentazioni multiple. *La Mathematica e la sua Didattica, 23*(3), 299-318.

[17] Deliyianni, E. and Gagatsis, A. (2013). Tracing the development of representational flexibility and problem solving in fraction addition: A longitudinal study. Educational Psychology : *An International Journal of Experimental Educational Psychology*, 33 (4), 427 – 442.

[18] Deliyianni, E., Gagatsis, A., Elia, I., and Panaoura, A. (2016). Representational flexibility and problem-solving ability in fraction and decimal number addition: A structural model. *International Journal of Science and Mathematics*, 14, 397 – 417.

[19] Duval, R. (2006). A Cognitive Analysis of Problems of Comprehension in a Learning of Mathematics. *Educational Studies in Mathematics*, 61, 103-131.

[20] Elia., I., Gagatsis, A., & Gras, R. (2005). Can we "trace" the phenomenon of compartmentalization by using the I.S.A.? An application for the concept of function. In R. Gras, F. Spagnolo & J. David (Eds.), *Proceedings of the Third International Conference I.S.A. Implicative Statistic Analysis* (pp. 175-185). Palermo, Italy: Universita degli Studi di Palermo.

[21] Elia, I., Gagatsis, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. *Learning and Instruction*, *17*, 658-672.

[22] Elia, I., Panaoura, A., Eracleous, A., & Gagatsis, A. (2007). Relations between secondary pupils' conceptions about functions and problem solving in different representations. *International Journal of Science and Mathematics Education, 5,* 533-556.

[23] Gagatsis, A., Shiakalli, M., & Panaoura, A. (2003). La droite arithmétique comme modèle géométrique de l'addition et de la soustraction des nombres entiers. *Annales de didactique et de sciences cognitives, 8,* 95-112.

[24] Gagatsis, A., Kyriakides, L., & Panaoura, A. (2004). Assessing the cross-cultural applicability of number line in conducting arithmetic operations using structural equation modeling: A comparative study between Cypriot, Italian and Greek primary pupils. *World Studies in Education, 5*(1), 85-101.

[25] Gagatsis, A. and Shiakalli, M. (2004). Ability to translate from one representation of the concept of function to another and mathematical problem solving. *Educational Psychology*, 24(5), 645-657.

[26] Gagatsis, A., Elia, I., & Mousoulides, N. (2006). Are registers of representations and problem-solving processes on functions compartmentalized in students' thinking? *Revista Latinoamericana de Educación Matematica, Special Volume,* 105-132.

[27] Gagatsis, A., Agathangelou, S., & Papakosta, V. (2010). Conceptualizing the role of pictures in problem solving by using the implicative statistical analysis. Acta Didactica Universitatis Comenianae – Mathematics, 10, 19-34.

[28] Gagatsis, A., Monoyiou, A., Deliyianni, E., & Philippou, A. (2010). Tracing 10th and 11th graders approaches in function tasks. Acta Didactica Universitatis Comenianae – Mathematics, 10, 51-67.

[29] Gagatsis, A., Deliyianni, E., Elia, I., & Panaoura, A. (2010). Tracing primary and secondary school students´ representational flexibility profiles in decimals. *Mediterranean Journal for Research in Mathematics Education, 9*, 211-222.

[30] Gagatsis, A., Deliyianni, E., Elia, I. & Panaoura, A. (2011). Explorer la flexibilité : le cas du domaine numérique. *Annales de didactique et de sciences cognitives, 16,* 21-38.

[31] Gagatsis, A., Christodoulou, T. & Elia, I. (2013). How young students deal with recognition and conversion tasks of graphs. *Acta Didactica Universitatis Comenianae – Mathematics*, 13, 1-16.

[32] Gagatsis, A., & Deliyianni, E. (2015). Mathematical Working Space relations with conversions between representations and problem solving in fraction addition.

Revista Latinoamericana de Investigación en Matemática Educativa – Relime. 17 (4-II), 249-267. DOI: 10.12802/relime.13.17412

[33] Gagatsis, A., Deliyianni, E., Elia, I., Panaoura, A., & Michael-Chrysanthou, P. (2016). Fostering Representational Flexibility in the Mathematical Working Space of Rational Numbers. *Bolema: Boletim de Educação Matemática, 30* (54), 287

[34] Gagatsis, A., Deliyianni, E., Elia, I., and Panaoura, A. (2017). Representational flexibility in fractions and decimals: A synthesis of recent studies. *Communication and Cognition*, 50 (3-4), 93-120.

[35] Gagatsis, A., Christodoulou, Th. & Elia, I. (2017). Young students' comprehension of data represented in graphs: The influence of technology and realistic problem solving. *Scientia Paedagogica Experimentalis*, LIV (1-2), 179-222.

[36] Gagatsis, A., Panaoura, A., Deliyianni, E., Nicolaou, S., & Elia, I. (2020). Are representations useful in Economic Mathematics? Students' beliefs and self-efficacy beliefs in the case of exponential and logarithmic functions. *La Matematica e la sua didattica*, Anno 28, n. 1, 63-85.

[37] Gras, R. (1979). *Contribution à L'étude Expérimentale et à L'analyse de Certaines Acquisitions Cognitives et de Certains Objectifs En Didactique Des Mathématiques*. Thèse de L'Université de Rennes 1.

[38] Gras, R., Couturier R., Blanchard J., Briand, H, Kuntz.P., and Peter P. (2004). Quelques critères pour une mesure de qualité de règles d'association. *Revue des nouvelles technologies de l'information* RNTI E-1, 3-30.

[39] Gras, R., Suzuki E., Guillet F., and Spagnolo F. (Eds) (2008). *Statistical Implicative Analysis, Theory and Applications.* Vol. 127. Studies in Computational Intelligence. Springer-Verlag, Berlin-Heidelberg Consultable

[40] Gras, R., Regnier, J. C., and Guillet, F. (2009*). Analyse statistique implicative : Une méthode d'analyse de données pour la recherche de causalités* Toulouse : Cépaduès Editions.

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

[42] Gras, R., and Couturier R. (2013). Spécificités de l'Analyse Statistique Implicative par rapport à d'autres mesures de qualité de règles d'association. *Educação Matemática Pesquisa*, 15(2).

[43] Goldin, G. A. (1998). Representational systems, learning, and problem solving in mathematics. *Journal of Mathematical Behavior*, 17(2), 137-165.

[44] Hamdan, N. and Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, 53(3), 587–596. https://doi.org/10.1037/dev0000252

[45] Janvier, C. (1987). Representation and understanding: The notion of functions as an example. In C. Janvier (ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 67-72). Hillsdale, N.J.: Lawrence Erlbaum Associates.

[46] Kuzniak, A., Tanguay, D., and Elia, I. (2016). Mathematical Working Spaces in schooling: an introduction, *ZDM Mathematics Education*, 48(6), 721-737.

[47] Markou, A., & Gagatsis, A. (2003). Rappresentazioni e apprendimento matematico: applicazioni nel campo delle frazioni. *La matematica e la sua didattica*, 2, 124-138.

[48] Michael – Chrysanthou, P., Gagatsis, A. & Papandreou, V. (2014). Solving addition tasks through the use of different modes of representations in primary school. *Acta Didactica Universitatis Comenianae* – Mathematics, 14, 19-42.

[49] Michaelidou, E., & Gagatsis, A. (2005). The geometrical model of number line as a representation of equivalence and addition of fractions. *Scientia Paedagogica Experimentalis, 42*, 185-204.

[50] Nicolaou, S., Gagatsis, A, Deliyianni, E., Panaoura, A., Elia, I., & Anastasiadou, S. (2017). Tracing the Beliefs and Self-efficacy Beliefs of Undergraduate Economic Sciences Students: The Case of the Représentations of Functions. In Régnier, J. C., Gras, R., Couturier, R., & Bodin, A. (Eds.), *Analyse Statistique Implicative* (pp. 459-476). Université Bourgogne Franche-Comté –Besançon : icar. ISBN: 978-2-9562045-1-0

[51] Nicolaou, S., Gagatsis, A., Panaoura, A., Deliyianni, E., Elia, I., & Televantou, I. (2020). Economic sciences students' understanding on representation tasks concerning the concept of function. In J.-C. Régnier, R. Gras, M. Henry, R. Couturier, & G. Brousseau (Eds.), *Analyse Statistique Implicative. Cadre théorique en relation étroite et au service de multiples disciplines* (pp. 281-302). Besançon: Université Bourgogne Franche-Comté-Besançon: icar. ISBN: 978-2-9562045-3-4.

[52] Panaoura, A., Gagatsis, A., Deliyianni E., & Elia, I. (2009). The structure of students' beliefs about the use of representations and their performance on the learning of fractions. *Educational Psychology*, 29(6), 713-728.

[53] Panaoura, A., Gagatsis, A., Deliyianni, E., & Elia, I. (2010). A model on the cognitive and affective factors for the use of representations at the learning of decimals. *Educational Psychology,30*(6), 713-734.

[54] Panaoura, A., Michael-Chrysanthou, P., Gagatsis, A., Elia, I., & Philippou, A. (2017). A structural model related to the understanding of the concept of function: Definition and problem solving. *International Journal of Science and Mathematics Education*, *15*(4), 723-740.

[55] Pedersen, P.L. and Bjerre, M. (2021). Two conceptions of fraction equivalence. *Educational Studies in Mathematics*, 107, 135–157. https://doi.org/10.1007/s10649-021-10030-7

[56] Rau, M. A., and Matthews, P. G. (2017). How to make 'more' better? Principles for effective use of multiple representations to enhance students' learning about fractions. *ZDM—International Journal on Mathematics Education*, 49(4), 531–544. https://doi.org/10.1007/s11858-017-0846-8

[57] Reading, C. (1999). Understanding data tabulation and representation. In O.Zaslavsky (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education* (Volume 4, pp.97-104). Haifa, Israel.

[58] Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification-The case of function. In E. Dubinsky, & G. Harel (Ed.), *The concept of function: Aspects of epistemology and pedagogy* (pp. 59-84). United States: Mathematical Association of America.

[59] Shiakalli, M., & Gagatsis, A. (2005). The geometrical model of the number line in the teaching of whole number addition and subtraction. *Scientia Paedagogica Experimentalis, 42*, 167-184.

[58] Sierpinska, A. (1992). On understanding the notion of function. In E. Dubinsky, & G. Harel (Ed.), *The concept of function: Aspects of epistemology and pedagogy* (pp. 25-58). United States: Mathematical Association of America.

[60] Theodoulou, R., Gagatsis, A., & Theodoulou A. (2004). Un'immagine vale più di mille parole…Ma che tipo di immagine risulta più efficace nelle attività di problem solving matematico degli studenti? *La matematica e la sua didattica*, 2, 4-32.

[61] Yerushalmy, M. (1997). Designing representations: Reasoning about functions of two variables. *Journal for Research in Mathematics Education*, 27(4), 431-466.

TRIANGULATION EN RECHERCHE QUALITATIVE À L'AIDE DE L'ANALYSE STATISTIQUE IMPLICATIVE[1]

Miguel Wilhelmi[2], Olga Belletich, Haritz Iribas, Jaione Abaurrea, Aitzol Lasa

TRIANGULATION IN QUALITATIVE RESEARCH USING SIA

RÉSUMÉ

En Sciences de l'Éducation et, plus généralement, en Sciences Humaines et Sociales, les approches qualitatives et descriptives sont utilisées abondamment. Du point du vue méthodologique, la triangulation est à son tour largement utilisée. Cette triangulation peut intégrer des méthodes quantitatives se référant à des variables catégorielles (ou des variables qualitatives ou des variables « attribut »). Cette spécificité est souvent mal maîtrisée par les chercheurs, qui peuvent manquer, dans ces domaines de recherche, de la formation technique nécessaire. Des considérations générales sont fournies et appliquées à une étude particulière, à savoir : « décrire le profil des étudiants pour devenir enseignants à l'École maternelle ou à l'École élémentaire ».

Mots-clés : *recherche qualitative, triangulation, variables catégorielles, formation des enseignants.*

ABSTRACT

In Education Sciences and, more generally, in Human and Social Sciences, qualitative and descriptive approaches are used extensively. From a methodological point of view, triangulation is in turn widely used. This triangulation can integrate quantitative methods referring to categorical variables (or qualitative variables or 'attribute' variables). This specificity is often not well understood by researchers, who may lack the necessary technical training in these research areas. General considerations are provided and applied to a particular study, namely: "describing the profile students to become kindergarten or primary schools' teachers".

Keywords: *qualitative research, triangulation, categorical variables, teacher training.*

1 Introduction

En Sciences de l'Éducation et, plus généralement, en Sciences Humaines et Sociales, les approches qualitatives et descriptives sont utilisées abondamment. Du point du vue méthodologique, la triangulation est à son tour largement utilisée. « De manière générale, la triangulation méthodologique peut être définie comme le fait d'appréhender un objet de recherche d'au moins deux[3] points de vue différents » (Caillaud et Flick, 2016, 227). Ainsi, on peut distinguer au moins cinq types de triangulation :

---

[1] *Reconnaissances*. Ce travail a été effectué dans le cadre du projet : CENEDUCA6/2019, Gouvernement de Navarre - Département de l'Éducation.

[2] Université Publique de Navarre, Département d'Informatique, Statistique et Mathématiques, Campus Arrosadia s/n, 31006 Pampelune, Espagne, miguelr.wilhelmi@unavarra.es; olga.belletich@unavarra.es; haritz.iribas@unavarra.es; jaione.abaurrrea@unavarra.es; aitzol.lasa@unavarra.es

[3] Même si la triangulation suggère la prise en compte d'au moins trois points de vue, la définition classique évoque « au moins deux … » ; cette définition « paradoxale » est une généralisation de la pratique standard des chercheurs. Ainsi, du point de vue technique, les chercheurs font recours normalement au moins trois instruments de récolte de données ou statistiques ; mais, du point de vue théorique, une

1. *Théorique*. Utilisation d'au moins deux approches théoriques pour l'analyse des données.

2. *Outils de cueillette*. Utilisation d'au moins deux outils de cueillette (des entrevues, des observations naturalistes, des analyses de documents, des questionnaires, etc.)

3. *Méthodes statistiques*. Utilisation d'au moins deux méthodes statistiques pour l'analyse des données.

4. *Jugement d'experts*. Analyse des données par jugement d'experts, c'est-à-dire, contraster la stabilité et consistance des appréciations d'au moins deux personnes réputées dans le domaine objet d'étude.

5. *Sources*. Utilisation d'au moins deux sources (étudiants, enseignants, livres de texte, programmes éducatifs, etc.)

Dans la recherche qualitative, la triangulation vise fondamentalement la *validité interne*, c'est-à-dire, le degré de confiance en les résultats établis à propos de la connexion entre les variables d'étude. Si on a confiance, alors on affirme que la connexion entre variables est fiable : aucune autre variable, sauf celle que l'on a été prise en compte, ne peut pas être considérée comme la cause probable du phénomène objet d'étude.

> « La validité interne implique de vérifier si les observations sont effectivement représentatives de la réalité ou crédibles, c'est-à-dire d'évaluer si ce que le chercheur observe est vraiment ce qu'il croit observer. » (Drapeau, 2004, 81)

Dans ce travail, on va montrer comment une triangulation des méthodes statistiques permet d'aboutir à des résultats consistants et crédibles. En particulier, on va à montrer comment on peut établir des résultats par le biais de l'Analyse Statistique Implicative (ASI).

2 L'objet d'étude et l'objectif de la recherche

Les programmes éducatifs pour l'École primaire ont beaucoup évolué. On y trouve une intégration de contenus pédagogiques, des indicateurs des apprentissages et des standards d'évaluation. Mais, il faut absolument souligner que le changement des programmes n'implique pas nécessairement un changement de la gestion et fonctionnement des *milieux didactiques* (Brousseau, 1998). D'ailleurs, si on n'envisage pas une « révolution » dans la formation initiale et continue des enseignants, alors le changement des pratiques d'enseignement ne se produira que par des raisons aléatoires ou fallacieuses.

> « Si le discours institutionnel a profondément évolué, les pratiques, elles, ont du mal à se transformer car mettre en place des situations de recherche de problèmes dans la classe suppose des compétences de gestion de classe qui ne peuvent se révéler qu'avec une solide formation. Il faudrait donc continuer à faire évoluer le rapport personnel et médiatif des professeurs et des formateurs […] aux problèmes […] Mais tout ceci ne sera possible que si la formation des maîtres est préservée et outillée par la recherche en éducation. » (Coppé et Houdement, 2009, 23)

---

comparaison de méthodes est aussi considérée comme une triangulation.

Alors, ce changement du rapport personnel et *médiatif*[4] des professeurs aux problèmes ne sera-t-il possible sauf que l'on développe des véritables compétences mathématiques chez les enseignants. Pour que cela se passe, il faut absolument et préalablement connaitre les caractéristiques de la population « étudiant pour devenir professeur à l'École Maternelle (0-6ans)[5] ou à l'École Élémentaire (6-12ans)[6] » ; c'est-à-dire, il faut répondre à la question : Quel est le profil de ces étudiants universitaires ?

En résumé, l'objet d'étude est «étudiant universitaire pour devenir professeur à l'École Élémentaire » et l'objectif de la recherche est « décrire le profil de ces étudiants». En outre, le travail a une vocation « pédagogique » pour la communauté des chercheurs en Sciences de l'éducation, c'est-à-dire, de rendre possible un schéma d'analyse pratique où le discours est centré sur les possibilités d'interprétation des instruments statistiques et non sur les instruments eux-mêmes.

3 L'instrument de construction des données : questionnaire

Le questionnaire complet peut être consulté à l'annexe A. Les réponses des étudiants sont croisées avec les valeurs disponibles à la base institutionnelle de données : âge, sexe, études préalables d'entrée à l'université (baccalauréat ou formation professionnelle), profil d'enseignement pré-universitaire (technico-scientifique ou en Sciences sociales et humaines), langue d'étude (espagnol ou basque), Diplôme (pour devenir enseignant à l'École Maternelle ou à l'École Élémentaire).

Le questionnaire est un questionnaire d'opinion ou de perception relative aux différentes disciplines requises à la formation initiale des enseignants. Le questionnaire est organisé en 3 blocs : a) questions sur l'intérêt ou le goût pour les différentes disciplines ; b) questions sur le milieu socio-familial ; c) questions à propos de l'intérêt ou le goût pour les mathématiques. Plus précisément, l'information récoltée est la suivante :

a) À propos du goût pour les disciplines
   - [RCM] Information sur les disciplines plus appréciées dans le parcours pré-universitaire.
   - [RCM] Information sur la motivation pour devenir enseignant.
   - [RCM] Information sur l'opinion de l'obligation d'étudier des différentes disciplines au Lycée.

[4] *Médiatif.* « La catégorie du médiatif […] s'organise autour de trois valeurs fondamentales : 1) faits rapportés, y compris selon le ouï-dire ; 2) faits inférés ; 3) faits de surprise » (Zlatka, 1994, 9). Autrement dit, le changement du rapport personnel et médiatif des professeurs implique l'évolution de l'*épistémologie des enseignants* (Brousseau, 1997), notamment de l'*épistémologie spontanée*, vers une compétence mathématique et didactique par le biais de véritables *transpositions didactiques* (Chevallard, 1989).

[5] Bien entendu, l'« École Maternelle » comprend l'ensemble : « Toute Petite Section » (2-3ans), « Petite Section » (3-4ans), « Moyenne Section » (4-5ans) et « Grande Section » (5-6ans) et en plus l'attention aux plus petits (0-2ans).

[6] En Espagne, l'École Élémentaire inclut la classe de sixième, c'est-à-dire, les enseignants de cette étape éducative peuvent développer sa carrière auprès des élèves du Cours Préparatoire (CP), Cours Élémentaire 1re année (CE1), Cours Élémentaire 2e année (CE2), Cours Moyen 1re année (CM1), Cours Moyen 2e année (CM2) et Sixième ; en somme, auprès des élèves de 6-12 ans.

- [RELF] Information sur l'opinion de l'obligation d'étudier des différentes disciplines à l'Université pour devenir enseignant.

b) À propos de l'environnement socio-familial

- [RCMO] Information sur le niveau de formation des parents.
- [RCM] Information sur la perspective à propos de ce que la formation universitaire apportera d'intéressant.
- [RCM] Information sur l'influence du contexte familial sur l'intérêt ou goût pour les différentes disciplines.
- [RCMO] Information sur l'activité professionnelle des parents : sont-ils des enseignants ?
- [ROC] Information sur la formation STIM des parents.
- [ROC] (Optionnel : si les parents n'ont pas de formation STIM) Information sur la l'influence du contexte familial sur l'intérêt ou goût pour les disciplines STIM.

c) À propos des mathématiques

- [RCMO] En tournant vers l'avenir, information sur l'intention d'étudier des Mathématiques.
- [ROC] Des commentaires ou des suggestions additionnelles.

Les types de réponses dans le questionnaire permettent de contrôler la dispersion des réponses, notamment par le biais de réponses à choix multiple (RCM) ou à échelle Likert à choix forcé (RELF), mais de ne pas contraindre les réponses reliées à la réalité personnelle et socio-familiale, en ayant recours à des réponses ouvertes partiellement (RCMO) ou totalement (ROC).

4 Triangulation : les méthodes statistiques

On utilise quatre méthodes statistiques pour l'analyse des données :

1. Corrélation.
2. Analyse des correspondances.
3. Analyse de similarité.
4. Analyse statistique implicative.

La cohérence des résultats obtenus par ces 4 méthodes est indispensable à la validité. D'une manière générale, la validité d'un « instrument » indique le degré de précision dans la mesure du « construit » qu'il est censé mesurer —ici, profil des étudiants universitaires en formation initiale des enseignants (FIE) — et, encore, la validité indique si l'instrument peut être utilisé pour l'objectif visé. Mais, cette précision dans la mesure ne peut pas être atteinte sauf si l'on contrôle l'adéquation des méthodes statistiques à la nature des données récoltées. Dans les prochaines sections, on discutera à propos de cette adéquation.

4.1 Corrélation

Les logiciels statistiques ont des outils pour obtenir automatiquement les graphes de corrélations ; bien sûr, cela offre de grands avantages ! Mais, d'une certaine manière, il faut « se méfier » de ces graphes et faire recours aux données. On va s'expliquer.

D'une part, certains graphiques sont réalisés en fonction d'une échelle de catégories. En fait, il existe un consensus assez répandu selon lequel l'interprétation générale des valeurs du coefficient de corrélation linéaire de Pearson est la suivante (en considérant les valeurs absolues) :

- Entre –0,09 et 0,09 : corrélation inexistante.
- Entre 0,10 et 0,19 (entre –0,19 et –0,10) : corrélation faible directe (indirecte).
- Entre 0,20 et 0,49 (entre –0,49 et –0,20) : corrélation modérée directe (indirecte).
- Entre 0,50 et 0,69 (entre –0,69 et –0,50) : corrélation forte directe (indirecte).
- Entre 0,70 et 1,00 (entre –1,00 et –0,70) : corrélation très forte directe (indirecte).

Toutefois, cette échelle est subjective. La définition d'une corrélation faible ou forte dépend de chaque cas. Une même valeur peut être interprétée comme faible ou forte, en fonction de la situation et de la connaissance que l'on a de celle-ci. En plus, bien sûr le caractère significatif dépend de la taille de l'échantillon.

D'une autre part, les graphes peuvent montrer la puissance des relations par le biais d'un tableau à double entrée où on représente la puissance par la dimension des aires des petites disques. Ainsi, une corrélation parfaite (si directe, puissance = 1 ; si inverse, puissance = –1) est représentée par un disque de rayon 1 (sous l'échelle adoptée). La taille des disques qui représentent les autres relations établies « non-parfaites » est proportionnelle à leur puissance. Dans ces disques-ci, il est bien compliqué de relever les différences et nuances.

En résumé, il est nécessaire de scruter les données pour bien comprendre ces graphes et pour mieux classer les relations entre variables, selon la puissance et la direction de ces relations.

Finalement, il faut absolument retenir que tout ce que l'on a discuté est relié aux relations linéaires et, en général, il est donc convenable d'examiner d'autres possibles relations entre les variables. En plus, la corrélation n'est pas une relation « directionnelle et temporelle » est donc ne peut pas être interprétée comme « de cause à effet ». Autrement dit, on peut conclure que deux variables sont liées ; cependant, on n'a pas d'information pour assurer si une variable influence l'autre ou, encore, s'il n'y a pas d'influence car les deux variables surviennent conjointement. En bref, la corrélation statistique ne peut être le seul outil pour l'étude de la causalité.

4.2 Analyse des correspondances

L'Analyse des correspondances (AC) est un type particulier d'Analyse des Composantes Principales (ACP). L'ACP doit être réservée aux variables quantitatives qui

représentent un contrôle ou une mesure, par exemple, les résultats des examens ; ce n'est pas le cas dans cette recherche.

Une ACP « classique » ne peut pas donc être utilisée proprement puisque les variables ne sont pas quantitatives, mais catégoriques (ou variables qualitatives ou variables « attribut »). Bien entendu, « 1 » ou « 0 » ne sont pas de valeurs sur une échelle, mais la présence ou l'absence d'une caractéristique. En fait, les données catégorielles n'ont pas d' « ordre ». Il est donc nécessaire d'appliquer une « analyse des correspondances » (AC), qui n'est qu'une méthode ACP spécifique pour des variables catégorielles.

4.3 Analyse de similarité

Pour certaines études, on peut justifier « mélanger » les variables contingentes et les variables supplémentaires, notamment dans études exploratoires, c'est-à-dire, une étude dont la finalité est de préciser la nature de quelque chose (son type ou sa catégorie), d'identifier des comportements ou des faits non encore connus, de clarifier un problème mal ciblé ou de fournir d'information contrasté à propos d'un « construit » encore mal formulé ; ce n'est pas non plus le cas dans cette recherche.

On connait déjà des rapports entre le contexte socio-familial et la performance académique (Lambert, 2012). En plus, on connait bien l'influence du contexte sur le goût et l'intérêt pour les différentes matières (Cosnefroy, 2007). Encore plus, l'*épistémologie spontanée des enseignants* (Brousseau, 1997) sous-tend nombre de réflexions, analyses et comportements des enseignants en formation initiale (EFI). Finalement, on a appris que les EFI ont un *stéréotype sur l'Éducation* (Belletich et Villarreal, 2014).

Ce savoir sociologique et pédagogique suggère de faire la différence entre les variables contingentes et les variables supplémentaires. Alors, dans l'Analyse de similarité (AS), on étude les classes constituées par les variables contingentes et la contribution des variables supplémentaires à ces classes.

4.4 Analyse statistique implicative

L'ASI ne donne pas nécessairement de relations de cause à effet, c'est-à-dire, en général, l'implication statistique ne représente pas la causalité, mais un rapport d'« inclusion ».

> "*The statistical implication admits an interpretation (mathematically isomorphic) according to the Set Theory*. Given a set *E*, we defined *P* (*Q* respectively) like the set of elements *x* in *E* (*x*∈ *E*) that verify the proposition *p* (*q* respectively): $P = \{x \in E | p(x)\}$ ($Q = \{x \in E | q(x)\}$ respectively). Then "*p* implies *q*" is equivalent to affirm that "P is included in Q" ($P \subseteq Q$)." (Lacasta et Wilhelmi, 2008, pp. 111-112)

Autrement dit, même si dans certains cas on peut tirer des résultats de « causalité » entre variables ou, plus largement, on peut identifier l'influence d'une variable (cause) sur une autre variable (effet), en général la relation implicative n'est qu'un indice d'une causalité « à évaluer ». En résumé, étant donnée une relation implicative, il faut toujours essayer de répondre à la question clef suivante : Est-il possible d'affirmer qu'un fait, un objet, un processus ou un état dans un contexte contribue (de manière vérifiable) à un autre fait, objet, processus ou état dans ce même contexte ?

Avant de répondre à cette question capitale, il faut clarifier le terme « explication », car certains chercheurs créent recommandable une certaine modestie dans la présentation de résultats et ils choisissent des expressions « faibles » du type : <il est probable>, <en notre opinion>, <il est convenable>, etc. ; au contraire, d'autres chercheurs utilisent des expressions « forts » du type : <nous avons démontré>, <ça implique que>, <la cause est>, etc., et ils essayent une réponse aux questions génériques suivantes : *Quelle est la cause d'un certain phénomène ? Sous quelles lois et conditions générales un certain phénomène se produit-il ?* (Wilhelmi, Font et Godino, 2005).

Dans cette recherche qualitative, l'objectif ciblé —à savoir, la description du profil des étudiants en FIE— n'est pas catégorique, c'est-à-dire, immutable, absolue, inconditionnelle et universelle. De fait, l'objectif dernier, comme toujours dans le domaine éducatif, est d'identifier un certain « statut » pour agir et le transformer en un autre statut plus adéquat pour atteindre les apprentissages envisagés. En bref, en éducation les objectifs sont toujours, per se, conditionnels, c'est-à-dire, restreints par des conditions spécifiques ou par un contexte déterminé. En plus, on n'essaye pas d'établir une loi « générale » ; on veut tout simplement de donner une explication « significativement » plausible aux relations entre variables établies dans le contexte de référence.

Pour aboutir à cet objectif, on va à profiter de l'ASI y compris les variables supplémentaires. Ici, à différence de l'AS, on utilise les variables supplémentaires car l'objectif est la description de groupes d'étudiants par rapport à leur contexte socio-familial et à leur intérêt personnel. Il faut souligner que la méthode de recherche est *ex-post-facto* (et donc non-expérimentale et sans manipulation de variables indépendantes), où on ne dispose pas d'information suffisante pour faire la différence entre les variables contingentes et les variables supplémentaires du point de vue de la *causalité*, car on ne peut pas distinguer « temporalité » et « direction » [7] dans les relations établies.

5 Résultats et discussion

5.1 Population, échantillon et variables

La population est « étudiant pour devenir professeur à l'École Maternelle (0-6ans) ou à l'École Élémentaire (6-12ans) en Espagne ». L'échantillon est 170 étudiants d'une université espagnole. Les variables retenues sont au nombre de 37 (annexe B) : 13 supplémentaires (tableau B.1) et 24 contingentes (tableau B.2). Les distributions des fréquences absolues et les fréquences relatives exprimées en pourcentage des variables figurent dans le tableau 1.

[7] Les caractéristiques fondamentales de la *relation causale* sont : a) la *temporalité*, c'est-à-dire que la cause précède l'effet ; b) la *direction*, c'est-à-dire que la relation va de la cause à l'effet ; et c) l'*association*, qui est comprise comme une quantification de la constance de la relation.

Variable	IDI s	SEX s	EDA s	INF s	PRI s	ESO s	BFP s	CAM s	PES s	SES s
Fréquence absolue	111	133	16	70	119	75	39	66	57	26
Pourcentage	64,9	77,8	9,4	40,9	69,6	43,9	22,8	38,6	33,3	15,2
Variable	IPS s	PPD s	PEC s	GPL	GPH	GPM	GPC	GPA	GPD	GPX
Fréquence absolue	26	112	44	69	83	58	59	31	19	22
Pourcentage	15,2	65,5	25,7	40,4	48,5	33,9	34,5	18,1	11,1	12,9
Variable	QEI	QET	QEP	NEM	NEC	NEX	FNT	FNC	FCP	FCT
Fréquence absolue	131	96	53	34	32	24	109	63	70	67
Pourcentage	76,6	56,1	31	19,9	18,7	14	63,7	36,8	40,9	39,2
Variable	FCM	FGC	FGL	YMG	YMP	YMO	YMN			
Fréquence absolue	48	73	56	62	82	74	28			
Pourcentage	28,1	42,7	32,7	36,3	48	43,3	16,4			

Tableau 1 – Fréquences absolues et fréquences relatives (pourcentage)

5.2 Corrélation

En nous basant sur l'échelle établie ci-avant, nous constatons que l'intensité de la corrélation entre variables est en général faible ou inexistante. En effet, la distribution de 666 relations entre variables est : a) 611 corrélations inexistantes ou très faibles (<0,2 en valeur absolue) ; b) 55 corrélations modérées (entre 0,2 et 0,5), mais seulement 16 ont une puissance supérieure à 0,3 (dont 10 directes et 6 inverses) ; et, finalement, c) il n'y a que 4 corrélations fortes (>0,5 en valeur absolue), dont 3 directes et 1 inverse.

La figure 1 est obtenue avec le logiciel R (https://www.r-project.org/). On peut y voir la difficulté à discerner la puissance des corrélations. C'est pourquoi que l'on a fait recours directement aux données. Le tableau 2 monstre ces corrélations. Finalement, la figure 2 est construite par révision systématique des données.

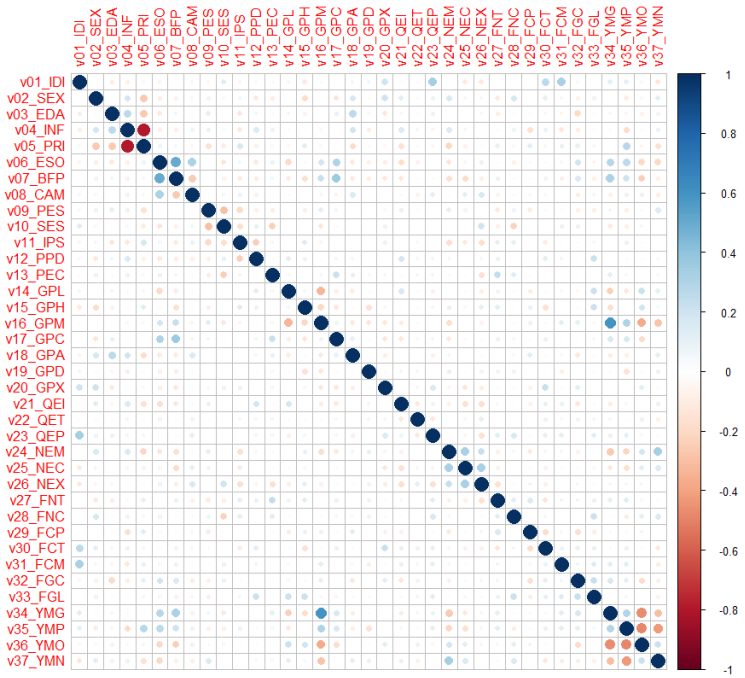


Figure 1 – Graphe 1 des intensités de corrélation entre variables (supplémentaires incluses)

Corrélation	DIRECTE	INVERSE
Forte	v01_IDI *vs.* v07_BFP : 0.5026 v06_ESO *vs.* v07_BFP : 0.5026 v16_GPM *vs.* v34_YMG : 0.5902	v04_INF *vs.* v05_PRI : -0.7940
Modérée	v01_IDI *vs.* v23_QEP : 0.3338 v01_IDI *vs.* v31_FCM : 0.3229 v06_ESO *vs.* v08_CAM : 0.3160 v06_ESO *vs.* v17_GPC : 0.3005 v07_BFP *vs.* v17_GPC : 0.3678 v07_BFP *vs.* v34_YMG : 0.3148 v16_GPM *vs.* v35_YMP : 0.3013 v24_NEM *vs.* v25_NEC : 0.3245 v24_NEM *vs.* v37_YMN : 0.3340 v25_NEC *vs.* v26_NEX : 0.3241	v34_YMG *vs.* v36_YMO : –0.4623 v35_YMP *vs.* v36_YMO : –0.4840 v35_YMP *vs.* v37_YMN : –0.4248 v14_GPL *vs.* v16_GPM : –0.3375 v16_GPM *vs.* v36_YMO : –0.3764 v34_YMG *vs.* v37_YMN –0.3009

Tableau 2 – Corrélations fortes et modérées (> 0,3)

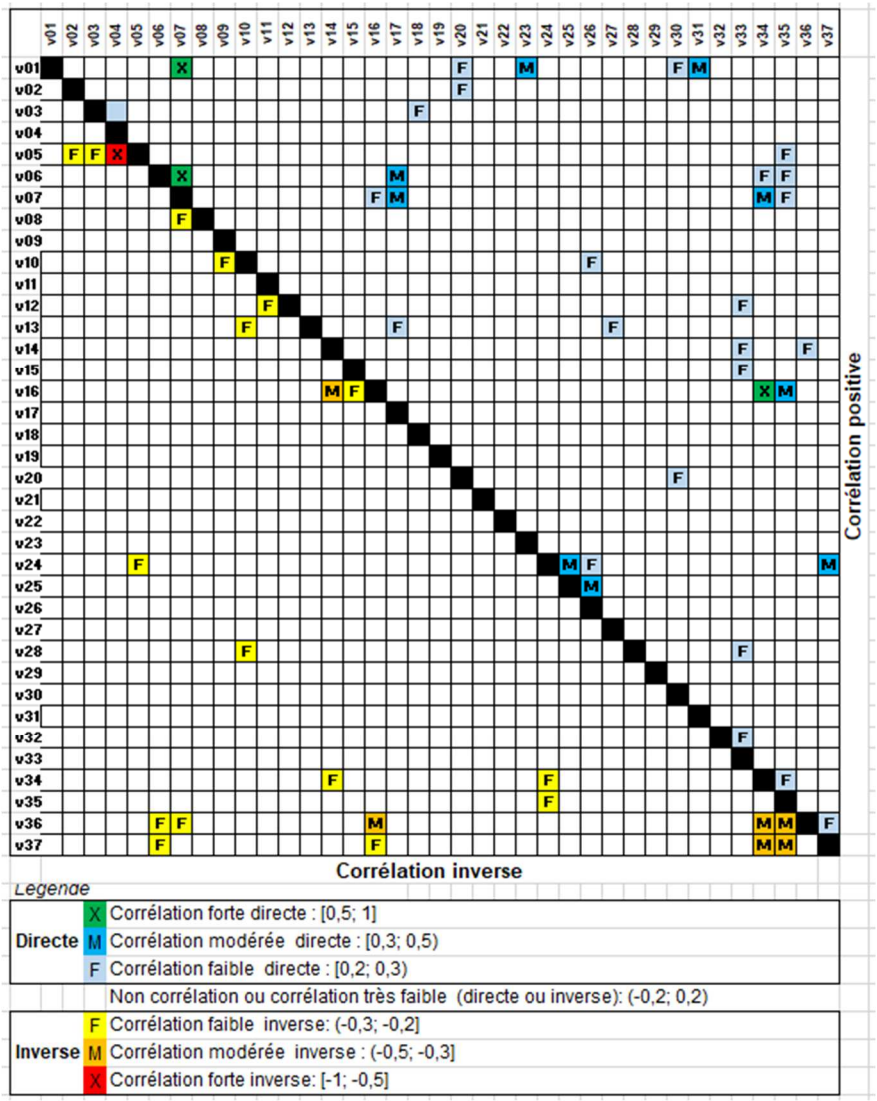


Figure 2 – Graphe 2 des intensités de corrélation entre variables (supplémentaires incluses)

Les corrélations montrent quelques relations entre variables tout-à-fait prévisibles et d'autres relations révélatrices. En particulier, on relève l'importance de la variable

« idiome (IDI) », c'est-à-dire, la langue véhiculaire commune à toutes les matières, qui est la même que celle utilisée à l'étape pré-universitaire. Par exemple, la variable IDI est liée à la variable « BFP » (voie technico-scientifique en Première et Terminale, 16-18 ans) et cette variable BFP est à son tour liée au goût pour les mathématiques et les sciences.

Les corrélations permettent aussi d'identifier deux groupes d'étudiants : d'une part, les « *scientophobes* », ceux qui n'aiment pas les sciences, y compris les mathématiques, et, d'autre part, les « *scientophiles* » qui aiment bien les sciences. Même si l'on aurait pu identifier *a priori* ces deux groupes, il faut retenir qu'ils se sont constitués « contre » l'espoir de devenir un professionnel exemplaire, car, bien entendu, il est impossible de devenir un bon enseignant en mathématiques ou en sciences si l'on ne maîtrise pas ces disciplines.

5.3 Analyse des correspondances

La figure 3 de l'Analyse de correspondances (AC) est obtenue aussi avec le logiciel R. Avant de tirer des conclusions de ce graphe, il faut observer que les dimensions 1 et 2 n'expliquent que le 19,4% de variabilité des données : 10,2% et 7,2%, respectivement. Alors, ce graphe « classique » ne donne pas d'information « fiable », car il y a un pourcentage de variabilité beaucoup plus grand (80,6%) qui n'est pas y compris et donc on est censé faire des analyses additionnelles. En fait, on accepte généralement comme un bon ensemble de dimensions celui-ci qui regroupe au moins le 70% de la variabilité. Ici, il faut au moins 14 dimensions ! (figure 4).

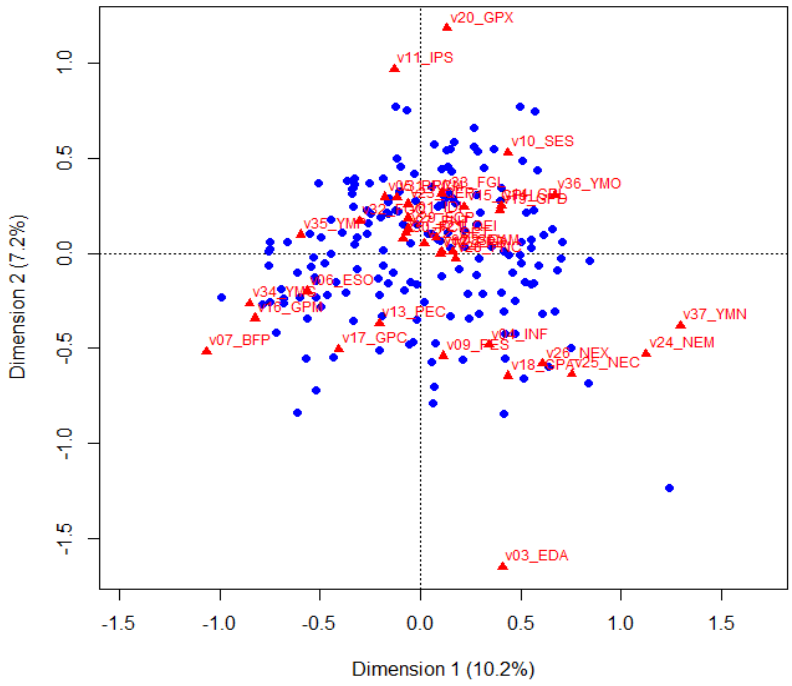


Figure 3 – Graphe de correspondances

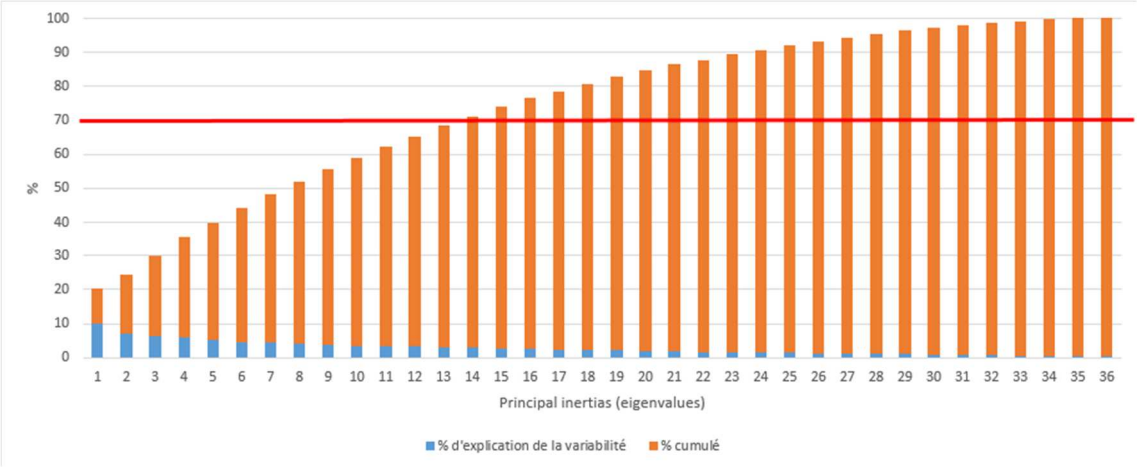


Figure 4 – Explication des dimensions et pourcentage cumulé

On peut dire donc sans risque de se tromper que l'analyse des données par le biais des dimensions de l'AC est la chronique d'un naufrage ou, au mieux, d'une frénésie de chiffres. Lorsqu'une dimension suggère une relation entre des variables, une autre dimension suggère d'autres relations qui ne sont pas cohérentes avec la première. En plus, on ne peut pas proposer un certain classement hiérarchique de ces relations, car la contribution des dimensions successives est graduelle et sans sauts substantiels et, par conséquent, aucune information n'est disponible pour choisir une relation plutôt qu'une autre. Alors, on a besoin d'autres moyens d'analyse.

5.4 Analyse de similarité

La figure 5 montre le graphe de similarité où on peut voir les nœuds significatifs.

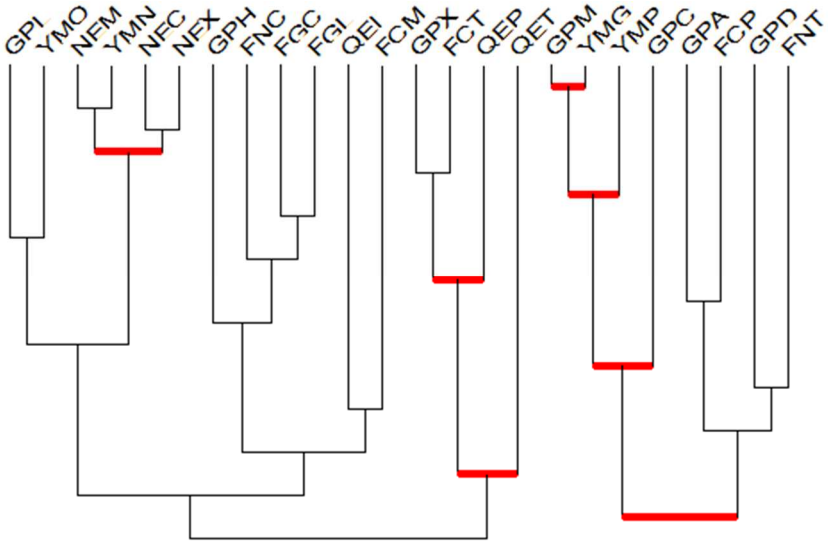


Figure 5 – Arbre des similarités

Avant de « se pencher » sur ce graphe, il faut analyser les indices de similarité. Le tableau 2 montre les niveaux de similarité avec un indice supérieur à 0,7.

Niveau	IS	Classe	VS+	Risque
1	1,0000	(GPM YMG)	BFP	0,0017
2	0,9998	(NEM YMN)	PES	0,0412
3	0,9998	(NEC NEX)	CAM	0,0176
4	0,9988	((NEM YMN) (NEC NEX))	EDA	0,0896
5	0,9851	(GPX FCT)	IDI	0,0763
6	0,9793	((GPM YMG) YMP)	BFP*	0,0763
7	0,9686	(FGC FGL)	ESO	0,1510
8	0,9529	(GPL YMO)	PPD	0,2180
9	0,9356	(FNC (FGC FGL))	IPS	0,1550
10	0,8937	((GPX FCT) QEP)	CAM*	0,0767
11	0,8868	(GPA FCP)	EDA*	0,0672
12	0,8700	(GPH (FNC (FGC FGL)))	IPS*	0,0324
13	0,8234	((GPL YMO) ((NEM YMN) (NEC NEX)))	PES*	0,2010
14	0,7875	(((GPM YMG) YMP) GPC)	BFP*	0,0003
15	0,7064	(GPD FNT)	CAM*	0,1220
16	0,7028	(QEI FCM)	IDI*	0,0122

Légende :
IS : Indice de similarité.
VS+ : Variable supplémentaire qui contribue le plus à la classe correspondante.
Risque : Probabilité de ce que la contribution de la variable supplémentaire soit due au hasard.
*Variable supplémentaire qui contribue à une classe de niveau supérieur.

Tableau 2 – Contribution des variables supplémentaires aux niveaux de similarité d'au moins 0,7.

On peut classer les 8 variables supplémentaires les plus explicatives en deux types: d'une part, celles *inhérentes aux sujets ou propres de choix personnels* ; d'une autre part, celles *extrinsèques aux sujets ou reliées à la formation des parents et, par conséquence, propres du contexte socio-familial* ; ces variables sont :

a) *Variables supplémentaires inhérentes aux sujets* :

- IDI : Langue véhiculaire à l'enseignement (1 : espagnol ; 0 : basque)
- EDA : Âge (1 : plus de 22 ans ; 0 : moins de 22 ans)
- ESO : Voie d'études à la fin de l'École Secondaire en Espagne (15-16ans) (1 : Seconde technico-scientifique ; 0 : Seconde en Sciences sociales)
- BFP : Voie d'études en Première et Terminale (16-18ans) (1 : Technico-scientifique ; 0 : Sciences sociales et humaines)
- CAM : Change de voie de Seconde à Première (1 : oui ; 0 : non)

b) *Variables supplémentaires extrinsèques aux sujets* :

- PES : Au moins un des parents a fait des études universitaires supérieures non-technico-scientifique (1 : oui ; 0 : non)
- IPS : Au moins un parent enseignant (à l'École maternelle, à l'École maternelle, élémentaire, au Collège ou au Lycée) (1 : oui ; 0 : non)
- PPD : Une personne proche au contexte socio-familial est enseignant (1 : oui ; 0 : non)

5.5 Analyse Statistique implicative

La figure 6 montre le graphe implicatif. On peut y voir 5 groupes des variables bien définis :

1) Groupe A – « *Les socio-scientophiles* ». Dans ce groupe A, les variables « goût » pour les disciplines sont reliées à un contexte socio-familial « pour ». Il faut souligner qu'au cœur de ce groupe A est la variable « FGL » (« Dans mon contexte socio-familial, on m'a appris à aimer les lettres »), ce qui relève l'importance de la lecture sur l'adaptation au contexte éducatif et, en fin de compte, l'importance sur le développement cognitif, car les étudiants qui affirment leur goût pour les matières sont finalement les plus performantes.

2) Groupe B – « *Les scientophobes* ». Dans ce groupe B, les variables « non » ou les variables « contre » les disciplines sont reliées. Même si l'on peut voir dans le graphe que la variable « SEX » est aussi présente, il faut être prudente dans l'interprétation de ce résultat car le genre majoritaire est « féminin ».

3) Groupe C – « *Les artistes* ». Femmes un peu plus âgées qui aiment l'art, en tant qu'exécutantes, et qui envisagent leur travail auprès de plus petits à l'École Maternelle.

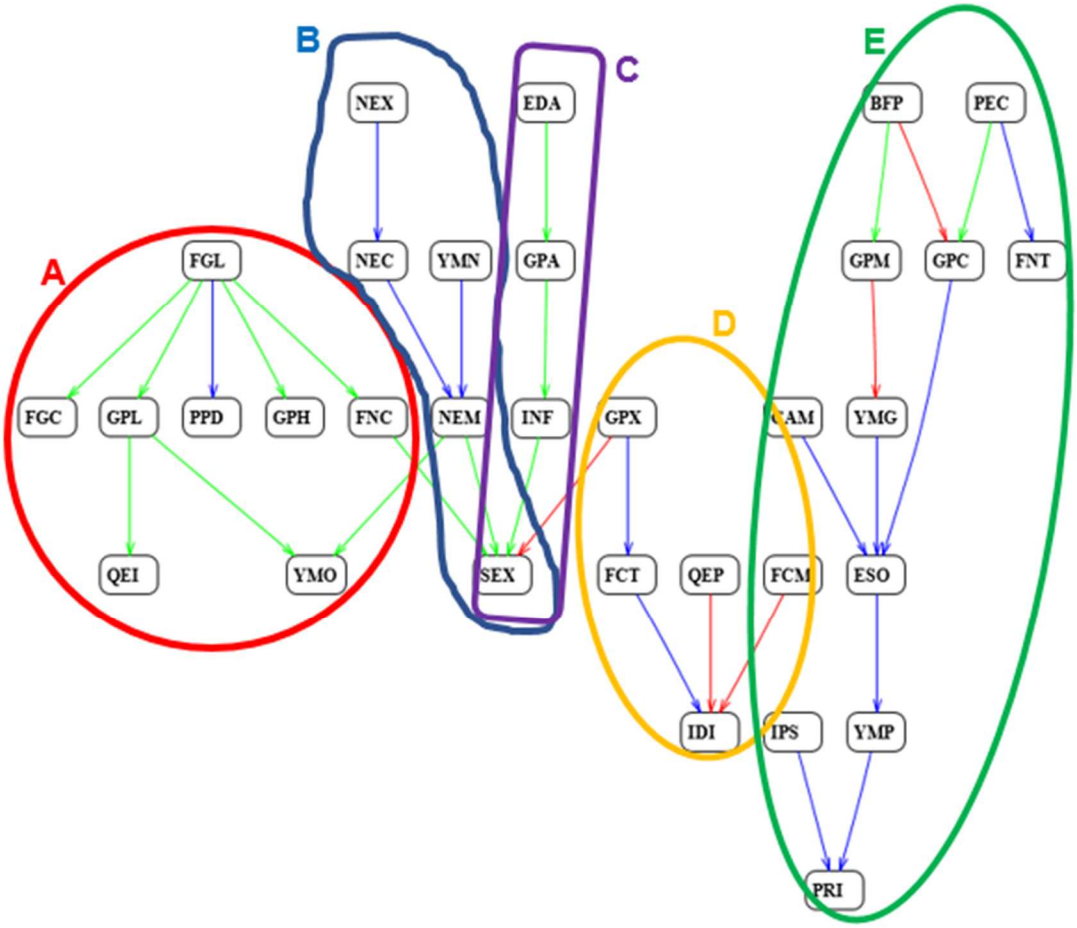


Figure 6 – Graphe implicatif (variables supplémentaires incluses)

4) Groupe D – « *Les travailleurs* ». Dans ce groupe D, les variables « travail » sont toutes reliées. Il faut souligner qu'au cœur de ce groupe D est la variable « GPX » (« Plaisir pré-universitaire pour les langues étrangères »), ce qui relève la compréhension sociale de l'importance des langues étrangères dans l'École ; en effet, maîtriser une langue étrangère est un atout pour commencer à travailler promptement.

5) Groupe E – « *Les scientophiles* ». Dans ce groupe E, les variables « goût » pour les sciences, y compris les mathématiques, sont toutes reliées. Il faut souligner qu'au cœur de ce groupe E sont les variables :

   – « BFP » (« Voie d'études en Première et Terminale (16-18ans) »), ce qui relève l'importance de l'étape pré-universitaire sur, à long cours, sur la professionnalisation des futurs enseignants ; et

   – « PEC » (« Au moins un des parents a une formation technico-scientifique »), ce qui renforce l'importance du contexte socio-familial.

Finalement, aussi important que l'analyse des variables reliées dans le graphe implicatif, est la détermination de celles qui ne sont pas là. Il y a 4 variables absentes :

– INF : Diplôme pour devenir enseignant à l'École Maternelle (0-6ans).

– PRI : Diplôme pour devenir enseignant à l'École Élémentaire (6-12ans).

– PES : Au moins un des parents a fait des études universitaires supérieures non-technico-scientifiques.

– SES : Aucun des deux parents n'a fait d'études universitaires.

Cette absence permet d'affirmer que les groupes A, B, C, D et E se forment indépendamment de :

a) *La Diplôme*. Ce qu'il faut retenir pour la compréhension des groupes sont les caractéristiques communes des étudiants, mais non leurs différences par rapport aux étapes éducatives de référence dans les Diplômes.

b) *La formation universitaire*. Sauf si la formation universitaire est technico-scientifique, elle n'a aucune influence sur le goût et l'intérêt pour les diverses disciplines ou sur la professionnalisation des futurs enseignants.

L'ASI a permis d'obtenir une organisation des variables selon un certain nombre de caractéristiques générales. Les relations ainsi obtenues ne doivent pas être interprétées comme « cause-effet », mais en termes de Théorie des ensembles. Dans l'analyse descriptive et qualitative, cette interprétation ensembliste permet de trouver des relations d'appartenance qui doivent être étudiées a posteriori par rapport à une éventuelle relation de causalité.

L'analyse des implications en préservant leur direction n'a pas de sens ici. On illustre cette affirmation par l'analyse des groupes C et D (tableau 3).

Groupe	Variables
C	EDA : Âge (1 : plus de 22 ans ; 0 : moins de 22 ans) GPA : Plaisir pré-universitaire pour le dessin artistique INF : Diplôme pour devenir enseignant à l'École Maternelle (0-6ans) SEX : Genre (1 : féminin ; 0 : masculin)
D	GPX : Plaisir pré-universitaire pour les langues étrangères FCT : Dans mon contexte socio-familial, je pourrai partager mon expérience de travail QEP : Je veux devenir enseignant car je trouve que ce métier passionnant FCM : Dans mon contexte socio-familial, je recevrai des apports théoriques utiles pour mon activité professionnelle IDI : Langue véhiculaire à l'enseignement (1 : espagnol ; 0 : basque)

Tableau 3 – Variables impliquées dans les groupes C et D.

Toutes les variables du groupe C, sauf le GPA, sont complémentaires. Il n'est donc pas logique d'établir une relation de cause à effet. En fait, l'interprétation des implications découle plutôt d'une analyse purement ensembliste. Par exemple, l'implication « INF -> SEX » signifie simplement que la majorité des étudiants INF sont des femmes, mais l'implication inverse n'est pas vraie car il y a un grand groupe de femmes qui étudient la Diplôme pour devenir enseignant à l'École Élémentaire (6-12ans) (variable « PRI »).

Dans le groupe D, en ce qui concerne la nature des variables impliquées, c'est la situation inverse à celle du groupe C, c'est-à-dire, toutes les variables sont contingentes, à l'exception de la variable IDI. Ici est aussi trop risqué d'établir une interprétation des directions des implications qui va au-delà de l'analyse ensembliste.

En plus, il faut souligner que cette recherche est un étude ex-post-facto, et, donc, non expérimentale et sans manipulation de variables. Ainsi, on ne dispose pas d'informations permettant de situer dans le temps la genèse de l'opinion exprimée dans le questionnaire. En résumé, le résultat principal est donc la constitution des groupes ci-dessus décrits, qui déterminent les « typologies d'étudiants ».

6 En guise de conclusion

Dans les études qualitatives et descriptives, l'analyse des données a besoin d'une triangulation. Dans l'application des différentes méthodes, il est nécessaire de contrôler leur efficacité et, en particulier, de transcender les représentations graphiques et de recourir à l'étude détaillée des valeurs numériques.

Ce travail est de « *validité interne* », c'est-à-dire, on a donné des indicateurs de qualité de la *fiabilité* des résultats obtenus et, par conséquent, de la *certitude* des conclusions que l'on peut tirer de données associées à l'échantillon (*conclusions internes*). Alors, la cohérence des résultats obtenus dans d'autres universités reste ouverte à l'analyse, ce qui permettrait d'envisager une généralisation des résultats à l'ensemble de la population (*validité externe*).

Références

[1] Belletich, O., Villarreal, M. P. de (2014). El imaginario colectivo de la educación en la formación inicial de maestros. En del Valle, M. E.; Padilla, G., *Construyendo la nueva enseñanza superior*, pp. 95-106. Madrid: McGraw Hill.

[59] Brousseau, G. (1998). *Théorie des Situations Didactiques en Mathématiques*. Grenoble : La pensée sauvage.

[60] Caillaud, S., Flick, U. (2016). Triangulation méthodologique. Ou comment penser son plan de recherche. En G. Lo Monaco, S. Delouvée et P. Rateau (Eds.), *Les représentations sociales. Théories, méthodes et applications*, pp. 227-238. De Boeck, Bruxelles. https://urlz.fr/fibS

[61] Chevallard, Y. (1985). La Transposition Didactique. Du savoir savant au savoir enseigné. Grenoble : La pensée sauvage.

[62] Coppé, S., Houdement, C. (2009). Résolution de problèmes à l'École Primaire française : perspectives curriculaire et didactique. *Colloque de la COPIRELEM*. https://urlz.fr/g1Og

[63] Cosnefroy, L. (2007). Les sens multiples de l'intérêt pour une discipline. *Revue française de pédagogie*, *159*, 93-102. https://doi.org/10.4000/rfp.1080

[64] Drapeau, M. (2004). Les critères de scientificité en recherche qualitative. *Pratiques Psychologiques*, *10*(1), 79-86. https://doi.org/10.1016/j.prps.2004.01.004

[65] Lacasta, E., Wilhelmi, M. R. (2008). The graphic illusion of high school students. In R. Gras, E. Suzuki, F. Guillet and F. Spagnolo (Eds.), *Statistical Implicative Analysis. Theory and Applications* (pp. 99-117). Springer. https://urlz.fr/g1O9

[66] Lambert-Le Mener, M. (2012). La performance académique des étudiants en première année universitaire : Influence des capacités cognitives et de la motivation. [Thèse doctorale, Université de Bourgogne]. https://urlz.fr/g1OO

[67] Wilhelmi, M. R., Font, V., Godino, J. D. (2005). Bases empiriques de modèles théoriques en didactique des mathématiques : réflexions sur la théorie de situations didactiques et l'approche ontologique et sémiotique. En A. Rouchier, *Actes du Colloque International « Didactiques : quelles références épistémologiques ? »* (CD-ROM) (pp.1-10). Bordeaux : AFIRSE et IUFM Aquitaine. Version espagnole en : https://urlz.fr/g1Oe

[68] Zlatka, G. (1994). Manifestations de la catégorie du médiatif dans les temps du français. *Langue française*, *102* (Nº thématique : Les sources du savoir et leurs marques linguistiques), 8-23. https://doi.org/10.3406/lfr.1994.5711

Annexe A : questionnaire

(Version originale en espagnol)

Gusto por las disciplinas	***Respuestas múltiples***
¿Cuál o cuáles de las siguientes asignaturas te han gustado más en la formación previa a la universidad? *Puedes elegir hasta tres opciones.*	– *Matemáticas* – *Lengua y Literatura* – *Física* – *Química* – *Historia* – *Biología* – *Tecnología* – *Dibujo Artístico* – *Dibujo Técnico* – *Otras (por favor especifique)*
¿Por qué decidiste estudiar el Grado en Maestro? *Por favor, marca hasta dos opciones.*	– *Me gusta mucho interactuar con los niños y niñas* – *Desde hace tiempo me veía trabajando como maestra o maestro* – *Me parece una carrera apasionante* – *Me parece una carrera más fácil para estudiar* – *Creo que en esta carrera tengo mayores posibilidades de graduarme* – *Otro (por favor especifica tus motivos)*
¿Qué ámbito te gustaría que NO fuera obligatoria en el Grado en Maestro? Por favor elige una o dos opciones	– *Lengua* – *Historia y Geografía* – *Biología y Geología* – *Física y Química* – *Matemáticas* – *Lenguas extranjeras* – *Otra (señala cuál)*
Por favor, indica en cada caso tu punto de vista. (Donde 1 es "Totalmente en desacuerdo"; 2 es "En desacuerdo", 3 es "De acuerdo" y 4 es "Totalmente de acuerdo")	*En los Grados en Maestro NO se debería(n) estudiar:* – *Ciencias Sociales* – *Matemáticas* – *Ciencias Experimentales* – *Lengua y Literatura* – *Lenguas extranjeras*
Entorno socio-familiar	***Respuestas múltiples o abiertas***
Por favor cuéntanos si en tu entorno socio-familiar se da alguna de las siguientes realidades: *Puedes marcar más de una opción.*	– *Al menos mi padre/tutor o mi madre/tutora tiene estudios superiores universitarios* – *Al menos mi padre/tutor o mi madre/tutora tiene estudios técnicos no universitarios* – *Mis padres/tutores no tienen estudios ni superiores ni técnicos* – *Otra circunstancia*
Lo que he aprendido de mi entorno familiar y social (sociedad-escuela-instituto) y que es lo que espero en los estudios de Grado es: *Puede señalar más de una opción.*	– *La novedad de los contenidos* – *La aportación de un punto de vista nuevo sobre cuestiones de mi trabajo* – *La posibilidad de compartir mi trabajo con otros* – *La aportación de un marco teórico a mi actividad profesional* – *La posibilidad de hacer consultas sobre problemas cotidianos*

Por favor dinos si en tu familia se da alguna de las siguientes circunstancias:	– *Me han enseñado a amar la Ciencias* – *Me han enseñado a amar las Letras* – *Se aborrecen las Ciencias* – *Se aborrecen las Letras* – *Ninguna de las anteriores*
¿Cuál de las siguientes afirmaciones es cierta?	– *Mi madre/padre/tutor/a es o ha sido maestra/o en Educación Infantil o Educación Primaria* – *Mi madre/padre/tutor/a es o ha sido maestra/o en Educación Secundaria (ESO, Bachillerato, FP o Escuela de Idiomas)* – *Una persona próxima con la que tengo trato frecuente es docente* – *Otra circunstancia relacionada que quieras hacer notar*
¿Tu padre/madre/tutor/a tiene un título en Ciencias o Tecnología (matemáticas, física, química, biología, ingeniería, etc.), en arquitectura o en ciencias de la salud? Por favor, especifica tu respuesta.	*[Respuesta abierta]*
Si las personas con las que has convivido (padre/madre/tutor/a) NO tienen formación en Ciencias o Tecnología, contesta a la siguiente pregunta: ¿Te han inculcado gusto, curiosidad, interés, etc., por estas disciplinas? Describe brevemente tu experiencia.	*[Respuesta abierta]*

Matemáticas y comentarios finales	***Respuestas múltiples o abiertas***
Respecto a las Matemáticas, ¿Qué podrías afirmar? *Por favor, elige hasta 3 afirmaciones.*	Quiero seguir estudiándolas para prepararme mejor como maestro o maestra No quiero volver a verlas en mi vida Me gustan mucho y creo que son importantes en la formación de los niños. Si tengo que estudiarla en el Grado lo haré, muy a mi pesar. *Otro (por favor especifique otra respuesta)*
¿Tienes algún cometario o sugerencia? *Si es así, por favor, escribe en el cuadro que siguiente*	*[Respuesta abierta]*

Annexe B : variables supplémentaires et contingentes

Nº	Code	Description
1	IDI	Langue véhiculaire à l'enseignement (1 : espagnol ; 0 : basque)
2	SEX	Genre (1 : féminin ; 0 : masculin)
3	EDA	Âge (1 : plus de 22 ans ; 0 : moins de 22 ans)
4	INF	Diplôme pour devenir enseignant à l'École Maternelle (0-6ans) (1 : oui ; 0 : non)
5	PRI	Diplôme pour devenir enseignant à l'École Élémentaire (6-12ans) (1 : oui ; 0 : non)
6	ESO	Voie d'études à la fin de l'École Secondaire en Espagne (15-16ans) (1 : Seconde technico-scientifique ; 0 : Seconde en Sciences sociales)
7	BFP	Voie d'études en Première et Terminale (16-18ans) (1 : Technico-scientifique ; 0 : Sciences sociales et humaines)
8	CAM	Change de voie de Seconde à Première (1 : oui ; 0 : non)
9	PES	Au moins un des parents a fait des études universitaires supérieures non-technico-scientifique (1 : oui ; 0 : non)
10	SES	Aucun des deux parents n'a fait d'études universitaires (1 : oui ; 0 : non)
11	IPS	Au moins un parent enseignant dans une étape pré-universitaire (1 : oui ; 0 : non)
12	PPD	Une personne proche au contexte socio-familial est enseignant (1 : oui ; 0 : non)
13	PEC	Au moins un des parents a une formation technico-scientifique (1 : oui ; 0 : non)

Tableau B.1 – Variables supplémentaires

Nº	Code	Description (1 : oui ; 0 : non)
1	GPL	Plaisir pré-universitaire pour la langue et la littérature
2	GPH	Plaisir pré-universitaire pour l'histoire et l'art
3	GPM	Plaisir pré-universitaire pour les mathématiques
4	GPC	Plaisir pré-universitaire pour la physique, la chimie, la biologie ou la technologie
5	GPA	Plaisir pré-universitaire pour le dessin artistique
6	GPD	Plaisir pré-universitaire pour l'éducation physique et sportive (EPS)
7	GPX	Plaisir pré-universitaire pour les langues étrangères
8	QEI	Je veux devenir enseignant car j'aime interagir avec les enfants
9	QET	Je veux devenir enseignant car j'ai envisagé depuis longtemps ce métier
10	QEP	Je veux devenir enseignant car je trouve que ce métier passionnant
11	NEM	On ne devrait PAS continuer à étudier les mathématiques à l'université
12	NEC	On ne devrait PAS continuer à étudier des Sciences empiriques à l'université
13	NEX	On ne devrait PAS continuer à étudier des Langues étrangères à l'université
14	FNT	Dans mon contexte socio-familial, je recevrai des points de vue de la vie professionnelle
15	FNC	Dans me contexte socio-familial, je pourrai discuter les matières que j'apprendrai
16	FCP	Dans me contexte socio-familial, je pourrai poser des questions sur les matières que je serai censé d'apprendre
17	FCT	Dans mon contexte socio-familial, je pourrai partager mon expérience de travail
18	FCM	Dans mon contexte socio-familial, je recevrai des apports théoriques utiles pour mon activité professionnelle
19	FGC	Dans mon contexte socio-familial, on m'a appris à aimer les sciences
20	FGL	Dans mon contexte socio-familial, on m'a appris à aimer les lettres
21	YMG	J'aime les mathématiques et je pense qu'elles sont importantes
22	YMP	Je veux continuer à étudier les mathématiques pour devenir un(e) bon(ne) enseignant(e)
23	YMO	Si je dois étudier les mathématiques, alors je le ferai, à mon grand regret
24	YMN	Je ne veux plus jamais étudier les mathématiques dans ma vie

Tableau B.2 – Variables contingentes

ANÁLISE ESTATÍSTICA IMPLICATIVA E ANÁLISE DE VARIÂNCIA: ESTUDO ESTATÍSTICO COMPARATIVO SOBRE O DESEMPENHO ESCOLAR EM SALA DE AULA MULTICULTURAL

Rafael Santos De Aquino[1], Ana Maria Dos Anjos Carneiro-Leão[2], Nadja Maria Acioly-Régnier[3], Vladimir Lira Véras Xavier De Andrade[4], Jean-Claude Régnier[5].

ANALYSE STATISTIQUE IMPLICATIVE ET ANALYSE DE VARIANCE : ÉTUDE STATISTIQUE COMPARATIVE SUR LA PERFORMANCE SCOLAIRE DANS UNE CLASSE MULTICULTURELLE

STATISTICAL IMPLICATIVE ANALYSIS AND ANALYSIS OF VARIANCE: COMPARATIVE STATISTICAL STUDY ON SCHOOL PERFORMANCE IN A MULTICULTURAL CLASSROOM

RESUMO

Este trabalho objetiva investigar as influências da cultura dos estudantes no desempenho escolar sob uma metodologia analítica quantitativa mista, composta pela análise de variância (ANOVA) e a análise estatística implicativa (ASI). O estudo se deu pela análise estatística de dados do registro escolar das notas de estudantes urbanos, sertanejos, indígenas e quilombolas do Ensino Médio Integrado ao Técnico em Agropecuária, desde o ano de 2011 até 2018 no Instituto Federal do Sertão Pernambucano na cidade de Salgueiro-PE, Brasil. Os resultados revelaram que há influência da cultura no rendimento escolar dos estudantes, onde os quilombolas compõem o perfil vinculado com os rendimentos de conceito D e consequentemente com a evasão escolar. Quanto a ANOVA e a ASI são instrumentos que podem se complementar enriquecendo a análise dos dados, justamente por serem diferentes. Consideramos que a utilização da ASI em pesquisas na área social, especificamente educacional, pode favorecer os estudos por possibilitar a compreensão do objeto de pesquisa como um fenômeno comportamental, não estático, que está sujeito às mudanças conforme as influências das diversas variáveis que compõem a realidade complexa.

Palavras-chave: *A.S.I., ANOVA, Rendimento Escolar, Multiculturalidade*

RÉSUMÉ

Ce travail vise à étudier les influences de la culture des élèves sur les performances scolaires dans le cadre d'une méthodologie analytique quantitative mixte, composée d'une analyse de variance (ANOVA) et d'une analyse statistique implicative (ASI). L'étude a été réalisée à travers l'analyse statistique des données du dossier scolaire des notes des élèves urbains, ruraux, indigènes et quilombola du Lycée Intégré au Technicien Agricole de 2011 à 2018 à l'Institut Fédéral du Sertão Pernambucano dans la ville de Salgueiro-PE, Brésil. Les résultats ont révélé qu'il existe une influence de la culture sur le rendement scolaire des élèves, où les quilombolas constituent le profil lié au concept D de revenu et, par conséquent, au décrochage scolaire. Quant à l'ANOVA et à l'ASI, ce sont des instruments qui peuvent se compléter,

[1] Instituto Federal do Sertão Pernambucano – IFSertãoPE, Salgueiro, Brasil, rafael.aquino@ifsertao-pe.edu.br.

[2] Universidade Federal Rural de Pernambuco – UFRPE, Recife, Brasil, ana.acleao@gmail.com.

[3] Université Claude Bernard Lyon 1 – INSPE – Lyon, France. EA 4571 – Laboratoire Éducation Cultures Politiques, École Doctorale 485 EPIC, nadja.acioly-regnier@univ-lyon1.fr.

[4] Universidade Federal Rural de Pernambuco – UFRPE, Recife, Brasil, vladimir.andrade@ufrpe.br.

[5] Professeur des universités émérite. Membre du laboratoire UMR 5191 ICAR, Université Lumière Lyon2. Professeur invité à l'Université d'Etat de Tomsk (Russie) – National Research Tomsk State University, jean-claude.regnier@univ-lyon2.fr.

enrichir l'analyse des données, précisément parce qu'ils sont différents. Nous considérons que l'utilisation de l'ASI dans la recherche dans le domaine social, notamment dans l'éducation, peut favoriser les études en permettant la compréhension de l'objet de recherche comme un phénomène comportemental, non statique, qui est sujet à des changements selon les influences des différentes variables qui composent la réalité complexe.

Mots-clés : *A.S.I., ANOVA, Rendement Scolaire, Multiculturalité*

ABSTRACT

This work aims to investigate the influences of students' culture on school performance under a mixed quantitative analytical methodology, consisting of analysis of variance (ANOVA) and Statistical Implicative Analysis (SIA). The study was carried out through the statistical analysis of data from the school record of the grades of urban, rural, indigenous and quilombola students from High School Integrated to the Agricultural Technician from 2011 to 2018 at the Federal Institute of Sertão Pernambucano in the city of Salgueiro-PE, Brazil. The results revealed that there is an influence of culture on the students' academic performance, where the quilombolas make up the profile linked with the concept D income and consequently with school dropout. As for ANOVA and SIA, they are instruments that can complement each other, enriching data analysis, precisely because they are different. We consider that the use of SIA in research in the social area, specifically in education, can favor studies by enabling the understanding of the research object as a behavioral phenomenon, not static, which is subject to changes according to the influences of the various variables that make up the complex reality.

Keywords: *S.I.A., ANOVA, School Achievement, Multiculturality*

1 Introdução

A globalização é um dos principais fatores utilizados para justificar a pesquisa e o desenvolvimento em educação multicultural, especificamente no ensino em cultura-cruzada devido a aproximação de diferentes culturas (Eijck e Roth, 2011; Grimberg e Gummer, 2013; Nadenka e Molina-Andrade, 2015), mas quando se trata de países historicamente colonizados como o Brasil, onde seus povos originários e demais povos subjugados foram escravizados e trazidos para o país quando colônia portuguesa, a educação multicultural se constitui como instituição da justiça social às etnias marginalizadas em todo o processo histórico.

Em um país como o Brasil, é natural que haja grande diversidade cultural, desde as diversas etnias indígenas à diáspora africana que espalhou as diversas culturas daquele continente mundo à fora, bem como a cultura dos diversos povos europeus que aqui se instalaram com prevalência do povo português e a consequente cultura cientificista ocidental. Essa diversidade de povos na formação social do país, fez surgir um povo brasileiro naturalmente sincrético e diverso, mas em nome da hegemonia da cultura eurocêntrica, as culturas tradicionais de indígenas, afro-brasileiros e demais mestiços como cafusos, mamelucos e caboclos que se mantiveram, transformaram e formaram as nossas mais diversas expressões culturais são historicamente marginalizadas.

Por isso, é preciso que os educadores atentem para a promoção de uma educação em cultura-cruzada para atender aos anseios históricos desses povos tradicionais. A consideração ou não da cultura dos estudantes, pelo professor, no processo de ensino e aprendizagem, refletirá também nos índices de desempenho escolar. Essa consideração ou o descaso desse fator não será variável exclusiva sob a influência do rendimento dos

estudantes, mas influenciará diretamente pela valorização do processo de ensino e aprendizagem contextualizado.

Em um trabalho clássico de Carraher e Schliemann (1983, p.3) onde os autores estudaram o fracasso escolar, particularmente em matemática em escolas brasileiras, verificaram que o fracasso na escola é seletivo das camadas pobres da população e requer explicações de ordem social, econômica e cultural. Mas os autores ainda fazem o contraponto de que a análise do rendimento escolar não deve seguir o determinismo mecânico direcionado apenas às ordens econômica, social e cultural.

Situando-se na perspectiva da multiculturalidade brasileira, desdobrada sobre o registro escolar do desempenho estudantil, pretendemos neste trabalho investigar a influência da cultura dos estudantes no desempenho escolar, sob uma metodologia analítica quantitativa mista, composta pela análise de variância (ANOVA) e a análise estatística implicativa (ASI).

1.1 Contextualização

A investigação do desempenho escolar ocorrerá no curso de Ensino Médio Integrado ao Técnico em Agropecuária, do Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano (IF SERTÃO-PE) na cidade de Salgueiro, Pernambuco, Brasil (figura 1).

Figura 1 – Localização de Salgueiro, Pernambuco, Brasil. Fonte: adaptado de Alves et al. (2008)

A comunidade estudantil é constituída multiculturalmente por indígenas das etnias *Atikum* e *Truká*, sertanejos tradicionais (habitantes do semiárido nordestino, que apresentam cultura própria formada pelo sincretismo cultural de indígenas, europeus portugueses e holandeses, judeus e cristãos, que praticam a pecuária caracterizando-se vaqueiros, aboiadores, cordelistas, confeccionam e usam vestimentas típicas de couro), quilombolas (povos habitantes de comunidades intituladas quilombos e que descendem dos povos africanos escravizados no Brasil colonial que constituíram, após fuga das condições escravistas, em comunidades no interior do país, algumas fundadas no século XVII e existentes até hoje mantendo a cultura dos antepassados) e ainda os estudantes urbanos.

1.2 Multiculturalismo

Como nosso trabalho trata das relações multiculturais, trazemos aqui uma breve contextualização sobre cultura e sobre o multiculturalismo. A variedade conceitual sobre o termo cultura foi discutida por Acioly-Régnier (2010, p. 9) quando nos informou que no ano de 1952 já haviam 164 definições diferentes catalogadas para o termo cultura apontando isso como obstáculo aos trabalhos que envolvem a temática cultural.

Para compreensão desta dificuldade conceitual, trazemos conceituações sobre cultura que pode ser definida antropologicamente como as normas, valores, crenças, expectativas e ações convencionais de um grupo (Phelan, Davidson e Cao, 1991), mas que pode apresentar outras definições como os aspectos materiais, sociais, cognitivos e linguísticos da cultura (Leavitt, 1995), ou ainda os aspectos ecológicos, sociais e cognitivos da cultura evidenciados por Stairs (1995).

Para conceituar multiculturalismo, que também é um termo que sofre com a polissemia, nos pautamos na definição de Candau e Moreira (2008) que caracteriza o multiculturalismo como constituído de elementos relacionais, que por influência da relação de poder entre as culturas, se divide em multiculturalismo assimilacionista, diferencialista e interculturalista, de modo que o interculturalismo é o multiculturalismo em sua condição mais desenvolvida com a influência das relações de poder entre as culturas próximas à nulidade, com um pleno relacionamento multicultural, democrático, respeitoso e igualitário.

1.3 Análise estatística implicativa – A.S.I.

A ASI designa um campo teórico centrado no conceito de implicação estatística ou, mais precisamente, no conceito de quase implicação para distingui-la da implicação lógica dos campos da lógica e da matemática (A.S.I. 11, 2021). O raciocínio que apoia a interpretação dos resultados da análise implicativa é essencialmente de natureza estatística e probabilística (Gras e Régnier, 2015); a força de uma associação entre duas variáveis é aferida pela probabilidade P (Nma < nma | independência) – que, no caso de independência, o número de ocorrência (Nma) é mais fraco do que o número efetivamente observado (nma) (Oris, Ritschard e Perroux, 2013).

A ASI se desenvolveu a partir da necessidade de instrumentos estatísticos que pudesse responder a uma questão da área das ciências humanas, mas especificamente da didática das matemáticas. Regis Gras em sua pesquisa de doutorado desenvolvia uma taxonomia de objetos matemáticos. Ele observou que na resolução de problemas de matemática se existia uma certa hierarquia. Na medida que a complexidade do conhecimento em jogo aumentava, o número de acerto diminuía. O que poderia se concluir que quem consegue resolver um problema mais complexo conseguia resolver um problema mais simples sendo a recíproca não verdadeira. Contudo, os contraexemplos aparecem, ainda mais quando estamos lidando com seres humanos. Mas em que medida esses contraexemplos invalidam essa regra? Com o objetivo de responder essa questão ele desenvolve a teoria estática e matemática da Análise Estatística Implicativa (Régnier e Andrade, 2020a). A ASI desde o surgimento até o presente momento vem se ampliando e se desenvolvendo. Trata-se de uma área de conhecimento em constante evolução. E sua aplicação vem se diversificando em diferentes áreas do conhecimento. Possibilitando o seu uso em áreas que antes não eram usadas ou não faziam parte das aplicações da estatística, favorecendo

o surgimento de novos conhecimentos em pesquisas de diferentes naturezas. Em relação ao desenvolvimento da ASI podemos observar que:

> "O desenvolvimento teórico da ASI foi principalmente estimulado por uma dialética entre a prática e a teoria, dentro de uma tensão entre dois quadros: estatística aplicada à alguma coisa e estatística matemática. Dentro de diversos campos científicos, tais como: a didática da matemática, a psicologia, a sociologia, a bioinformática etc., dados observados foram submetidos a esse critério de análise. Essa aplicação mostrou a eficiência do método dentro de sua capacidade a fazer emergir propriedades, que outros métodos não permitiam, mas ela permitiu também mostrar seus limites que suscitaram novas problemáticas em torno do conceito-objeto da quase implicação." (Gras e Régnier, 2015, p. 25).

Essa análise, assim como qualquer outro procedimento estatístico requer a utilização de um software que computacionalmente sistematize e processe os dados produzindo relatórios interpretativos. Para a ASI, utiliza-se o CHIC, cuja última versão para Windows é a 7.0 (2014). Segundo Di Paola e Spagnolo (2009), a sua multiplataforma gratuita é chamada R-CHIC®. Conhecer a história do desenvolvimento da ASI, requer o conhecimento do desenvolvimento do CHIC® que foi executado por uma equipe interdisciplinar de pesquisadores como relatado por Gras e Régnier (2015):

> "A ampliação do desenvolvimento da ASI tem também encontrado apoio dentro da assistência informática que aporta um software dedicado, designado pelo acrônimo CHIC (Classificação Hierárquica, Implicativa e Coesitiva) cuja construção foi iniciada por Régis Gras em 1985, aperfeiçoado em seguida dentro das teses de Saddo Ag Almouloud (1992) e de Harrison Ratsimba-Rajolin (1992). Sua manutenção e atualização são garantidos por Raphaël Couturier (2008)." (Gras e Régnier, 2015, p. 25).

Desta forma, Couturier, Bodin e Gras (2003, p.1) esclarecem que o software CHIC tem como objetivo essencial o de "extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecendo um índice de qualidade de associação e de representação de uma estruturação das variáveis obtidas por meio dessas regras". No CHIC podemos trabalhar com diferentes tipos de variáveis: binárias, modais, quantitativas ou efetivas e variáveis intervalo (Régnier e Andrade, 2020b). Em nossa pesquisa utilizamos as variáveis binárias. As variáveis binárias, que compreendem unicamente dois valores, 0 e 1 representando, respectivamente, a ausência e a presença, o verdadeiro e o falso, o sim e o não, etc. (Gras e Almouloud, 2002; Freitas et al. 2019). Os dados binários são dispostos em planilha do Microsoft Excel do tipo ".csv" ou "Separado por vírgula". Di Paola e Spagnolo (2009) destacam que as funções usuais do CHIC® são: árvore de similaridade, gráfico implicativo, árvore Coesitiva e redução; e uma das funções complementares é a entropia, aplicada quando se está usando um grande número de dados.

Gras e Régnier (2015) esclarecem que através do CHIC os dados construídos em uma planilha podem ser processados determinando os valores de intensidade de implicação. Os maiores valores vão indicar o sentido e as relações entre as variáveis que possibilita o software CHIC gerar o gráfico implicativo onde cada seta determina o sentido da

implicação (variável A → variável B) e a cor do vetor estará relacionada a um determinado nível de intensidade (figura 2).

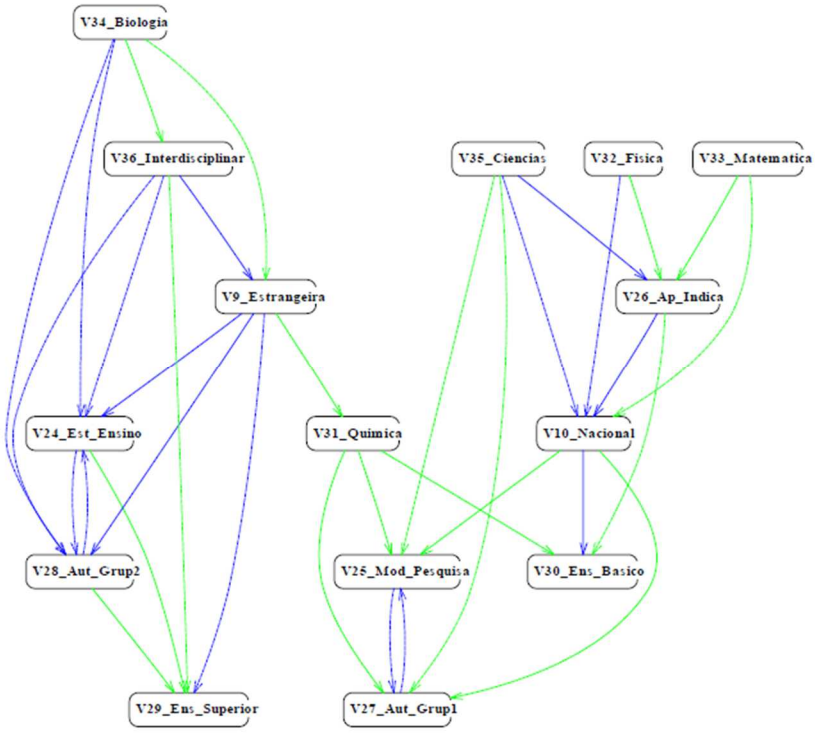


Figura 2 – Exemplo de gráfico implicativo. Fonte: Freitas et al. (2019).

No gráfico implicativo cada seta representa uma relação de implicação, onde a cor indica uma intensidade de implicação, que quanto mais próxima de 1,00 (0,99) será mais forte. No exemplo, podemos ver que a variável 31 - Química implica nas variáveis 25, 27 e 30, e a variável 9 implica na variável 31 – Química. Essas relações são importantes para revelar ao pesquisador como se comportam as variáveis sobre as outras confirmando ou negando hipóteses, realizando novas inferências, complementando o estudo qualitativo.

A teoria estatística e a ASI estão em contínua expansão e desenvolvimento, como apontam Di Paola e Spagnolo (2009), pois são diversas as áreas científicas de aplicação da ASI. Podemos citar Pazmiño-Maji, García-Peñalvo e Conde-González (2017) que ao estudarem e relacionarem as publicações sobre ASI com outros métodos estatísticos (KDD – *Knowledge Discovery in Databases and Data-Mining*) verificaram que a literatura sobre ASI é majoritariamente produzida na França correspondendo a 40% das publicações e que os estudos mais significativos foram nas áreas de Ciências da Computação (32%), Ciências Sociais (24%), Matemática (23,6%) e Decisões Científicas (6%) responsáveis por aproximadamente 80% das publicações sobre ASI.

Na temática que envolve a educação intercultural, tema central desta proposta de pesquisa, existem diversas utilizações da ASI como ferramenta de análise quantitativa em proposições metodológicas em pesquisas mista (quali-quantitativo). Por exemplo, podemos citar a pesquisa multicultural na sala de aula constituída por chineses e italianos onde a ASI contribuiu para integrar variáveis, em classes, avaliando a forma de pensar, o comportamento e as estratégias dos estudantes chineses e italianos no processo de aprendizagem da matemática relacionados às suas respectivas culturas (Di Paola e

Spagnolo, 2009). Ainda na mesma linha da interculturalidade apontamos o trabalho de Pérez-Caraballo, Acioly-Régnier e Régnier (2012) que em estudo na área de Psicologia, com indivíduos de populações internacionais da fronteira Brasil e Uruguai perceberam que a Análise Estatística Implicativa possibilitou a identificação da prevalência de identidades nacionais em detrimento das identidades de fronteira. Em outro estudo, Acioly-Régnier e Régnier (2005) citaram a influência da ASI na percepção do conhecimento cultural extraescolar e escolar, ocorrendo ao mesmo tempo no processo de interação de diferentes sistemas simbólicos revelando contrastes entre os conceitos cotidianos e científicos.

Na área de formação de conceitos e constituição paradigmática, também consideradas por nossa proposta de pesquisa, podemos citar Fazio e Spagnolo (2008) que ao investigarem as linhas filósofo-metodológicas (construtivismo, aristotélico-galileu, platonista, etc.) de estudantes de graduação do curso de Professores da Educação Secundária da Universidade de Palermo, Itália, que desejavam atuar no ensino da matemática e da física. Eles perceberam que o construtivismo presente na universidade e no ensino secundário influencia o perfil de futuros professores. Gras e Larher (1994) citam estudo de Londeix, em que foi reconstruído, a partir de um teste, uma hierarquia de estágios segundo Piaget que mostra uma defasagem diferencial devido aos contextos dos exercícios do teste, verificando hipóteses psicológicas que influenciam a formação de conceitos.

Noutra temática de nosso interesse, destacamos o estudo exploratório da formação humana nos trabalhos dos Colóquios Internacionais sobre ASI, realizado por Cunha, Ferreira e Acioly-Régnier (2018) que apontaram as áreas de formação de professores, formação profissional e formação humana como as temáticas mais presentes nas pesquisas. Todos esses exemplos da aplicabilidade da análise estatística implicativa só reforçam a importância dessa metodologia para a pesquisa em Ensino de Ciências e à elas ajuntamos tal estudo da aplicação estatística mista na análise do desempenho escolar sob a ótica da multiculturalidade.

1.4 Análise de variância - ANOVA

A análise de variância é, provavelmente, o método estatístico de maior repercussão na pesquisa científica, especialmente na experimentação agrícola (Memória, 2004). Hoje é aplicada em diversas áreas incluindo a educação (Plonsky e Oswald, 2017, Louro, 2018, Zeynali et al., 2019) permite explorar as relações entre fatores controlados em um experimento e uma única resposta. Em particular, pode separar formalmente a variabilidade nas respostas entre as diferentes amostras.

A ANOVA pode contribuir para um padrão característico quando todas as respostas são consideradas simultaneamente, o que é chamado de vantagem multivariada. Isso enriquece a análise do objeto de pesquisa em vez de trabalhar com cada produto ou variável separadamente (Bertinetto et a., 2020).

Essa metodologia estatística é muito difundida no meio científico em diversas áreas, por permitir análises de objetos de estudo sob uma perspectiva multivariada, onde diversas variáveis são consideradas conjuntamente, afim de conhecer efeitos entre elas.

A utilização da ANOVA, na prática, pode limitar a nossa interpretação e contextualização dos resultados quanto as abstrações relacionadas aos problemas reais do cotidiano associados ao objeto de pesquisa, principalmente nas ciências humanas. Isso se deve à lógica quantitativa determinista cartesiana que estrutura a ANOVA. Sobre essa interpretação, citamos Plonsky e Oswald (2017) que aplicaram a análise de variância para estudo do desempenho escolar e apontaram para uma possível dificuldade em nosso potencial de raciocínio empírico para informar e promover a teoria e a realidade prática, mais especificamente, em vez disso nos concentramos quase que exclusivamente na diferença entre as médias.

2 Materiais e métodos

Descrevemos aqui a metodologia quanto ao local em que se deu a experimentação e informações sobre a pesquisa, a caracterização amostral, a organização e tratamento dos dados na análise estatística implicativa e ANOVA.

2.1 Local

A pesquisa foi desenvolvida no Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, IFSertãoPE, na cidade de Salgueiro, Pernambuco, Brasil (Figura 1). A pesquisa é parte dos estudos de tese de doutorado em cotutela em Ensino das Ciências da Universidade Federal Rural de Pernambuco (UFRPE) no Brasil e do doutorado em *Science de l'Éducation* da Université Lumière Lyon 2 em Lyon, França. O projeto foi registrado e aprovado pelo sistema CEP/CONEP[6] sob o número CAEE: 32536620.2.0000.9547.

2.2 Caracterização amostral

Foram coletadas informações de notas escolares por disciplina, de estudantes do Ensino Médio Integrado ao Técnico em Agropecuária, de todas as séries (1º ao 4º ano) referentes aos anos letivos de 2011 até 2018, totalizando 232 estudantes e 517 perfis (um mesmo estudante do 1º ano do Ensino Médio em 2011 tem um segundo perfil referente ao 2º ano do Ensino Médio e assim sucessivamente).

Os estudantes foram agrupados conforme a cultura ou etnia caracterizadas por quatro grupos distintos: o sertanejo tradicional, o quilombola, os indígenas e urbanos. Estas representam as variáveis independentes do estudo, enquanto que as disciplinas e grupos disciplinares configuram as variáveis dependentes.

2.3 Organização e tratamento de dados para A.S.I.

Para tratamento dos dados aplicamos a base teórica da Análise Estatística Implicativa – A.S.I. (Gras, Régnier, Guillet, 2009; Gras, Régnier, Marinica, Guillet, 2013; Gras, Régnier, Lahanier-Reuter, Marinica, Guillet, 2017; Gras, 2018) em que destacamos a

[6] O sistema CEP/CONEP é o responsável por avaliar e deliberar sobre aprovação ou rejeição dos projetos de pesquisa quanto às exigências éticas na pesquisa com seres humanos no Brasil.

originalidade de estudar as relações não simétricas entre as variáveis ou entre grupos de variáveis.

Para analisar os dados de rendimento escolar, construímos as variáveis conforme o desempenho médio em níveis de performance, distribuídos em conceitos avaliativos de A à D (A, 8,0 a 10,00; B, de 6,0 à 7,9; C, de 3,0 à 5,9; e D, de 0,0 à 2,9 pontos). As disciplinas foram reagrupadas em domínios comuns afim de preservar a identidade dos professores (Tabela 1).

Grupos	Subgrupos (nº. disciplinas)	Disciplinas
Propedêutica (13 disciplinas)	Ciências Humanas (4)	História, Geografia, Filosofia e Sociologia
	Ciências Naturais (3)	Biologia, Física et Química
	Linguagem e Matemática (4)	Português, Espanhol, Inglês e Matemática
	Educação Física e Arte (2)	Educação Física e Arte
Técnica (24 disciplinas)	Zootecnia (7)	Introdução à Zootecnia, Apicultura, Produção de Animais Não-Ruminantes, Produção de Animais Ruminantes, Forragicultura, Sanidade e Higiene Animal, Alimentos e Alimentação
	Agronomia (8)	Introdução à Agricultura, Fitossanidade, Grandes Culturas, Horticultura, Solos e Fertilidade, Irrigação e Drenagem, Desenvolvimento Vegetal, Conservação de Solos
	Tecnologia Agrícola (4)	Mecanização Agrícola, Desenho e Topografia, Tecnologia de Produtos de Origem Agropecuária e Construções e Instalações Rurais
	Gestão Agrícola (5)	Gestão Ambiental, Administração e Agronegócio, Associativismo e Cooperativismo, Informática Aplicada à Agricultura e Segurança do Trabalho

Tabela 1: Organização analítica dos sujeitos em grupos e subgrupos. Fonte: autoria própria.

Para compreensão dos dados foram produzidos grafos implicativos e a árvore coesitiva. Os grafos implicativos foram gerados sob o índice de implicação superior a 0.80, já na árvore coesitiva, demos destaque ao índice de coesão superior a 0.72. O que influencia a opção por determinado índice implicativo ou índice de coesão são as relações entre as variáveis que desejamos enfatizar para compreender e explicar o objeto de estudo, por exemplo, para os grafos implicativos índices inferiores a 0.80 não expressaram relações importantes, para a árvore coesitiva, a consideração de índices acima de 0.72 eliminam nós importantes entre as variáveis limitando a análise das relações implicativas.

Os dados obtidos foram inseridos em uma planilha do Microsoft Excel (2010) formato ".csv" (separado por vírgulas) com identificação binária dos conceitos por disciplinas de acordo com cada perfil discente. Para o tratamento dos dados foi utilizado o software de

Classificação Hierárquica Implicativa e Coesitiva - CHIC®, v. 7.0, (2014), (Couturier, 2000). Foi escolhido a opção "nós significativos", foi empregada a implicação segundo a teoria clássica e a Lei binomial. Em sua versão 7.0 temos como opções a lei Binomial e a Lei de Poisson. Na escolha dessas leis deve-se considerar que:

> "[...] a modelagem probabilística da variável aleatória pode ser uma distribuição binomial ou uma de Poisson. Na modelagem com a distribuição de Poisson, supõe-se que o universo de estudo seja infinito numerável, senão tão grande (muitos sujeitos) que ele pode ser considerado como tal, enquanto na modelagem com a distribuição binomial o universo de estudo é finito (Régnier e Andrade, 2020b, p. 62).

Em função da amostra pequena, considerou-se como mais adequado o tratamento da modelagem probabilística da variável aleatória utilizando uma distribuição binomial.

2.4 Organização e tratamento de dados para ANOVA

Para o tratamento estatístico aplicado à metodologia de análise de variância (ANOVA) foram computados os valores exatos das notas de cada estudante em cada disciplina, subgrupos e grupos disciplinares, por série, conforme os grupos culturais (Sertanejo, Quilombola, Indígena e Urbano).

Os dados foram organizados em planilha do Microsoft Excel 2010® onde os valores numéricos foram registrados com duas casas decimais. A planilha foi inserida e tratada no software MAXQDA, 2020 (VERBI Software, 2019).

3 Resultados e discussão

Para a compreensão da análise estatística implicativa utilizamos grafos implicativos gerados através do CHIC v. 7.0 (2014) onde optou-se por cálculos com implicação segundo a teoria clássica e a Lei Binomial para focar nas relações implicativas associadas às culturas dos estudantes. Assim obtivemos três grafos implicativos que são apresentados na figura 3 e nos revelam que o baixo desempenho escolar, associado ao conceito D, está ligado aos estudantes quilombolas, seguido dos indígenas com o conceito C. Os urbanos que apresentam implicações com o desempenho escolar B e os sertanejos (identificados pela variável RURAL) com o conceito escolar A.

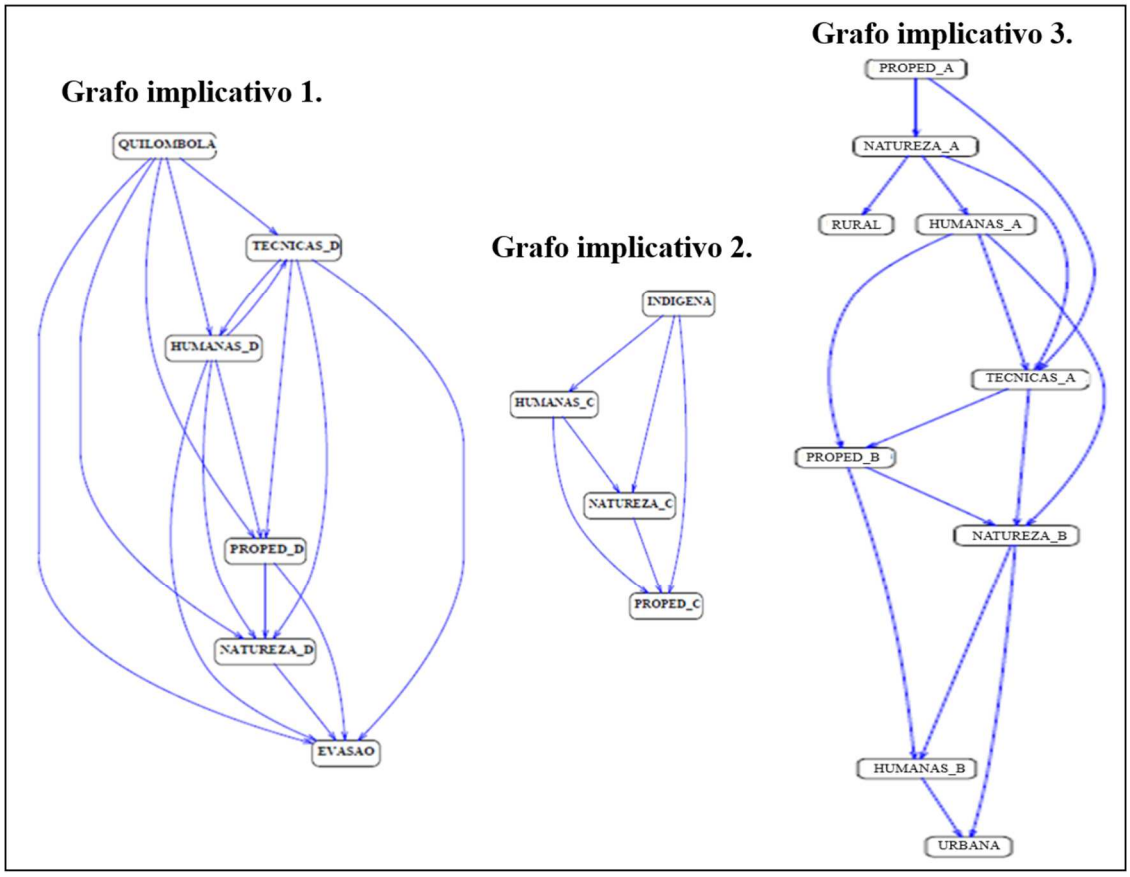


Figura 3: série de grafos implicativos ao nível superior à 0.80 de índice de implicação, onde o grafo 1 apresenta as relações com os estudantes quilombolas, o grafo 2, as relações da performance escolar dos indígenas e o grafo 3 as relações dos estudantes urbanos e sertanejos. Fonte: grafos implicativos obtidos através do software CHIC v. 7.0 (2014) e organizado sob a forma de ilustração no Microsoft Paint 3D®.

A figura 3 nos mostra uma série de grafos implicativos, onde o grafo 1 apresenta as relações do perfil cultural quilombola (QUILOMBOLA) com as performances escolares mais baixas, representadas pelo conceito D nas disciplinas de domínio técnico (TECNICA_D) e propedêutica (PROPED_D) em particular as ciências humanas (HUMANAS_D) e naturais (NATUREZA_D). Ser um estudante quilombola também implica em abandono escolar (EVASAO). O grafo 2 indica que os estudantes indígenas estão relacionados com conceitos do nível C, com implicação sobre as disciplinas propedêuticas (PROPED_C), as disciplinas de humanidades (HUMANAS_C) e das ciências naturais (NATUREZA_C). O grafo implicativo 3 revela a relação dos conceitos A com os estudantes sertanejos (RURAL) por implicação indireta do conceito A em ciências naturais (NATUREZA_A) sobre os estudantes sertanejos e diretamente do conceito A em ciências humanas (HUMANAS_A). O grafo 3 revela ainda que a performance escolar de conceito B nas disciplinas de ciências naturais (NATUREZA_B) e humanas (HUMANAS_B) implicam em estudantes urbanos (URBANA).

Na prática, podemos afirmar com base nos resultados, que os dados de desempenho escolar analisados a partir da etnicidade e cultura dos estudantes confirmam que há influência das questões socioculturais, inegáveis por revelarem segregação no ensino médio integrado ao técnico em agropecuária por resultados de baixo desempenho, que

contribuem, por exemplo, com a relação dos estudantes quilombolas serem o grupo étnico-cultural que mais evadem da escola.

A figura 4, nos apresenta uma relação sistemática entre as classes de variáveis a partir da árvore coesitiva que relaciona as variáveis em relações de implicação, onde podemos compreender as características das comunidades estudantis por grupos étnico-culturais. Priorizamos na figura os nós das classes de variáveis acima do índice de coesão de 0.72.

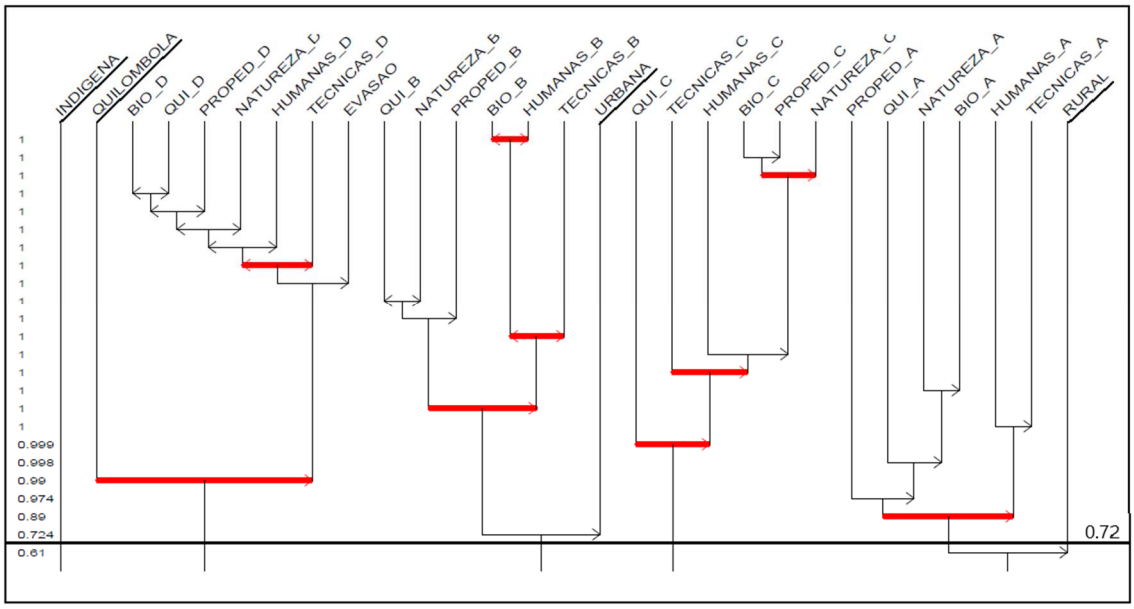


Figura 4: Árvore coesitiva das variáveis étnico-culturais (sublinhadas) e o desempenho escolar classificado por conceitos de A até D. Fonte: Software CHIC v. 7.0 (2014).

A árvore coesitiva nos mostra quatro classes onde destacamos (da esquerda para a direita) a classe do desempenho escolar D, que apresenta forte implicação dos estudantes quilombolas (QUILOMBOLA) ao nível de coesão de 0.99; a segunda classe que apresenta coesão do desempenho escolar de conceito B sobre estudantes urbanos; a terceira classe relacionam os conceitos C entre si sem haver coesão com variáveis étnico-culturais acima do índice de coesão de 0.72; e a quarta classe mostra as relações de coesão dos conceitos A que implicam em estudantes rurais (RURAL), que representam os sertanejos, mas a coesão com a cultura sertaneja que ocorre ao nível de coesão de 0.61.

Chamamos a atenção ao fato de que o desempenho nas disciplinas técnicas para os conceitos A, B e D são influenciados pelos desempenhos nas disciplinas da área propedêutica em que as classes apresentam índice de coesão acima de 0.72 com prevalência de índices acima de 0.90. Para desempenho escolar de conceito C observamos uma relação inversa, onde o desempenho escolar em disciplinas técnicas é que influenciam o desempenho escolar no grupo de disciplinas propedêuticas. Essas relações devem ser utilizadas para um planejamento didático escolar inter ou transdisciplinar dada a influência dos grupos disciplinares de ordem técnica e propedêutica entre si, significando dizer que as disciplinas devem ser tratadas na escola de maneira integrada e holística. Por outro lado, como os desempenhos estão relacionados aos grupos étnicos-culturais a compreensão do sentido das influências de um grupo disciplinar sobre outros pode favorecer um planejamento didático escolar pautado na decolonialidade e na educação multicultural por considerar as diferenças étnico-culturais na sala de aula

incorporando uma política escolar no âmbito didático-pedagógica de inclusão social efetiva e permanente o que favoreceria a aprendizagem mais próxima da equidade combatendo a exclusão social intraescolar.

A árvore coesitiva reforça a heterogeneidade entre as variáveis já apresentadas pelos grafos implicativos, com destaque aos resultados dos estudantes quilombolas associados ao baixo rendimento que afetará o processo de escolarização sob um olhar pautado na inclusão e na integração intercultural.

Como proposta de comparação entre os métodos estatísticos A.S.I. e ANOVA para a compreensão das relações entre o desempenho escolar e os perfis étnico-culturais a tabela 2 apresenta o rendimento escolar entre os grupos culturais e os grupos disciplinares. Informamos que o valor mínimo não foi expresso na tabela por corresponder à nota zero (0,0) em todas as disciplinas, havendo variação apenas sobre o valor máximo entre as médias escolares.

Natureza Disciplinar	Grupo étnico-cultural	Média	Desvio Padrão	Erro Padrão	Valor Máximo
Propedêutica (P<0,003)	Urbano	5,45	±2,41	0,13	8,80
	Sertanejo	5,60	±2,39	0,18	9,10
	Quilombola	0,54	±0,51	0,23	1,40
	Indígena	4,50	±2,45	0,36	7,90
	Geral	5,36	±2,46	0,10	9,10
Ciências Naturais (P<0,001)	Urbano	4,48	±2,96	0,17	9,00
	Sertanejo	4,65	±3,03	0,23	9,60
	Quilombola	0,39	±0,58	0,26	1,30
	Indígena	3,75	±2,47	0,36	7,90
	Geral	4,43	±2,96	0,13	9,60
Linguagens (P=0,0051)	Urbano	5,21	±2,52	0,14	8,70
	Sertanejo	5,29	±2,56	0,20	8,90
	Quilombola	0,60	±0,84	0,37	1,80
	Indígena	4,41	±2,66	0,39	7,80
	Geral	5,12	±2,58	0,11	8,90
Técnicas (P<0,005)	Urbano	6,19	±2,64	0,15	9,50
	Sertanejo	6,27	±2,67	0,21	9,90
	Quilombola	0,36	±0,70	0,31	1,60
	Indígena	5,55	±3,02	0,44	8,80
	Geral	6,10	±2,73	0,12	9,90

Tabela 2: ANOVA do rendimento escolar por grupo étnico-cultural. Fonte: autoria própria.

Os dados da tabela 2 revelam que as médias do desempenho escolar dos estudantes quilombolas foram as que menos variaram, com valores de desvio padrão abaixo de ±0,61, o que reforça a prevalência de notas baixas para este grupo. Os sertanejos apresentam as maiores médias dentre os grupos étnico-culturais. O erro padrão, de maneira geral, esteve abaixo de 0,50. Os índices de significância dos grupos disciplinares apresentaram apenas um índice acima de 0,05 (P=0,0051) para o grupo de linguagens, as disciplinas técnicas abaixo de 5% (P<0,005), o desempenho escolar para as disciplinas

do domínio propedêutico abaixo de 3% (P<0,003) e o desempenho nas disciplinas de ciências naturais abaixo de 1% (P<0,001).

A comparação do rendimento escolar nas disciplinas propedêuticas, técnicas e o rendimento escolar geral por grupo étnico-cultural é apresentada na figura 5 através de uma série de gráficos de barra.

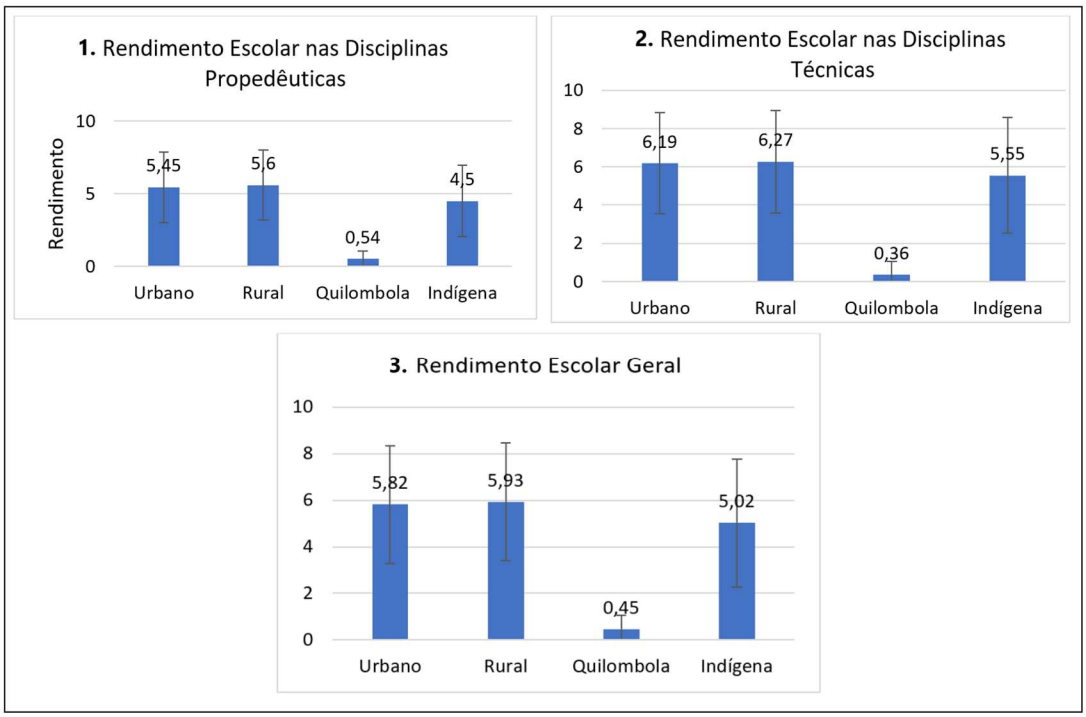


Figura 5: Rendimento escolar médio por grupo cultural nas disciplinas propedêuticas e técnicas além do rendimento geral. Fonte: gráficos produzidos através do Microsoft Excel® (2010) e organizados como única ilustração através do Microsoft Paint 3D®.

Os gráficos apresentados na figura 3 ilustram os resultados já evidenciados, facilitando uma análise comparativa do desempenho escolar médio dos estudantes de culturas diferentes no curso de Ensino Médio Integrado ao Técnico em Agropecuária. Os estudantes urbanos, sertanejos (RURAL), quilombolas e indígenas apresentam desempenho médios significativamente diferentes entre si (disciplinas do grupo de Linguagens P<0,002; disciplinas propedêuticas e técnicas P<0,001; e as disciplinas de ciências naturais P<0,005), mas são os estudantes quilombolas que apresentam a maior variação média no desempenho escolar para os grupos disciplinares investigados comparados com os demais grupos étnico-culturais.

Silva et al. (2016) ao estudar o desempenho de estudantes do Ensino Médio da cidade de Ouricuri-PE, Brasil, especificamente na disciplina de Química, verificou diferenças estatísticas significativas entre os tipos de escolas (públicas estaduais e federal e escola particular) e também entre escolas urbanas e rurais e atribuiu as diferenças de desempenho a diversos fatores, dentre eles a falta de formação do professor de Química, o nível educacional e a quantidade de disciplinas ofertadas na instituição federal na modalidade de Ensino Médio Integrado, a ação familiar determinante nos resultados das escolas particulares e a falta de contextualização no âmbito rural. Na prática, Silva revelou que os diversos grupos subculturais verificados influenciaram o desempenho dos alunos.

Silva et al. (2016) também verificou a tendência ao baixo desempenho médio dos alunos de escolas rurais quando comparados com a média daqueles das escolas urbanas de Ouricuri-PE. Corbett (2015) ao se debruçar sobre a educação rural na Austrália e compará-la com a de outros países como Canadá e Estados Unidos verificou que no geral, a juventude rural teve um desempenho inferior ao dos jovens urbanos e a este resultado defendeu a estigmatização do meio rural como atrasado e não moderno (atributos inversos aos ambiente urbano) que permite que políticas públicas não sejam elaboradas ao ambiente rural à mesma importância aquela conferida às grandes cidades respeitando as peculiaridades do campo que dificultam inclusive a sua própria definição e identidade.

O estudo da diferença do desempenho escolar de estudantes de culturas distintas através dos métodos estatísticos ANOVA, mais tradicionalmente utilizado, e a Análise Estatística Implicativa (A.S.I.) se revelam como metodologias quantitativas que podem ser utilizadas simultaneamente por apresentarem características analíticas complementares, que enriquecem a análise dos dados mesmo que estejam na mesma perspectiva de análise quantitativa. Santos et al. (2017) mostram que a pesquisa de métodos mistos possibilita a realização de estudos com desenhos metodológicos diferentes, mas que produzem dados mutuamente complementares, favorecendo uma interpretação mais completa e abrangente do fenômeno investigado.

Essa integração metodológica contribui com uma leitura mais próxima da realidade por trazer uma perspectiva complexa às análises de dados quantitativos referentes ao desempenho escolar. Podemos, nesse contexto, refletir não apenas sobre as vantagens, mas também sobre as limitações da integração metodológica:

> "Muitas vezes os desenhos experimentais não apresentam os mecanismos subjacentes que explicam como as variáveis independentes influenciam a variável dependente. Muitas vezes não é possível saber em que medida o caso escolhido representa um *outlier* ou uma observação típica de uma determinada distribuição. Em geral, pesquisas com N grande (*large N studies*) não apresentam informações suficientes para entender o contexto específico das unidades observadas, assim como é muito difícil garantir a replicabilidade em entrevistas e observação participante. É exatamente contra essas limitações que repousa o ganho analítico da abordagem multimétodo." (Paranhos et al., 2016, p. 406).

Para fundamentar essa interpretação de complementariedade entre os métodos quantitativos aqui apresentados, identificamos na tabela 3, as características entre a ANOVA e a A.S.I. quanto ao objeto de pesquisa proposto nesse trabalho.

Características	Análise de Variância – ANOVA	Análise Estatística Implicativa – A.S.I.
Natureza dos dados	Quantitativo Absolutista	Quantitativo Probabilístico
Interpretação do resultado	Determinista	Possibilista
Confiabilidade	Significância (P-valor)	Tendência (índice implicativo)
Referência espacial	Posição	Percurso
Natureza descritiva	Estática (retrato)	Cinética (filmográfico)

Tabela 3: Características dos métodos estatísticos ANOVA e ASI quanto ao estudo do desempenho escolar de estudantes de culturas diferentes. Fonte: autoria própria.

A tabela 3 nos apresenta as características da ANOVA e da ASI quanto à natureza dos dados, à interpretação do resultado, à confiabilidade, à referência espacial e à natureza

descritiva. O processamento dos dados pela ANOVA revela-se de natureza absolutista, de modo que os valores encontrados são interpretados como sendo a exatidão da realidade posta, enquanto que na ASI a interpretação estatística é de natureza probabilística o que provê maleabilidade interpretativa possibilitando a consideração da influência de outras variáveis, inclusive não consideradas no estudo inicial. Dessa maneira a interpretação dos resultados na ANOVA apresenta uma característica determinista conforme a base científica positivista, já com a ASI a interpretação se revela possibilista, onde se valoriza as mudanças impostas por variáveis outras, presentes na realidade que envolve o objeto estudado, e que impõe possíveis mudanças detectáveis com o tempo, com o local, e com outras condições importantes à uma interpretação no campo das ciências sociais e educação.

A confiabilidade na ANOVA é conferida pelo valor estatístico da significância, principalmente, que também é interpretada sob forte rigidez do paradigma cartesiano, com a ASI os resultados apresentam a confiabilidade no índice implicativo que reforça uma tendência. Na interpretação do objeto de pesquisa enquanto referência espacial para compreensão de um determinado segmento realístico, os resultados pela ANOVA posicionam o objeto de estudo em seu respectivo espaço, possibilitando as comparações do objeto em diferentes espaços, já a ASI proporciona ao objeto de estudo ser identificado como percurso ou movimento possível. Essa condição proporciona a diferenciação da natureza descritiva dos dados de uma ANOVA e da ASI, a primeira apresenta uma leitura estática do objeto, que mais se relaciona analogicamente, com um retrato, enquanto que a ASI possibilita uma descrição cinética do objeto, mais próxima de uma filmografia.

Assim, podemos inferir que a ASI permite uma compreensão plástica, fenomenológica do objeto de pesquisa, que aponta a tendência de um comportamento. Permitindo a compreensão não apenas do momento descrito, mas contribuindo com a compreensão do movimento e possíveis deduções desse comportamento social no futuro, bem como a possibilidade de lançar mão de planejamentos necessários para a correção dos problemas. Por exemplo, na figura 6, é possível compreender além da caracterização do rendimento escolar em si, quais são as relações do rendimento escolar de um grupo disciplinar com outros fatores.

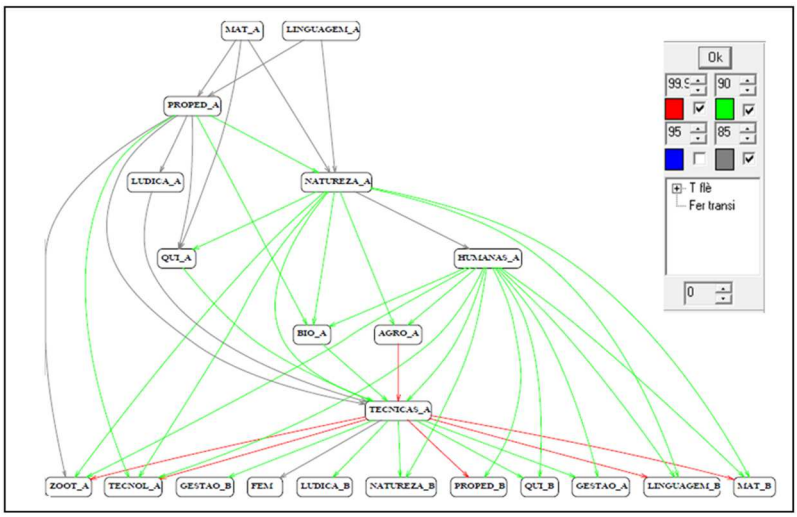


Figura 6: Grafo implicativo sobre as variáveis de conceito A. Os vetores de cores diferentes correspondem a distintos índices implicativos, os vermelhos correspondem a 0.99, os verdes a 0.90 e os cinzas a 0.85. Fonte: Software CHIC v. 7.0 (2014).

A figura 6 nos permite entender, por exemplo, que o ótimo desempenho escolar, caracterizado pelo conceito A em matemática (MAT_A) e em linguagens (LINGUAGEM_A) implica em ótimo desempenho nas disciplinas propedêuticas (PROP_A) e nas disciplinas da área de ciências da natureza (NATUREZA_A). O que isso nos diz além da descrição das relações entre as variáveis? É que, se a escola investir no aprendizado em Matemática e Linguagem tenderá a aumentar o rendimento escolar nas demais disciplinas. Isso, empiricamente, já é percebido por serem a matemática e as linguagens a base de conhecimento para as demais disciplinas, mas é revelado e validado pela ASI. Outro exemplo é a relação do rendimento A em ciências humanas (HUMANAS_A) e naturais (NATUREZA_A) que implicam em rendimento A nas disciplinas técnicas (TECNICAS_A). Isso pode ser interpretado em prol de um planejamento escolar didático-pedagógico interdisciplinar que melhor integre essas disciplinas entre si, pois isso tenderá a contribuir com a elevação do rendimento nas disciplinas de conhecimento técnico.

Do ponto de vista multicultural, a segregação dos quilombolas dentro da escola deve ser objeto de preocupação, estudo e planejamento para efetivar a inclusão étnico-cultural durante o processo escolar, combatendo o baixo desempenho e a evasão por meio de planejamento didático-pedagógico pautado na educação multicultural e interculturalidade. De um modo geral, descobrir quais são os motivos que favorecem os estudantes sertanejos e urbanos quanto ao desempenho escolar classificado entre B e A e buscar ofertar um sistema educacional igualitário e equitativo onde todos e todas possam aprender interculturalmente e se reconhecerem na cultura escolar.

Outras relações poderiam ser descritas para exemplificar que a ASI proporciona uma compreensão não estática do objeto de pesquisa, potencializando a sua interpretação e utilização dos estudos na prática do campo estudado, em nosso caso a escola.

4 Considerações finais

O estudo revelou que existem diferenças significativas no desempenho escolar dentre estudantes sertanejos, urbanos, indígenas e quilombolas. Os estudantes quilombolas são os que apresentam baixo rendimento escolar e consequentemente são, em sua maioria, os que são alijados do processo de ensino pela evasão. Carraher e Schliemann (1983, p. 3) afirmaram que o alto índice de fracasso escolar nas camadas de baixa renda garantem a reprodução do sistema de classes.

A investigação sobre o rendimento escolar não deve ser considerada totalitariamente, mas se revela e se constitui como importante meio de compreensão do processo de escolarização e educação multicultural. Silva et al. (2019) relacionou o desempenho escolar com os tipos escolares (particular, pública, urbana e rural), condições sociais, atuação familiar na vida escolar dos estudantes, formação docente, dentre outros fatores que foram considerados. O fracasso estudantil quanto ao rendimento escolar é antes de tudo o fracasso da escola que falhou no cumprimento do seu papel de maneira equitativa e plenamente inclusiva.

O desempenho escolar é influenciado culturalmente, porém o estudo requer maior profundidade afim de identificar fatores e variáveis que influenciam nesse processo e que não considere simplesmente a realidade numérica dos registros escolares, mas contemple as inúmeras questões sociais e econômicas, didático-pedagógicas, institucionais, etc. A

influência cultural sobre o desempenho escolar por nós afirmada com base nos dados desse estudo, não está e nem deve ser compreendida como fator exclusivo ao estudante.

Apesar do nosso foco estar centrado no perfil estudantil, ressaltamos que na relação social escolar, o professor também contribui diretamente no desempenho do estudante, por ser este também portador de uma cultura que pode exercer determinado grau de poder sobre a cultura dos estudantes. Assim mesmo, podemos ainda inferir a escola, tal qualquer comunidade, também exercerá influência cultural, visto que a metodologia politico-pedagógica da instituição de ensino reverbera sobre a práxis do professor e essa consequentemente sobre a aprendizagem estudantil, revelando relações de poder culturais diversas e difusas.

Quanto a análise dos dados de desempenho escolar em função da cultura dos estudantes através dos métodos estatísticos de ANOVA e ASI observamos características interpretativas distintas em que, a associação dos métodos quantitativos não anulam e nem causam conflitos à interpretação dos dados. Podemos afirmar que a utilização simultânea dos métodos pode contribuir para uma compreensão holística dentro do campo de análise quantitativa, visto que se complementam.

Embora a integração dos métodos atribua à pesquisa a possibilidade de uma análise multifacetada, que se aproxima de uma interpretação complexa da realidade, nos cabe ainda, destacar a Análise Estatística Implicativa para a área das ciências humanas, mais precisamente à educação, por ser um método de natureza probabilística que se constitui como ferramenta analítica nas ciências humanas por não promover o determinismo interpretativo.

Esse trabalho não é estanque, não se mostra e nem pretende estar acabado. Diante de todo o exposto sobre a complexidade das relações culturais no contexto escolar e da potencialidade de aplicação de uma metodologia analítica multimétodo, inclusive multiquantitativa no tocante ao uso simultâneo da ASI e da ANOVA, essa temática requer maior ampliação de estudos e consequentes aprofundamentos em prol do desenvolvimento e estabelecimento de uma educação multicultural.

Escolas com público discente multicultural devem estar centradas em projetos e planos de cursos e pedagógicos baseados na Educação Multicultural e Decolonial, para que as influências socioculturais coloniais sejam pouco a pouco dirimidas e a escola possa vivenciar relações de desempenho escolar que não estejam marcadas sob a estrutura étnico-cultural estudantil. É urgente realizar a inclusão sociocultural não apenas para o ingresso e acesso escolar, mas para manutenção do estudante na escola e pela conclusão da escolaridade de indígenas, quilombolas, sertanejos e urbanos de qualquer classe social e gênero de maneira igualitária.

5 Agradecimentos

Ao Instituto Federal do Sertão Pernambucano pela concessão do afastamento para doutoramento e à FACEPE pela concessão da bolsa de estudos.

Referências

[1] ASI 11 (2021) *11ème colloque international sur l'Analyse statistique implicative* https://sites.univ-lyon2.fr/asi/11/index.php

[2] Bertinetto, C., Engel, J., and Jansen, J. (2020). ANOVA simultaneous component analysis: A tutorial review. *Analytica Chimica Acta: X*, 100061.

[3] Candau, V. M. F.; Moreira, A. F. B. (Org.) (2008). *Multiculturalismo*: diferenças culturais e práticas pedagógicas. Petrópolis, RJ: Ed. Vozes.

[4] Carraher, T. N. e Schliemann, A. L. D. (1983). Fracasso escolar: uma questão social. *Caderno de Pesquisa*, v. 45, p. 3-19.

[5] CHIC – *Classificação Hierárquica Implicativa e Coesitiva*. (2014). Versão 7.0, Copyright. Método de Análise Implicativa de dados de Régis Gras: École Polytechnique – Université de Nantes. Colaboração: Saddo Ag Almouloud, Marc Bailleul, Anleine Bodin,Annie Larher, Harrison Ratsimba-Rajohm, Jean-Claude Régnier, André Totohasina. Versão Windows: Raphaël Couturier.

[6] Corbett, M. (2015). Rural Education: some sociological provocation for the field. *Australian and International Journal of Rural Education*, v. 25, nº. 3, p. 9-25.

[7] Couturier, R., Bodin, A. e Gras R. (2004). *A classificação hierárquica implicativa e coesiva*. Manual Curso CHIC versão 2.3. Acesso em jun. 2021. Disponível em: http://math.unipa.it/~grim/asi/asi_03_gras_bodin_cout.pdf.

[8] Cunha, P. C., Ferreira, A. L., Acioly-Régnier, N. M. (2018). Estudo exploratório da formação humana nos trabalhos dos Colóquios Internacionais sobre Análise Estatística Implicativa (A.S.I.): Meta-análise dos Colóquios A.S.I. 5, A.S.I. 6 e A.S.I. 7. *V Congresso Nacional de Educação – V CONEDU*, Olinda-PE, Brasil, 17 a 20 de outubro de 2018.

[9] Di Paola, B. and Spagnolo, F. (2009) Argumentation and proving in multicultural classes: a didactical experience with Chinese and Italian Students. *Journal of Mathematics Education*, v. 2, nº. 1, p. 1-4.

[10] Eijck, M., and Roth, W. M. (2011). Cultural diversity in science education through novelation: against the epicization of science and cultural centralization. *Journal of Research Science Teaching*, v. 48, nº. 7, p. 824-847.

[11] Fazio, C. and Spagnolo, F. (2008). Conceptions on modeling processes in Italian high-school prospective mathematics and physics teachers. *South African Journal of Education*, v. 28, nº. 4, p. 469-488.

[12] Freitas, L. P. S. R., Campos, A. F., Andrade, V. L. V. X., Régnier, J. C. (2019). Análise estatística implicativa da tendência de abordagens do método de estudo de casos no ensino de ciências. *Revista de Educação em Ciências e Matemáticas*, v.15, nº. 33, p. 5-21.

[13] Gras, R., Almouloud, S. A. (2002). A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais. *Revista Educação Matemática Pesquisa*, v. 4, n. 2, p. 75-88, 2002.

[14] Gras, R., Larher, A. (1994). Um exemplo de utilização de uma análise implicativa para o exame de questionários. *Em Aberto*, ano 14, nº. 62, p. 42-51.

[15] Gras, R., Régnier, J. C. (2015). Origem e desenvolvimento da Análise Estatística Implicativa (A.S.I.). In: VALENTE, J. A.; ALMEIDA, M. E. B. (Orgs.). *Uso do CHIC na formação de educadores*: à guisa da apresentação dos fundamentos e das pesquisas em foco. 1ª ed. Rio de Janeiro: Letra Capital. Cap. 2, p. 22-45.

[16] Grimberg, B. I., and Gummer, E. (2013). Teaching Science from cultural points of intersection. *Journal of Research Science Teaching*, v. 50, n. 1, pp. 12-32.

[17] Leavitt, S. C. (1995). Suppressed Meanings in Narratives About Suffering: A Case from Papua New Guinea. *Anthropology and Humanism*, v. 20, nº. 2, p. 1–20.

[18] Louro, A. F. R. (2018). *Understanding students' academic achievement in public High School: evidence for Portugal*. Master Degree dissertation in Information Management, Universidade Nova de Lisboa, Portugal. 47 p.

[19] Memória, J. M. R. (2004). *Breve história da estatística*. Brasília/DF: Embrapa Informação Tecnológica, 111 p. ISSN: 1677-5473.

[20] Nadenka, B.M.B., e Molina-Andrade, A. (2015). Discussões sobre a transição entre escola, conhecimento científico e conhecimento ecológico tradicional: uma análise para salas de aula culturalmente diferenciadas. *X Encontro Nacional de Pesquisa em Ensino de Ciências-ENPEC*, Águas de Lindóia, SP, Brazil.

[21] Oris, M., Ritschard, G., Perroux, O. (2013). Le pluralisme religieux croissant a Genève dans la première moite du XIXe siècle : une exploration des dynamiques sous-jacentes. In: Amster, F., Scholl, S. *L'apprentissage du pluralisme religieux*. Le cas genevois au XIXe siècle. Genève: Labor et Fides.

[22] Paranhos, R.; Figueiredo-Filho, D. B.; Rocha, E. C.; Silva-Júnior, J. A.; Freitas, D. (2016). Uma introdução aos métodos mistos. *Sociologias*, ano 18, nº. 42, p. 384-411.

[23] Pasmiño-Maji, R. A., Garcia-Peñalvo, F. J., Conde-González, M. A. (2017). Statistical Implicative Analysis approximation to KDD and Data Mining: a systematic and mapping review in knowledge Discovery database framework. In: Schmidt, A., Laux, F., Hristovski, D., Ohnish, S. (Orgs.). *Proceedings of Nineth International Conference on Advances in Databases, Knowledge, and Data Applications, DBKDA*, Barcelona, Spain, May 21 to 25, 2017. p. 70-77, DE, USA: IARIA XPS Press.

[24] Pérez-Caraballo, G., Acioly-Régnier, N. M., Régnier, J. C. (2012). L'analyse statistique implicative et son utilisation en sciences humaines : apports à la recherche en Psychologie Interculturelle. *VI Colloque International A.S.I. Analyse Statistique Implicative*, Caen, France, 7 a 10 Novembre 2012.

[25] Phelan, P.; Davidson, A.; Cao, H. (1991). Students multiple worlds: Negotiating the boundaries of family, peer, and school cultures. *Anthropology and Education Quarterly*, v. 22, p. 224-250.

[26] Plonsky, L., and Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, *39*(3), 579-592.

[27] Régnier, J.-C. e Andrade, V. L. V. X de (2020a) A Análise Estatística Implicativa e Análise de Similaridade. In: J.-C. Régnier e V. L. V. X. de Andrade (dir.), *Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC*. (p. 41-840. Recife: EDUFRPE, 2020. Acesso em: fev. 2021. Disponível em: http://www.editora.ufrpe.br/ASI.

[28] Régnier, J.-C. e Andrade, V. L. V. X. de (2020b). Usando o software CHIC. In: J.-C. Régnier e V. L. V. X. de Andrade (dir.), *Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC*. (p. 85-164). Recife: EDUFRPE. 2020. Acesso em: fev. 2021. Disponível em: http://www.editora.ufrpe.br/ASI.

[29] Santos, J. L. G., Erdmann, A. L., Meirelles, B. H. S., Lanzoni, G. N. M., Cunha, V. P., Ross, R. (2017). Integração entre dados quantitativos e qualitativos em uma pesquisa de métodos mistos. *Texto Contexto Enfermagem*, v. 26, nº. 3: e1590016.

[30] Silva, A.P.F., Santos de Aquino, R., Silva, T. S., Oliveira, A. D. (2019). Desempenho escolar na disciplina de química em escolas do interior de Pernambuco. p. 129-139. *In:* Baggio, Vilmar (Org.). *Rumos da Educação*, v. 6, 2ª. ed., Veranópolis: Editora Diálogo Freiriano. ISBN: 978-65-80183-54-8.

[31] Stairs, A. (1995). Learning processes and teaching roles in Native education: Cultural base and cultural brokerage. In M. Battiste & J. Barman (Eds.), *First Nations education in Canada:* The circle unfolds, Vancouver, Canada: University of British Columbia Press. pp. 139-153.

[32] VERBI Software. (2019). *MAXAQDA 2020* [computer software]. Berlin, Germany: VERBI Software. Available from www.maxqda.com.

[33] Zakaria, E., and Syamaun, M. (2017). The effect of realistic mathematics education approach on students' achievement and attitudes towards mathematics. *Mathematics Education Trends and Research*, v. 1, nº. 1, p. 32-40.

[34] Zeynali, S., Pishghadam, R., and Fatemi, A. H. (2019). Identifying the motivational and demotivational factors influencing students' academic achievements in language education. *Learning and Motivation*, v. 68, 101598.

ÉTUDE DE LA DISTANCE ENTRE GRAPHES IMPLICATIFS POUR LA PRISE EN COMPTE DE LA TEMPORALITÉ EN ANALYSE STATISTIQUE IMPLICATIVE

Régis Gras[1], Antoine Bodin[2], Jean-Claude Régnier[3],

RESUME

La comparaison entre des résultats obtenus dans des conditions où les populations observées sont différentes mais où ils portent sur les mêmes variables, présente, pour le didacticien, un intérêt manifeste. Par exemple, il peut s'interroger sur l'effet d'un changement de stratégie didactique au cours du déroulement d'une séquence d'enseignement-apprentissage. Il peut également s'interroger sur l'évolution des effets d'un apprentissage donné au cours du temps, ou encore sur les comportements de populations différentes soumises à de mêmes épreuves didactiques et pédagogiques. Dans cette communication, supposant connues des matrices d'intensités d'implication issues de traitements ASI d'un ensemble de variables portant sur des populations comparables, nous poursuivons nos recherches déjà anciennes sur la construction de normes sur l'espace vectoriel des matrices carrées d'intensités d'implication de même taille. Cette norme nous permet alors de définir une distance entre les matrices et, par suite, nous offre un outil numérique qui permet de comparer les matrices et fournir un appui pour interpréter globalement les changements ou les stabilités éventuelles des phénomènes étudiés .

***Mots-clés** : analyse a priori, analyse a posteriori, contingence, norme*

ABSTRACT

The comparison between results obtained under conditions where the populations observed are different but relating to the same variables is of obvious interest to the educator. For example, he may wonder about the effect of a change in didactic strategy during an apprenticeship. But it can also question the lasting effect of learning over time, but also the behavior of different populations subjected to the same tests. In this chapter, on the supposedly known basis of the implicative matrices of situations generated by the same variables on comparable populations, we establish a norm on the vector set of similar matrices. This then allows us to define a distance between the matrices and, consequently, offers us a numerical tool which allows us to compare the matrices and to answer the initial questions.

***Keywords** : a priori analysis, a posteriori analysis, contingency, norm*

1 Introduction

Conformément aux bases théoriques de l'ASI, à un ensemble *E* d'individus, de cardinal *n*, on fait correspondre les valeurs 0 et 1 ou, plus généralement, des valeurs numériques que l'on transforme pour les ramener à des valeurs dans l'intervalle [0 ; 1],

1 Ecole Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge, Laboratoire LS2N, UMR 6241, Site de la Chantrerie, rue C. Pauc, BP 44306, Nantes cedex 3, e-mail : regis.gras@univ-nantes.fr, regispor@hotmail.fr.

[2] Consultant en matière de curriculum et d'évaluation dans le domaine des mathématiques; (France) antoinebodin@mac.com ;antoinebodin@me.com

[3] Professeur des universités émérite. UMR 5191 ICAR http://www.icar.cnrs.fr/membre/jcregnier/, Université Lumière Lyon 2. Professeur invité à l'Université d'État de Tomsk (Russie) – National Research Tomsk State University – jean-claude.regnier@univ-lyon2.fr

par des fonctions "attributs" d'un ensemble *V*, de cardinal *m*. Ces "attributs", tels $a \in V$, peuvent être des traits de caractère présents ou absents, réalisation de variables binaires, mais également des variables modales ou plus généralement numériques qui, à l'individu x, associent des valeurs telles que $a(x) \in [\,0\,;\,1]$.

À la suite de la mise en œuvre de traitements de la base de données *E* x *V* par l'analyse statistique implicative – ASI, on suppose alors connus deux états ou même plus, au moyen des matrices d'intensités d'implication carrées de taille m^2, qui sont ensuite traduites sous forme graphique à l'aide des graphes implicatifs construits à un niveau d'intensité d'implication choisi par le chercheur de valeur minimale 0.50. Notons que les traitements peuvent être étendus à des variables-conjonctions de deux ou plus de variables de *V*

On rencontre une telle situation, que nous considérons comme prototypique, dans le champ de l'éducation, en particulier dans le cadre de la didactique des mathématiques, où une même population d'élèves ou encore deux échantillons extraits d'une même population-mère est soumise à deux moments différents à un même questionnaire, l'un avant une intervention didactique et l'autre après celle-ci.

État 1					État 2			
V	...	b_j	...		V	...	b_j	...
...			...		...			...
a_i		$\psi_1(a_i, b_j)$	...	→	a_i		$\psi_2(a_i, b_j)$	...

Tableau 1

Il est intéressant d'étudier si, par exemple relativement à la même population d'étude, la structure globale des réseaux de relations de quasi-implication reste stable ou quasi-stable ou s'avère modifiée entre l'état 1 et l'état 2 fournis par les réponses obtenues à l'issue de la passation du questionnaire. Si nous retenons l'interprétation selon laquelle il y a eu modification, cela nous conduit à poursuivre une analyse partant de l'hypothèse de l'influence de l'intervention didactique.

Une autre situation se présente en didactique des mathématiques. Il s'agit de comparer les résultats d'une analyse *a priori* et ceux d'une analyse *a posteriori* portant sur des variables didactiques ou des variables qui modélisent des situations d'enseignement-apprentissage de concepts mathématiques. Une telle approche a été déjà abordée par (Spagnolo, Gras, Régnier, 2009, 2013). Cette situation repose d'abord sur la connaissance d'une matrice d'intensités d'implication *a priori* qui résulte du traitement ASI d'une base de données *E* x *V* construites par le didacticien en fonction des comportements de réponse des élèves ou des réponses elles-mêmes fournies par ceux-ci, attendus. Il vient ensuite la matrice des intensités d'implication *a posteriori* qui, elle, résulte du traitement de la base de données construites à partir de l'observation des comportements de réponse et des réponses effectivement fournies par les élèves.

Nous pouvons encore imaginer une autre situation mettant en œuvre plusieurs matrices d'intensités d'implication issues de données construites à partir de réponses relatives à un ensemble d'items proposés à des élèves de plusieurs pays, comme dans le cadre de l'évaluation internationale PISA (Montpied *et al*, 2019) (Christodoulou *et al*, 2017). Il s'agit alors de comparer les structures globales des réseaux de relations de quasi-implication supposés traduire l'organisation du savoir observable sur des groupes

d'élèves immergés dans des contextes culturels différents déterminés par le pays d'appartenance.

De façon générale, nous disposons donc d'une suite d'au moins deux matrices carrées d'intensités d'implication, non symétriques dont les graphes implicatifs ne sont que des traductions partielles. Une analyse qualitative de deux ou plusieurs graphes implicatifs peut apporter de précieuses informations sur les différences éventuelles entre les structures des réseaux de relations de quasi-implication formés entre les variables étudiées. Une question méthodologique se pose alors concernant le caractère significatif ou non de ces différences identifiées et caractérisées. Pour ce faire, il convient de tenir compte de l'information complète contenue dans les matrices d'intensités d'implication car les graphes implicatifs n'en restituent qu'une partie et des formes géométriques voisines peuvent être issues d'intensités d'implication distinctes. Dit autrement, les liaisons entre les variables étant le plus souvent différentes, la comparaison n'est pas possible sans revenir au tableau des mesures de qualité des relations quasi-implicatives sur V.

État 1					État 2					État 3			
V	...	b_j |	...		V	...	b_j	...		V	...	b_j	...
...			...		...			...		...			...
a_i		$\psi_1(a_i, b_j)$	...		a_i		$\psi_2(a_i, b_j)$	...		a_i		$\psi_3(a_i, b_j)$	...

Tableau 2

2 Comparaison de deux états en recourant à un test d'hypothèse statistique

Dans un premier, nous nous intéressons à la comparaison de deux graphes implicatifs G_1 et G_2. Nous examinerons plus tard une généralisation à plusieurs graphes implicatifs. Ces deux graphes sont des représentations graphiques partielles de deux matrices d'intensités d'implication carrées d'ordre m. Nous rappelons que les composantes de ces matrices sont les valeurs des intensités d'implication obtenues dans le cadre de la théorie classique ou celles des intensités d'implication-inclusion dans le cadre de l'approche dite entropique. Elles sont calculées à partir de la variable coefficient ou indice d'implication notée $Q(a, \bar{b})$ qui est reliée aux contre-exemples à la règle implicative. Nous ne revenons pas sur cette approche théorique. En ce qui concerne l'intensité d'implication notée $\varphi(a, b)$, ou celle d'implication-inclusion notée $\psi(a, b)$, nous rappelons que nous pouvons les considérer comme des réalisations d'une variable aléatoire de loi uniforme continue sur [0 ; 1] (Gras et al, 2017 p.60)

Nous supposons donc que nous disposons de deux matrices carrées M_1 et M_2, d'ordre m issues du croisement des mêmes m variables en jeu dans les deux observations. Ces matrices sont *a priori* indépendantes sous l'hypothèse nulle H_o par laquelle nous conjecturons qu'il n'y a pas de lien entre les deux familles d'observations. L'ensemble de cardinal m^2 des valeurs de leur différence M_1-M_2 sur [-1 ; 1] suit *a priori* la loi de probabilité qui régit D_u la différence de deux variables uniformes. La densité de cette loi est définie comme suit :

$$f(x) = \begin{cases} 0\ si\ x < -1\ ou\ si\ x > +1 \\ 1 + x\ si - 1 \leq x < 0 \\ 1 - x\ si\ 0 \leq x \leq +1 \end{cases}$$

L'espérance $E(D_u)=0$ et la variance $V(D_u) = \frac{1}{6}$

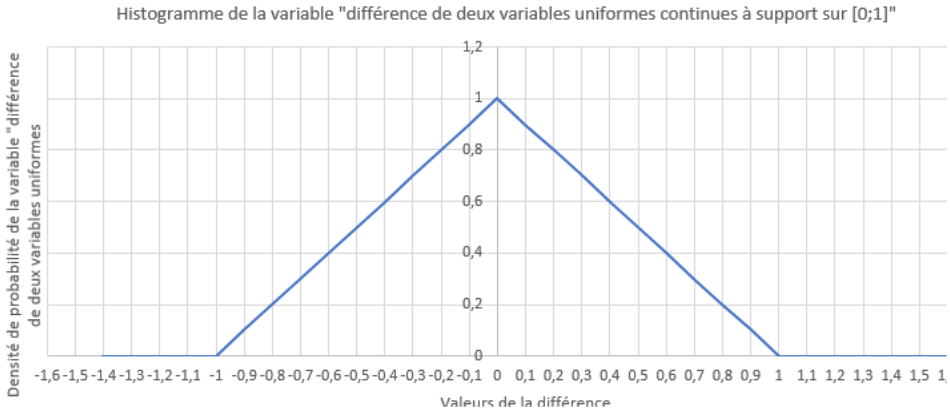


Figure 16 Histogramme et densité de la variable D_u

Une partition de ces valeurs sur [-1 ; 1] en k intervalles permet d'établir une comparaison entre la distribution observée, où l'intervalle I_i=[u, v[contient m_i observations, alors que la distribution théorique sous H_0 en contiendrait m^2p_i, et où p_i serait la probabilité de l'intervalle I_i sous la loi de la différence de deux variables uniformes chacune à support [0 ;1].

Nous choisissons une partition en k=20 intervalles d'amplitude 0,1 de l'intervalle [- 1 ; +1]. Le tableau ci-après fournit la distribution des probabilités :

[-1; -0,9[	[-0,9; -0,8[	[-0,8; -0,7[	[-0,7; -0,6[	[-0,6; -0,5[	[-0,5; -0,4[	[-0,4; -0,3[	[-0,3; -0,2[	[-0,2; -0,1[	[-0,1; 0[
0,005	0,015	0,025	0,035	0,045	0,055	0,065	0,075	0,085	0,095
[0; 0,1[	[0,1; 0,2[	[0,2; 0,3[	[0,3; 0,4[	[0,4; 0,5[	[0,5; 0,6[	[0,6; 0,7[	[0,7; 0,8[	[0,8; 0,9[	[0,9; 1[
0,095	0,085	0,075	0,065	0,055	0,045	0,035	0,025	0,015	0,005

Tableau 10 :distribution des probabilités sur les intervalles de la partition

On est donc ramené à la situation générale d'ajustement de deux distributions à l'aide de la variable de décision $D^2 = \sum_{i=1}^{i=k} \frac{(O_i - T_i)^2}{T_i}$ qui suit alors approximativement une distribution du χ^2 à *k-1* degrés de liberté. On compare alors la réalisation $d^2 = \sum_{i=1}^{i=k} \frac{(m_i - m^2 p_i)^2}{m^2 p_i}$ à la valeur critique $\chi^2_{critique}$ telle que $Prob\{\chi^2 > \chi^2_{critique}\} \leq \alpha$. Dans ce cas où k=20, il s'agit de la variable du χ^2 à *19* degrés de liberté pour la valeur critique $\chi^2_{critique} = 30{,}14$ pour un risque α=0.05.

L'hypothèse H_0 sera rejetée avec un risque de 1ère espèce de niveau α, si la valeur $d^2 > \chi^2_{critique}$

Si H_o est réfutée, cela traduit le fait que la différence entre les deux états ou entre l'analyse *a priori* et l'analyse *a posteriori*, est significativement grande selon le niveau de risque α choisi. Nous interprétons cet écart comme traduisant que les graphes G_1 et G_2 ont globalement des structure différente.

Nous pouvons aussi tenir compte du fait que les matrices d'intensités d'implication ou d'implication-inclusion ont leur diagonale principale constituées de valeurs nulles par construction de ces indices, c'est-à-dire que $V_j(i) = 0\ si\ i = j$. Ainsi les termes de la diagonale principale de la matrice M_1-M_0 sont toujours nuls et n'apportent aucune information sur la proximité entre les deux matrices.

Nous pouvons alors ne considérer que les m^2-m=m(m-1) valeurs non nécessairement nulles. En conséquence le calcul de la distance entre les effectifs théoriques espérés et les effectifs contingents se réalise par la formule :

$$d^2 = \sum_{i=1}^{i=k} \frac{(m_i - m(m-1)p_i)^2}{m(m-1)p_i}$$

3 Espace vectoriel normé de matrices d'indices d'implication ou d'indices d'implication-inclusion

On cherche à définir une topologie sur les matrices carrées M d'ordre m^2 dont les colonnes sont les vecteurs ($\vec{V}_1$, $\vec{V}_2$, …, $\vec{V}_m$). Pour tout j=1, 2, …, m, les composantes du vecteur $\vec{V_j}$ sont les m valeurs d'intensités d'implication $V_j(i) = \varphi(a_i \Rightarrow a_j)$ ou d'intensités d'implication-inclusion $V_j(i) = \psi(a_i \Rightarrow a_j)$. Dans notre cas, ce sont donc les valeurs des intensités d'implication-inclusion de chacune des variables croisées avec la variable relative au vecteur colonne. L'ensemble de telles matrices peut être muni d'une structure d'espace vectoriel bien que les opérations somme et produit par un scalaire n'aient pas de signification interprétable dans le domaine des règles de quasi-implication. En effet ces valeurs numériques appartenant à l'intervalle [0 ; 1] ont une nature comparable aux probabilités, ce qu'elles sont d'ailleurs si ce sont des intensités d'implication $\varphi(a_i \Rightarrow a_j)$ Cet espace vectoriel de matrices peut lui-même être normé de la façon suivante à partir des normes de ses vecteurs-colonnes. En choisissant la norme euclidienne, nous obtenons :

$$\left\|\overrightarrow{V_J}\right\|_2 = \left[\sum_{i=1}^{i=m} V_j^2\,(i)\right]^{\frac{1}{2}}$$

Nous pourrions ensuite choisir la norme matricielle suivante : $\|M\|_{sup} = \frac{Sup}{j}\left\{\left\|\vec{V_j}\right\|_2\right\}$

Nous pouvons aussi choisir la norme de Frobenius. Le calcul de la norme de la matrice M résulte de la formule suivante en notant la matrice adjointe M* qui, dans notre contexte, n'est autre que la matrice transposée ${}^tM$:

$$\|M\|_F = [TrM^*M]^{\frac{1}{2}} = \sqrt{\sum_{i,j} V_j^2(i)}$$

La relation entre les deux graphes implicatifs G_1 et G_2 peut alors être étudiée à partir de l'examen de la distance $d(M_1, M_2)$ entre les matrices d'intensités d'implication ou

d'implication-inclusion à partir desquelles ils sont construits. Cette distance permet d'étudier également les positions respectives d'autres matrices par rapport à un état de la base de données qui conduit à la matrice M_0. Par conséquent, si l'on recueille une suite temporelle de données qui déterminent une suite de matrices respectives M_0, M_1, …, M_k, il est alors possible les proximités relatives de chaque état par rapport à l'état initial.

Nous pouvons aussi tenir compte du fait que les matrices d'intensités d'implication ou d'implication-inclusion ont leur diagonale principale constituées de valeurs nulles par construction de ces indices, c'est-à-dire que $V_j(i) = 0\ si\ i = j$. Ainsi les termes de la diagonale principale de la matrice M_1-M_0 sont toujours nuls et n'apportent aucune information sur la proximité entre les deux matrices.

Par ailleurs, nous rappelons la propriété suivante $0 \leq V_j(i) \leq 1, \forall\ (i,j) \in \{1, \dots, m\}^2$ de laquelle nous pouvons déduire que les m^2 composantes de la matrice M_1-M_0 vérifient la propriété :

$$-1 \leq V_j^1(i) - V_j^0(i) \leq +1, \forall\ (i,j) \in \{1, \dots, m\}^2$$

$$0 \leq \left(V_j^1(i) - V_j^0(i)\right)^2 \leq +1, \forall\ (i,j) \in \{1, \dots, m\}^2$$

$$0 \leq \sum_{i,j} \left(V_j^1(i) - V_j^0(i)\right)^2 \leq m^2$$

Si nous tenons compte de la nullité des termes de la diagonale principale, nous pouvons améliorer la précision de l'encadrement :

$$0 \leq \sum_{i,j} \left(V_j^1(i) - V_j^0(i)\right)^2 \leq m^2 - m = m(m-1)$$

De là, nous en déduisons que la distance $d(M_0, M_1)$ vérifie la propriété suivante :

$$d(M_0, M_1) = \sqrt{\sum_{i,j} \left(V_j^1(i) - V_j^0(i)\right)^2} \leq \sqrt{m(m-1)}$$

Nous pourrions alors créer un indice qui mesurerait la proximité des deux matrices indépendamment de l'ordre m de la matrice carrée des différences de la manière suivante :

$$prox(M_0, M_1) = 1 - \frac{1}{\sqrt{m(m-1)}} d(M_0, M_1)$$

4 Conclusion

Dans le cadre de la didactique des disciplines, mais également dans toute recherche dite scientifique, se pose de manière récurrente le problème de la question légitime, voire incontournable, de la validation ou de la réfutation d'hypothèses émises par le chercheur. Se donner un moyen d'évaluer l'écart en termes quantitatifs entre les hypothèses *a priori* et les observations, la contingence, le contenu *a posteriori* est l'objet de cet article pour lequel il reste à mettre en œuvre sur des données de recherche les propositions théoriques explicitées.

Références

[2] Christodoulou T., Gagatsis A., Anastasiadou S. (2017), Enquête internationale comparative : croyances et attitudes des élèves du secondaire inférieur relatives aux erreurs en mathématiques, *Analyse Statistique Implicative, Points de vue conceptuels, applicatifs et* méta*phoriques,* (2017) Eds. Régnier J.-C., Gras R., Couturier R., Bodin A., ASI 9, IUT Belfort-Montbéliard

[3] Gras, R., Régnier, JC., Lahanier-Reuter, D. Marinica, C., Guillet, F. (Eds) (2017). *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales*. Toulouse : Cépaduès

[4] Montpied P., Duclos M., Le Hébel F. , Tiberghien A. [2019], Inégalité de la culture scientifique en France, Étude des caractéristiques impliquées dans les sources de difficulté des items Pisa Sciences, Ed. Régnier J.-C., Gras R., Henry M., Couturier R., Brousseau G., *Analyse Statistique Implicative, Cadre théorique en relation étroite et au service de multiples disciplines*, IUT Belfort Montbéliard.

[5] Spagnolo F., Gras R. y Régnier J.-C., (2009), *Una medida comparativa de las matemáticas entre el análisis a priori y la contingencia.* 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, p 143-158.

[6] Spagnolo F., Gras R. et Régnier J.-C., (2013) Mesurer l'écart entre une analyse a priori et la contingence en didactique*, L'Analyse Statistique Implicative, Méthode exploratoire et confirmatoire à la recherche de causalités,* sous la direction de Régis Gras. Toulouse : Ed. Cépaduès. pp.219-227

UNE CONSTRUCTION CONDENSÉE ET INTERACTIVE DU GRAPHE IMPLICATIF D'UNE BASE DE DONNÉES

Parfait Bemarisika,[1] Raphaël Couturier,[2] André Totohasina[3]

A CONDENSED AND INTERACTIVE CONSTRUCTION OF THE IMPLICATIVE GRAPH IN DATA BASE

RÉSUMÉ

Ce papier aborde la question de construction condensée et interactive d'un graphe implicatif, basé sur une mesure la plus discriminante M_{GK}, qui étend la théorie de l'Analyse Statistique Implicative (ASI) développée par Gras et ses collaborateurs. Le présent papier propose différentes optimisations pour la configuration du graphe en faisant éliminer les règles d'association non-intéressantes et redondantes. Basé sur ces optimisations, le papier présente trois algorithmes, dénommés GC2M, CONCISE et IGRAPH, permettant respectivement l'extraction des motifs fréquents (nœuds du graphe), la collection concise de l'ensemble des règles intéressantes et non-redondantes, et la construction des graphes implicatifs à partir de cet ensemble des règles valides. Il développe aussi un nouvel outil *rchicmgk* interface R implantant l'ensemble des graphes implicatifs à partir de cet ensemble de règles valides. Les résultats expérimentaux montrent la faisabilité de cette approche.

Mots-clés : *Règles d'association, graphe implicatif, rchicmgk.*

ABSTRACT

This paper addresses the problem of condensed and interactive construction of an implicative graph, based on a most discriminating measure M_{GK}, which extends the theory of Statistical Implicative Analysis (SIA) developed by Gras and his collaborators. This paper proposes different optimizations for the configuration of graphs by removing uninteresting and redundant association rules. Based on these optimizations, the paper introduces three algorithms, called GC2M, CONCISE and IGRAPH, respectively allowing the extraction of frequent patterns (nodes of the graph), the generation of interesting and non-redundant rules, and the construction of implicative graphs from these concise rules. He is also developing a new *rchicmgk* interface R tool implementing the implicative graphs thus obtained. The experimental results show the feasibility of the approach in SIA.

Keywords : *Association rules, implicative graph, rchicmgk.*

1 Introduction

Ce travail aborde la question de construction condensée et interactive d'un graphe implicatif. Un graphe implicatif est un graphe représentant l'ensemble de règles d'association d'une base de données. Une règle d'association est une implication logique de la forme $X \rightarrow Y$ (dite règle d'association positive), où X et Y sont deux itemsets disjoints, appelés respectivement la prémisse et le conséquent de la règle. Les règles d'association de la forme $X \rightarrow \overline{Y}$, $\overline{X} \rightarrow Y$ et $\overline{X} \rightarrow \overline{Y}$, qui représentent les relations *cachées* entre les items, sont appelées règles d'association négatives. De nombreux problèmes peuvent être résolus sous l'angle du graphe implicatif, par exemple Visualisation,

[1] ENSET, Université d'Antsiranana, Madagascar, bemarisikap7@yahoo.fr
[2]Institut FEMTO-ST, Univ. Bourgogne Franche-Comté, Belfort, raphael.couturier@univ-fcomte.fr
[3] ENSET, Université d'Antsiranana, Madagascar, andre.totohasina@gmail.com

Recherche du flot minimum dans un réseau de transport, des pages Internet, des serveurs, des Système d'embarquement, etc. Nous nous intéressons aux problèmes de visualisation. Ces derniers sont catégorisés en deux familles d'approches : Approche déterministe, et Approche statistique (ou stochastique). Cette dernière prend son essor considérable dans la communauté de l'Analyse Statistique Implicative (ASI).

Dans le cadre de l'approche déterministe, nous citons entre autres, Bruno *et al.*, 2007; Cheikh *et al.,* 2008 ; Mouakher et Ben Yahia, 2011 ; Hahsler, 2017 ; Carlos *et al.,* 2019. Cependant, ces approches sont insuffisantes quant au type des règles restituées. En effet, elles ne concernent que des règles d'association positives, les règles négatives sont absentes, et ce, avec le couple moins sélectif support-confiance (Agrawal *et al.,* 1994), alors que les règles d'association négatives peuvent être très utiles dans de nombreux domaines (Cao *et al.,* 2016 ; Dong *et al.,* 2018 ; Guyet et Quiniou, 2018 ; Rodriguez-Jimenez *et al.* 2014 ; Xu *et al.,* 2018), à savoir Classification, Biologie, Didactique de discipline, Data mining, Intelligence artificielle, Médecine, Système d'embarquement, Web Mining, etc. Par ailleurs, les techniques utilisées par ces approches nous ont permis de déceler un problème lié au volume important des règles d'association extraites dont plusieurs sont inintéressantes et redondantes qui conduisent souvent à des graphes denses. Par conséquent, l'utilisateur rencontre des difficultés pour effectuer convenablement l'interprétation de leurs résultats pour une prise de décision.

Dans le cadre de l'approche statistique, le modèle s'appuie sur l'intensité d'implication (Gras *et al.,* 1979). Comme le soulignent Gras et Kuntz (2008), Gras *et al.* (2013), Couturier et Ghanem (2015), Gras *et al.* (2015), Delacroix *et al.* (2017), Bodin et Giovannini (2019), l'intensité d'implication présente des défauts notables malgré son intérêt incontestable à la description sémantique d'une règle positive. En effet, elle a tendance à ne plus être suffisamment discriminante lorsque les données sont grandes et denses, et n'est donc pas à l'abri des règles d'association non-intéressantes (Cf. Section 3) qui pourraient bruiter exponentiellement les graphes. Comme dans les approches déterministes, les règles d'association négatives restent encore dans l'ombre dans les approches statistiques. Récemment, l'implifiance (Gras *et al.,* 2015) vient adjoindre l'intensité d'implication pour nuancer l'étonnement statistique de la règle $X \rightarrow Y$ et sa contraposée $\bar{Y} \rightarrow \bar{X}$. En ce sens, les règles négatives du type $X \rightarrow \bar{Y}$, $\bar{X} \rightarrow Y$ et $\bar{X} \rightarrow \bar{Y}$ restent encore ignorées. D'ailleurs, la règle $\bar{Y} \rightarrow \bar{X}$ est une règle redondante par rapport à la règle $X \rightarrow Y$, ce qui bruite tout simplement les graphes. Par ailleurs, étant donné sa définition sémantique, l'*implifiance* (Déf. 6) est la fusion de l'intensité d'implication (Gras, 1979) et de la confiance (Agrawal *et al.,* 1994). En conséquence, elle hérite naturellement les défauts majeurs de son ancêtre primitif en termes de sélection.

Pour pallier ces notables limites, nous proposons une nouvelle méthode de construction condensée et interactive du graphe implicatif très lisible et interprétable, basée sur la mesure M_{GK} introduite initialement dans (Guillaume, 2000) puis affinée par Totohasina et Ralambondrainy (2005). L'approche intègre simultanément les règles positives et négatives caractérisées par des règles exactes (i.e. $M_{GK} = 1$) et approximatives (i.e. $M_{GK} \neq 1$). Les principales contributions sont les suivantes. 1) Nous proposons l'algorithme GC2M (*Generator itemsets, Closed itemsets, Maximal itemsets, and Minimal infrequent Itemsets)*, un algorithme autonome pour l'extraction simultanée l'ensemble des générateurs, fermés fréquents, maximaux fréquents fermés, et minimaux non-fréquents d'une base de données. 2) Nous introduisons une définition formelle d'une règle d'association inintéressante (UAR), et proposons ensuite une stratégie d'élagage de

ce type UAR. 3) Nous proposons un nouveau modèle de sélection d'une règle potentiellement valide. 4) Nous développons une nouvelle stratégie pour l'espace de recherche des règles valides. 5) Nous étendons Pasquier *et al.,* 2005 auquel nous définissons une nouvelle base pour les règles positives exactes, appelée *Base concise pour les règles positives exactes (CBE^{+}).* 6) Nous proposons aussi trois nouvelles bases : *Base concise pour les règles positives approximatives (CBA^{+}), Base concise pour les règles négatives exactes (CBE^{-}), et Base concise pour les règles négatives approximatives (CBA^{-}).* 7) Basé sur ces formalisations, nous introduisons un nouvel algorithme permettant la collection concise de règles non-redondantes. 8) A partir de cet ensemble des règles non-redondantes, nous développons aussi un algorithme, dénommé IGRAPH, permettant la construction efficace des graphes implicatifs à partir de cet ensemble de règles valides. 9) Nous développons un nouvel outil *rchicmgk* interface R implantant l'ensemble des graphes implicatifs. 10) Nous présentons les expérimentations et les différents résultats obtenus avec des interprétations.

Le papier est organisé comme suit. La Section 2 présente le cadre formel. La Section 3 détaille l'approche proposée. Section 4 décrit l'évaluation expérimentale de l'approche proposée. La conclusion et les perspectives sont données dans la Section 5.

2 Cadre formel

Soit $\mathcal{B} = (\mathcal{T}, I, \mathcal{R})$ un contexte de fouille de données, où $\mathcal{T}$ et I sont des ensembles finis non vides, composés respectivement de n transactions et de m items. Un élément de I est appelé item et un élément de $\mathcal{T}$ est appelé transaction représentée par TID, et $\mathcal{R} \subseteq \mathcal{T} \times I$ une relation binaire entre $\mathcal{T}$ et I. On note $i\mathcal{R}t$ si un item i est présent dans t. Le Tableau 1 donne un exemple d'une base de données ayant 6 transactions et 5 items.

TID	Item
1	ACD
2	BCE
3	ABCE
4	BE
5	ABCE
6	BCE

Tableau 1– Contexte formel $\mathcal{B}$

L'élément $X \subseteq I$ est appelé itemset, et $Y \subseteq \mathcal{T}$ est appelé tidset. Un k-itemset est appelé itemset de longueur (ou de taille) k. Soit $X \subseteq I$, on a $\bar{X} = \neg X = \{t \in \mathcal{T} | \exists\, i \in X : (i,t) \notin \mathcal{R}^2\}$ c'est complémentaire de X dans I *(i.e. $I \backslash X$)*. Soit 2^I (resp. $2^{\mathcal{T}}$) l'ensemble des parties de I (resp. l'ensemble des parties de $\mathcal{T}$). Pour $X \subseteq I$ et $Y \subseteq \mathcal{T}$, nous définissons deux fonctions $\phi(X)$ et $\psi(Y)$ qui définissent la connexion de Galois (Ganter & Wille, 1999) entre 2^I et $2^{\mathcal{T}}$, elles s'appellent aussi extension de X et intension de Y respectivement. Etant donnée la relation binaire $\mathcal{R} \subseteq \mathcal{T} \times I$, $X \in 2^I$ et $Y \in 2^{\mathcal{T}}$, nous avons : $2^I \to 2^{\mathcal{T}},\ \phi(X) = X' = \{t \in \mathcal{T} | i\mathcal{R}t, \forall i \in X\}$ et $2^{\mathcal{T}} \to 2^I,\ \psi(Y) = Y' = \{i \in I | i\mathcal{R}t, \forall t \in Y\}$. Le support de l'itemset X s'écrit de la façon suivante :

$$supp(X) = P(X') = \frac{|\phi(X)|}{|\mathcal{T}|}, \quad (1)$$

où $|\ell|$ est la cardinalité de ℓ et P la probabilité discrète. Considérons un seuil *minsup* tel que $0 < minsup \leq 1$, X est dit fréquent si $supp(X) \geq minsup$. On notera par :

$$\mathcal{F} = \{X \subseteq I | X \neq \emptyset \wedge supp(X) \geq minsup\}, \quad (2)$$

l'ensemble des motifs fréquents dans $\mathcal{B}$. L'opérateur de fermeture est défini par :

$$\gamma(X) = \psi o \phi(X) \quad (3)$$

Définition 1 (Classe d'équivalence). X *et* Y sont dits équivalents, dénotés $X \cong Y$, si $\gamma(X) = \gamma(Y)$. La classe d'équivalence de X est définie par $[X] = \{Y \subseteq X | X \cong Y\}$.

Propriété 1 (Agrawal et al., 1994). Pour tous $X, Y \subseteq I$, on a $X \subseteq Y \Rightarrow \phi(X) \supseteq \phi(Y)$.

Définition 2 (Motif fermé fréquent). X est fermé si $\gamma(X) = X$. Nous définissons $\mathcal{FC} = \{\mathcal{C} \subseteq I | \mathcal{C} \neq \emptyset \wedge supp(\mathcal{C}) \geq minsup\}$ l'ensemble des motifs fermés fréquents.

Définition 3 (Générateur minimal). Un motif G est générateur s'il est minimal (au sens de l'inclusion) dans sa classe d'équivalence. Nous définissons $\mathcal{G}_{\mathcal{C}} = \{\mathrm{G} \in \mathrm{I} | \mathcal{C} \in \mathcal{FC}, \emptyset \neq \mathrm{G} \subseteq \mathrm{C} \subseteq \mathrm{I}, \mathrm{G} \cong \mathrm{C}, \nexists G^* \subset G \in \mathcal{F}\}$ l'ensemble des générateurs.

Propriété 2 (Pasquier et al., 2005). Pour tout $X \subseteq I$, on a $supp(X) = supp(\gamma(X))$.

Définition 4 (Motif maximal fréquent). X est maximal fréquent si $\mathrm{X} \in \mathcal{F}$ et $\nexists$ Y tel que $\mathrm{Y} \supset \mathrm{X}$. On définit $\mathcal{MFC} = \left\{\mathcal{C} \in \mathcal{FC} \middle| \nexists \tilde{\mathcal{C}} \supset \mathcal{C}, \tilde{\mathcal{C}} \in \mathcal{FC}\right\}$ l'ensemble des maximaux.

En d'autres termes, X est maximal si aucun de ses sur-ensembles n'est fréquent. En remarquant que les motifs maximaux, notés $\mathcal{M}$, sont moins nombreux que les motifs fermés, notés $\mathcal{C}$. Par définition, tout motif maximal est nécessairement fermé.

Traditionnellement, une règle d'association est associée par des mesures de qualité. Nous présentons ensuite quelques mesures les plus classiques afin de quantifier une règle d'association. Le support et la confiance de $X \rightarrow Y$ s'écrivent respectivement :

$$supp(X \cup Y) = \frac{|\phi(X \cup Y)|}{|\mathcal{T}|} \text{ et } conf(X \rightarrow Y) = P(Y'|X') = \frac{|\phi(X \cup Y)|}{|\phi(X)|} \quad (4)$$

Dans la suite, nous introduisons les quantités $n = |\mathcal{T}|$, $n_X = |\phi(X)|$, $n_Y = |\phi(Y)|$, $n_{XY} = |\phi(X \cup Y)|$ et $n_{X\bar{Y}} = |\phi(X \cup \bar{Y})|$. Aussi, $N_{X\bar{Y}}$ est la variable aléatoire qui génère $n_{X\bar{Y}}$, et N_{XY} celle qui génère n_{XY}. L'intensité d'implication (Gras *et al.,* 1979) est le complément à 1 de la p-value unilatérale du nombre de contre-exemples (cf. Définition 5). Son calcul dépend de la distribution de la variable aléatoire $|\phi(Z \cup \bar{T})|$ qui suit une loi de Poisson de paramètre $\frac{n_X n_{\bar{Y}}}{n}$, où Z et T sont deux variables de I, choisies aléatoirement et indépendamment de mêmes cardinaux respectifs que X et Y.

Définition 5 (Intensité d'implication). Soient $X, Y \subseteq I$ tels que $X \cap Y = \emptyset$. On appelle intensité d'implication de la règle d'association $X \rightarrow Y$, telle que $n_Y \neq n$, le nombre :

$$\varphi(X, Y) = 1 - P(N_{X\bar{Y}} \leq n_{X\bar{Y}}) \quad (5)$$

L'implifiance (Gras *et al.,* 2015) est la mesure de l'implication statistique qui prend en compte l'implication directe et sa contraposée ainsi que la confiance (cf. Définition 6).

Définition 6 (Implifiance). On appelle implifiance de la règle $X \rightarrow Y$, le nombre :

$$\Phi(X, Y) = \varphi(X, Y).\left[P(\mathrm{Y}'|\mathrm{X}').\mathrm{P}(\bar{X}'|\bar{\mathrm{Y}}')\right]^{1/4} \quad (6)$$

Définition 7 (Mesure M_{GK}). Soient X et Y de I, la mesure M_{GK} (Guillaume, 2000 ; Totohasina et Ralambondrainy, 2005) de la règle d'association $X \rightarrow Y$ est définie par :

$$M_{GK}X \rightarrow Y) = \begin{cases} \frac{P(Y'|X') - P(Y')}{1 - P(Y')}, si\ P(Y'|X') > P(Y') \text{ et } P(Y') \neq 1 \\ \frac{P(Y'|X') - P(Y')}{P(Y')}, si\ P(Y'|X') \leq P(Y') \text{ et } P(Y') \neq 0. \end{cases} \quad (7)$$

M_{GK} varie dans l'intervalle [-1,+1]. Elle mesure la distance de la confiance à l'indépendance. Plus M_{GK} s'éloigne de 0, plus le couple (*X,Y)* est fortement dépendant. Plus M_{GK} est proche de 1 (resp. -1) plus le couple (*X,Y*) est positivement dépendant (resp. négativement dépendant). Si $M_{GK}(X \rightarrow Y) = -1$, alors *X* défavorise *Y*, ce qui correspond à la répulsion limite entre *X* et *Y*. Si $M_{GK}(X \rightarrow Y) = 0$, alors *X* défavorise aussi *Y*, ce qui caractérise alors l'indépendance statistique entre *X* et *Y*. Si $M_{GK}(X \rightarrow Y) = 1$, alors *X* favorise *Y*, ce qui dénote l'attraction forte entre les motifs *X* et *Y* .

3 Construction condensée d'un graphe implicatif

Notre approche sera réalisée en trois étapes. La première étape consiste à extraire les motifs fréquents associés aux sommets du graphe implicatif. Il s'agit de déterminer l'ensemble $\mathcal{G}_C$ des générateurs, l'ensemble $\mathcal{FC}$ des fermés fréquents, l'ensemble $\mathcal{MFC}$ des motifs maximaux fréquents, et l'ensemble $\bar{\mathcal{F}}_{min}$ des motifs minimaux non-fréquents d'une base de données. La seconde étape se focalise sur la collection concise des règles d'association significativement valides à partir de ces quatre sous-ensembles des motifs fréquents. La troisième étape porte sur la configuration topologique et algorithmique d'un graphe implicatif à partir de cette famille des règles d'association valides.

3.1 Extraction des motifs fréquents

La principale motivation repose sur l'absence dans la littérature d'un algorithme autonome permettant la génération simultanée de sous-ensembles $\mathcal{G}_C$, $\mathcal{FC}$, $\mathcal{MFC}$ et $\bar{\mathcal{F}}_{min}$ des motifs fréquents. Ainsi, nous proposons un algorithme, dénommé GC2M, qui collecte simultanément ces quatre sous-ensembles $\mathcal{G}_C$, $\mathcal{FC}$, $\mathcal{MFC}$ et $\bar{\mathcal{F}}_{min}$. Notons que l'extraction simultanée de ces sous-ensembles pose un énorme problème lorsqu'on est en présence d'une base de données dense. L'un des problèmes repose sur le calcul de support pour chaque candidat. Ainsi, nous proposons une nouvelle méthode permettant de calculer le support d'un non générateur que le Théorème 1 ci-après l'expose.

Théorème 1 (Bemarisika et Totohasina, 2020). Le support d'un k-itemset ($k \geq 3$) non-générateur est exactement un minimum de support de ses sous-ensembles $(k-1)$-itemsets. Soit X un k-itemset ($k \geq 3$) non-générateur, $\exists \tilde{X} \subseteq I$ tel que $\tilde{X} \subset X$, on a :

$$supp(X) = min\{supp(\tilde{X}) | \tilde{X} \subset X\} \quad (5)$$

A noter que le Théorème 1 n'est utilisé qu'à un itemset non-générateur de longueur supérieure ou égale à 3. En effet, les 1-itemsets n'ont pas de sous-ensembles ; les 2-itemsets ont exactement 2 sous-ensembles de taille 1 qui sont générés a partir de deux 1-motifs fréquents, donc forcément fréquents, inutile d'utiliser le Théorème 1. Alors, l'originalité de ce Théorème 1 repose sur le fait qu'il permet d'éviter l'accès systématique à la base de données pour calculer les supports candidats : le support d'un k-itemset non-

générateur est dérivable de ses sous-ensembles $(k-1)$-itemsets. Par exemple, du tableau 1, on a *supp*(*AC*) = *supp*(*A*) = 3/6 et *supp*(*BE*) = *supp*(*B*) = *supp*(*E*) = 5/6 impliquent *AC* and *BE* sont non-générateurs. Leur sur-ensemble *ABC* l'est aussi, son support est alors dérivable de ses sous-ensembles *AC* et *BE,* sans avoir passée à la base de données : *supp*(*ABC*) = *min*{*sup*(*AB*), *sup*(*AC*), *sup*(*BC*)} = *min*{2/6, 3/6, 4/6} = 2/6. Dans le cadre des 1-motifs, 2-motifs et d'un générateur de type supérieur, l'algorithme GC2M se fonctionne comme Apriori (Agrawal *et al.,* 1994) : il utilise le support classique de l'équation 1. En termes d'espace de recherche, l'algorithme GC2M utilise une stratégie d'élagage efficace que les Théorèmes 2 et 3 ci-après l'expliquent.

Théorème 2 (Agrawal & Srikant, 1994). (1) Tout sous-ensemble d'un motif fréquent est fréquent. (2) Tout sur-ensemble d'un motif non-fréquent est non-fréquent.

Théorème 3 (Bemarisika & Totohasina, 2020). Soit $X \subseteq I$. Si X est un générateur, alors pour tout $\tilde{X} \subset X$, $\tilde{X}$ est aussi générateur.

Ces résultats seront synthétisés dans l'algorithme 1 ci-après. GC2M (Algorithme 1)

Algorithm 1 GC2M

Require: A database $\mathcal{B}$, A minimum support threshold $minsup \in]0, 1]$.
Ensure: $\mathcal{G}_{\gamma(.)}$, $\mathcal{FC}$ and $\mathcal{MFC}$.

1: $\mathcal{FCC}_1$.GENERATORS $\leftarrow$ {1-itemsts}
2: **for all** $(k \leftarrow 1; \mathcal{FCC}_k$.GENERATORS $\neq \emptyset; k++)$ **do**
3: $\quad \mathcal{FCC}_k$.CLOSURE $\leftarrow \emptyset$; $\mathcal{FCC}_k$.SUPPORT $\leftarrow 0$;
4: $\quad \mathcal{FCC}_k \leftarrow$ GENCLOSURES$(\mathcal{FCC}_k)$
5: $\quad$ **for all** (candidate itemsets $c \in \mathcal{FCC}_k$) **do**
6: $\quad\quad$ Calculate $supp(c)$;
7: $\quad\quad$ **if** $(supp(c) \geqslant minsup)$ **then**
8: $\quad\quad\quad \mathcal{FC}_k \leftarrow \mathcal{FC}_k \cup \{c\}$
9: $\quad\quad$ **end if**
10: $\quad$ **end for**
11: $\quad \mathcal{FCC}_{k+1} \leftarrow$ GENGENERATORS$(\mathcal{FC}_k)$
12: $\quad \mathcal{FCC}_{k+1} \leftarrow$ GENMAXIMAL$(\mathcal{FC}_k)$
13: **end for**
14: $\zeta \leftarrow \bigcup_{j=1}^{k-1} \{\mathcal{FC}_j$.CLOSURE, $\mathcal{FC}_j$.GENERATORS, $\mathcal{FC}_j$.MAXIMAL, $\mathcal{FC}_j$.SUPPORT$\}$
15: **return** U

Figure 1 : Schéma de l'algorithme 1 GC2M

prend en entrée une base de données $\mathcal{B}$, un *minsup*, et retourne en sortie les ensembles $\mathcal{G}_C$, $\mathcal{FC}$, $\mathcal{MFC}$ et leurs supports respectifs. L'algorithme 1 commence par générer les 1-itemsets (ligne 1). Les listes $\mathcal{FCC}_k$.GENERATORS des k-candidats générateurs et $\mathcal{FCC}_k$.CLOSURE des k-candidats fermés sont initialisées à vide, alors que la liste $\mathcal{FCC}_k$.SUPPORT des supports de ces ensembles est initialisée à zéro (lignes 2-3). Dans un premier temps, GC2M recherche les k-motifs fermés en utilisant la fonction GENCLOSURES (lignes 4-10). Durant la recherche, une liste $\mathcal{FC}_k$ des k-fermes fréquents est maintenue lorsqu'un candidat c est fréquent (lignes 5-6). Une fois que les k-motifs fermés sont trouvés, l'algorithme GC2M procède des deux procédures secondaires GENGENERATORS et GENCMAXIMAL qui génèrent respectivement les k-générateurs et k-maximaux à partir de ces k-fermés (lignes 11-12). Enfin, la liste ζ classe les k-motifs générateurs, fermés, maximaux et leurs supports selon leur ordre de sélection (ligne 14).

Le tableau 2 ci-après représente l'exemple d'exécution de cet algorithme 1 mené sur une petite base de données du tableau 1, avec un *minsup=2/6*. L'ensemble $\mathcal{FCC}_1$ est initialisé avec les 1-motifs. Dans ce cas, le 1-motif D et son sur-ensemble ACD sont

supprimés depuis $\mathcal{FCC}_1$, car ils ne sont pas fréquents. Le 1-motif *C* est à la fois générateur et fermé, il sera alors ajouté dans $\mathcal{G}_C$ et $\mathcal{FC}$. Six candidats *AB*, *AC*, *AE*, *BC*, *BE* et *CE* sont ensuite crées. On voit que *AC* et *BE* sont respectivement des fermés de *A*, et des (*B* et *E)*, ils seront alors ajoutés dans $\mathcal{FC}$; tandis que *BC* et *CE* sont des générateurs, et ajoutés à leur tour dans $\mathcal{G}_C$. Enfin, *ABE* et *ABCE* sont créés. Le 3-motif

Scan $\mathcal{B}$

$\mathcal{FCC}_1$		Supp
A	AC	3/6
B	BE	5/6
C	C	5/6
D	ACD	1/6<2/6
E	BE	5/6

Supprimer les infréquents

$\mathcal{G}_{\gamma(.)}$	$\mathcal{FC}$	$\mathcal{MFC}$	Supp
A	AC	-	3/6
B	BE	-	5/6
C	C	-	5/6
E	BE	-	5/6

Scan $\mathcal{B}$

$\mathcal{FCC}_2$		Supp
AB	ABCE	2/6
AE	ABCE	2/6
BC	BCE	4/6
CE	BCE	4/6

Supprimer les infréquents

$\mathcal{G}_{\gamma(.)}$	$\mathcal{FC}$	$\mathcal{MFC}$	Supp
AB	ABCE	ABCE	3/6
AE	ABCE	ABCE	5/6
BC	BCE	-	5/6
CE	BCE	-	5/6

Tableau 2– Liste des $\mathcal{G}_{\gamma(.)}$, $\mathcal{FC}$ et $\mathcal{MFC}$ de la base $\mathcal{B}$ du Tableau 1 sous *minsup=2/6*

ABE est supprimé (car il est sur-ensemble de $BE \notin \mathcal{FCC}_2$), alors que le 4-motif *ABCE* est ajouté dans $\mathcal{FC}$ et $\mathcal{MFC}$ car il est à la fois fermé et maximal des *AB* et *AE.*

Nous définissons dans ce qui suit l'ensemble $\bar{\mathcal{F}}_{min}$ des motifs minimax non-fréquents (Minimal Infrequent Subset) appelé MIS. L'ensemble d'un motif minimal non fréquent peut être dérivé d'un ensemble des maximaux fréquents $\mathcal{MFC}$.

Définition 8. Soit $\mathcal{MFC}$ l'ensemble des maximaux fréquents, $\mathcal{F}$ celui des fréquents. On appelle l'ensemble des motifs minimaux non fréquents associé à un maximal, noté $\bar{\mathcal{F}}_{min}$, l'ensemble complémentaire d'un $\mathcal{MFC}$ dans I, défini de la façon suivante :

$$\bar{\mathcal{F}}_{min} = \{X \in I \backslash \mathcal{MFC} | \nexists Y \subset X, Y \notin \mathcal{F}\} \quad (6)$$

Théorème 4. L'ensemble $\mathcal{MFC}$ des maximaux associé à l'ensemble $\bar{\mathcal{F}}_{min}$ est donné :

$$\mathcal{MFC} = \{X \in I | \bar{X} \in \bar{\mathcal{F}}_{min}\} \quad (7)$$

où $\bar{X}$ est le complémentaire d'un ensemble $\mathcal{MFC}$ dans I.

Preuve. (1) Soit $X \subseteq I$. Nous montrons tout d'abord que $X \in \mathcal{MFC} \Leftrightarrow \bar{X} \in \bar{\mathcal{F}}_{min}$. En effet, $X \in \mathcal{MFC} \Leftrightarrow \forall Y \in \bar{\mathcal{F}}_{min} Y \subsetneq X \Leftrightarrow \forall Y \in \bar{\mathcal{F}}_{min}, Y \cap \bar{X} \neq \emptyset$, ce qui prouve que $\bar{X}$ est un minimal d'un $\bar{\mathcal{F}}_{min}$. Nous montrons maintenant que X est maximal fréquent dans $\mathcal{MFC} \Leftrightarrow \bar{X}$ est un minimal non-fréquent de $\bar{\mathcal{F}}_{min}$. Soit $X \in \mathcal{MFC}$. Puisque $X \in \mathcal{MFC}$, équivaut alors à $\bar{X}$ est minimal de $\bar{\mathcal{F}}_{min}$. Supposons qu'il ne soit pas minimal : il existe $Y \in I$ tel que $Y \in \bar{\mathcal{F}}_{min}$ et $Y \subset \bar{X} \Leftrightarrow X \subset \bar{Y}$. Donc, $\bar{Y} \in \mathcal{MFC}$ et $X \subset \bar{Y}$, ce qui contredit le fait que $X \in \mathcal{MFC}$. Enfin, soit $X \in \bar{\mathcal{F}}_{min}$, donc $\bar{X} \in \mathcal{MFC}$. Supposons que $\bar{X}$ ne soit pas maximal : $\exists Y \in \mathcal{MFC}$ tel que $\bar{X} \subset Y$. Donc, $\bar{Y}$ est un minimal non-fréquent avec $\bar{X} \subset Y$, soit $\bar{Y} \subset X$, ce qui contredit le fait que X est un minimal dans $\bar{\mathcal{F}}_{min}$.

L'algorithme 2 ci-après représente l'extraction des motifs minimaux dans une base de données. Il prend en entrée une base de données $\mathcal{B}$ et un *minsup*, et retourne en sortie l'ensemble $\bar{\mathcal{F}}_{min}$ des minimaux. D'après les résultats de la figure 1, nous avons trouvé

Algorithm 2 GC2M-MINIMAL INFREQUENT ITEMSETS

Require: $\mathcal{B}$, $\mathcal{MFC}$ and $minsup \in]0, 1]$.
Ensure: $\overline{\mathcal{F}}_{min}$ the set of minimal infrequent itemsets.
1: $\overline{\mathcal{F}}_{min} \leftarrow \emptyset$
2: **for all** $(X \in 2^{\mathcal{I}} \setminus \mathcal{MFC})$ **do**
3: **if** $(\nexists Y \subset X \mid supp(Y) \leq minsup)$ **then**
4: $\overline{\mathcal{F}}_{min} \leftarrow \overline{\mathcal{F}}_{min} \cup \{X\}$
5: **end if**
6: **end for**
7: **return** $\overline{\mathcal{F}}_{min}$

Figure 2 :: Schéma de l'algorithme 2 GC2M

que $D \notin \mathcal{F}$ et $\nexists \widetilde{D} \subset D$ tel que $\widetilde{D} \notin \mathcal{F}$ (i.e. $supp(\widetilde{D}) \leq minsup$) $\Rightarrow \overline{\mathcal{F}}_{min} = \{D\}$.

3.2 Collection concise des règles d'association valides

Communément, la taille conséquente des règles d'association pèse lourd sur les coûts de construction d'un graphe implicatif. Ainsi, nous proposons un nouveau critère de sélection des règles d'association avant la représentation graphique.

3.2.1 Critère de sélection d'une règle d'association valide

Nous utilisons, au gré de sélection d'une règle valide, la mesure M_{GK}. Notons que la première composante de cette mesure M_{GK} (Equation (1)) est implicative mais la seconde n'en est. Seulement la première sera active dans la modélisation. Cette procédure s'avère ainsi plus cohérente à l'interprétation causale d'une règle d'association. Signalons aussi que le calcul d'un seuil de sélection des règles valides selon M_{GK} est un peu moins immédiat (voire difficile) en raison de sa normalisation. La version actuelle (Feno et Totohasina, 2008) est basée sur la valeur critique définie par :

$$\eta_\alpha = \sqrt{\frac{1}{n}\frac{n - n_X}{n}\frac{n}{n - n_Y}\chi^2(\alpha)} \tag{9}$$

où χ^2 est une statistique de Khi-deux à un degré de liberté, qui dépend de α un seuil critique choisi par l'utilisateur dans l'intervalle $]0,1]$. Dans ce cas, une règle $X \rightarrow Y$ sera valide lorsque $M_{GK}(X \rightarrow Y) \geq \eta_\alpha$. Malgré son intérêt incontestable, ce modèle présente des limites notables. Sa principale faiblesse est d'ignorer les règles d'association ayant une valeur forte, dites règles potentiellement intéressantes (i.e. règles proches de 1), mais inférieure à la valeur critique η_α. A l'inverse, elle accepte des règles d'association faiblement intéressantes (i.e. règles proches de 0) mais supérieure à la valeur critique.

Afin de pallier ces notables défauts, nous proposons une méthode permettant d'éliminer les mauvaises règles facilement enregistrées, et de récupérer les bonnes règles pourraient être écartées facilement. L'idée est d'affiner efficacement les règles retenues en utilisant M_{GK} renormalisée, et avec un nouveau paramètre de sélection d'une règle valide. En effet, la sensibilité de cette mesure M_{GK} aux variations des occurrences du nombre de contre-exemples se lit avec la dérivée partielle ci-après :

$$\frac{\partial M_{GK}}{\partial n_{X\bar{Y}}} = -\frac{1}{\frac{n_X n_{\bar{Y}}}{n}} \tag{10}$$

Ce qui relève que M_{GK} décroît quand le nombre de contre-exemples augmente et d'autant plus vite que la quantité $\frac{n_X n_{\bar{Y}}}{n}$ est relativement faible. A cet effet, le nombre $n_{X\bar{Y}}$ des contre-exemples doit être pris en compte pour accepter ou rejeter statistiquement une règle d'association $X \to Y$. Cela conduit à observer quelques transactions dans la base de données contenant X et non Y sans avoir la tendance générale d'avoir Y lorsque X est présent contesté. Le modèle que nous proposons pour cela est similaire à celui proposé dans (Gras *et al.*, 2013). Il s'en distingue cependant en termes de mesure de qualité utilisée. Nous utilisons cette fois-ci une autre mesure M_{GK} en faisant mesurer la probabilité que le nombre $N_{X\bar{Y}}$ soit inférieur à celui observé $n_{X\bar{Y}}$ sous l'hypothèse H_0 d'indépendance entre X et Y. Pour ce faire, associons aux X et Y deux ensembles U et Z de I, tirés aléatoirement et indépendamment, de mêmes cardinaux respectifs que X et Y (i.e. $|U| = n_X$ et $|Z| = n_Y$). Le nombre de contre-exemples $N_{X\bar{Y}}$ attendu sous l'hypothèse H_0 est le nombre $|\phi(U \cup \bar{Z})|$ qui suit loi de Poisson de paramètre $\frac{n_X n_{\bar{Y}}}{n}$ (I. C. Lerman, 1981), et notée $N_{X\bar{Y}}$. Il s'agit d'une variable aléatoire dont $n_{X\bar{Y}}$ est une variable observée. Une telle règle d'association $X \to Y$ est alors dite *admissible* au seuil $\alpha \in]0,1]$ si la probabilité $P(N_{X\bar{Y}} \leq n_{X\bar{Y}})$ sous l'hypothèse H_0 est faible, i.e. si :

$$P(N_{X\bar{Y}} \leq n_{X\bar{Y}} \mid H_0) \leq \alpha \tag{11}$$

Il s'agit d'une p-valeur (ou degré de signification) de la règle d'association $X \to Y$. Par suite, nous considérons $\widetilde{M_{GK}}$ comme la réalisation de la variable M_{GK}. $\forall X, Y \subseteq I$, on a $M_{GK}(X \to Y) = -M_{GK}(X \to \bar{Y})$ (Feno et Totohasina, 2008), ce qui donne :

$$\widetilde{M_{GK}}(X \to Y) = -\widetilde{M_{GK}}(X \to \bar{Y}) = -\frac{n_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{\frac{n_X n_{\bar{Y}}}{n}}} \sqrt{\frac{n}{n_X n_{\bar{Y}}}} \tag{12}$$

C'est l'opposée de la contribution orientée de la cellule $X \cup \bar{Y}$ à $\frac{\chi^2}{n}$. De (11), on a :

$$P(N_{X\bar{Y}} \leq n_{X\bar{Y}} \mid H_0) = P\left(\frac{N_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{\frac{n_X n_{\bar{Y}}}{n}}} \leq \frac{n_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{\frac{n_X n_{\bar{Y}}}{n}}}\right)$$

$$= \left(\frac{N_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{n_X n_{\bar{Y}}}} \frac{n}{\sqrt{n_X n_{\bar{Y}}}} \leq \widetilde{M_{GK}}(X \to \bar{Y})\right)$$

Sous la condition peu contraignante $\frac{n_X n_{\bar{Y}}}{n} \geq 3$, la loi de Poisson peut être approximée par une loi normale. Notant F la fonction de répartition de la loi normale standardisée $\mathcal{N}(0,1)$, on a $\frac{N_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{\frac{n_X n_{\bar{Y}}}{n}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ si n grand, soit

$\frac{N_{X\bar{Y}} - \frac{n_X n_{\bar{Y}}}{n}}{\sqrt{n_X n_{\bar{Y}}}} \frac{n}{\sqrt{n_X n_{\bar{Y}}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$, d'où :

$$P(N_{X\bar{Y}} \leq n_{X\bar{Y}} \mid H_0) \simeq \int_{-\infty}^{\widetilde{M_{GK}}(X \to \bar{Y})} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = F\left(\widetilde{M_{GK}}(X \to \bar{Y})\right) \tag{13}$$

Notre modèle de sélection d'une règle d'association significative devient :

Définition 10 (Règle d'association valide). Soient $X, Y \subseteq I$. Une règle d'association $X \rightarrow Y$ est dite valide au seuil $\alpha \in]0,1]$, notée aussi $(1-\alpha)$-valide, si et seulement si :

$$\pi_{X,Y} = 1 - F\left(-\widetilde{M_{GK}}(X \rightarrow \bar{Y})\right) \geq 1 - \alpha \quad (14)$$

Lors de la phase de sélection des règles valides, de nombreuses règles d'association inintéressantes (UAR), qui perturbent la qualité des graphes, peuvent être générées. Ainsi, nous proposons une stratégie efficace permettant d'élaguer ces UARs (§3.2.2).

3.2.2 Élagage des règles d'association inintéressantes (UAR)

A notre connaissance, aucune technique d'élagage des règles d'association inintéressantes n'est trouvée dans (Bruno *et al.*, 2007 ; Cheikh *et al.,* 2008 ; Ben Yahia et Mouakher, 2011 ; Gras *et al.,* 2013 ; Couturier et Ghanem, 2015 ; Gras *et al.,* 2015 ; Hahsler, 2017 ; Delacroix *et al*., 2017 ; Carlos *et al.,* 2019). Ainsi, nous proposons deux critères (Définition 9) permettant de définir une règle d'association inintéressante.

Définition 11 (Règles inintéressantes). Soient $X, Y \subseteq I$ tels que $X \cap Y = \emptyset$. Une règle $X \rightarrow Y$ est dite inintéressante si (i) $P(Y'|X') = P(Y')$) ou (ii) $P(Y'|X') < P(Y')$.

Le premier critère permet de supprimer les UARs dont le conséquent Y est statistiquement indépendant de la prémisse X. Le deuxième critère consiste à élaguer les UARs dont les deux motifs X et Y sont négativement dépendants. Dans ce cas, les règles d'association du type $X \rightarrow \bar{Y}$ pourraient être intéressantes pour l'utilisateur.

Afin de supprimer ce type UAR, nous définissons une métrique, $\Delta_{X,Y} = P(Y'|X') - P(Y')$, qui représente la distance de la fréquence de $Y|X$ et celle de Y. En effet, une métrique $\Delta_{X,Y}$ positive indique une dépendance positive entre X et Y, et exprime que la règle $X \rightarrow Y$ est pertinente. Une métrique $\Delta_{X,Y}$ zéro indique une indépendance statistique entre X et Y, et signifie que $X \rightarrow Y$ n'est pas non plus pertinente. Une métrique $\Delta_{X,Y}$ négative exprime une dépendance négative entre X et Y, et décrit que la règle $X \rightarrow Y$ n'est pas pertinente. Il en résulte que la métrique $\Delta_{X,Y}$ est un témoin efficace permettant de localiser les règles inintéressantes définies dans la Définition 11. C'est un indicateur permettant de filtrer les règles candidates en ne conservant que celles potentiellement intéressantes, règles pour lesquelles $\Delta_{X,Y} > 0$ (i. e. $P(Y'|X') > P(Y')$). Cette technique sera utilisée le long de ce papier afin de minimiser non seulement l'espace de recherche mais aussi d'accélérer le temps d'exécution. Comme nous nous intéressons à l'élagage systématique des règles inintéressantes, nous utilisons ensuite la mesure M_{GK}. Le choix de cette mesure M_{GK} est motivé par son pouvoir discriminant en termes de sélection d'une règle valide dans une base de données. Étant donné sa définition, M_{GK} (Déf. 7) construit en fonction de la métrique $\Delta_{X,Y}$. Elle garantit donc l'élagage systématique des UARs de la Définition 11. En effet, si $\Delta_{X,Y} < 0 \Rightarrow M_{GK}(X \rightarrow Y) < 0$ exprime que $X \rightarrow Y$ n'est pas pertinente. Si $\Delta_{X,Y} = 0$, alors $M_{GK}(X \rightarrow Y) = 0$ montre que $X \rightarrow Y$ est aussi non-pertinente. Dans ce cas, aucune règle de la base de données ne pourrait être pertinente. Enfin, si $\Delta_{X,Y} > 0$, alors $M_{GK}(X \rightarrow Y) > 0$ confirme que $X \rightarrow Y$ est pertinente. Ces optimisations seront illustrées dans le tableau 2 ci-après. Ce dernier illustre non seulement le pouvoir discriminant de M_{GK} mais aussi la perte du pouvoir discriminant des confiance, intensité d'implication, et implifiance à la lecture d'une règle valide. Dans ce tableau 2, les quatre premières colonnes correspondent aux informations des A et B, les quatre dernières indiquent celles du café et du thé.

	A	$\bar{A}$	Somme		café	$\overline{café}$	Somme
B	72	8	90	thé	20	5	25
$\bar{B}$	8	2	10	$\overline{thé}$	70	5	75
Somme	80	10	100	Somme	90	10	100

Tableau 3– Tableau de contingence des (*A,B*) et (*café, thé*)

Ces informations peuvent être utilisées afin d'évaluer les associations $A \rightarrow B$ et *thé→café*. Pour le couple *(A,B),* on trouve que la confiance de $A \rightarrow B$ est 0.9, et son implifiance Φ est 0.4. Ces valeurs raisonnablement élevées nous invitent à penser que la règle $A \rightarrow B$ est pertinente. Cependant, $\Delta_{A,B} = 0.9 - 0.9 = 0 \Rightarrow M_{GK}(A \rightarrow B) = 0$ signifie bien que B est indépendant de *A*. Donc, la règle $A \rightarrow B$, jugée valide par la confiance et l'implifiance, est loin d'être intéressante, donc non valide selon M_{GK}. Cela met en relief le pouvoir discriminant de la mesure M_{GK} en termes de sélection d'une règle d'association valide. Ce n'est pas le cas pour la confiance et l'implifiance, elles restent encore significatives (voire proche de 1, cas de la confiance) même s'il n'y a pas de dépendance statistique entre la prémisse et le conséquent de la règle. Elles ne sont pas capables de distinguer les règles inintéressantes définies dans la Définition 11.

Pour le couple *(thé, café),* on s'ensuit que la confiance de la règle $thé \rightarrow café$ est égale à 0.8, son implifiance Φ est égale à 0.48, et son intensité d'implication φ est égale à 0.99. Similairement au couple *(A,B)*, ces valeurs raisonnablement élevées nous invitent à penser aussi que *thé* favorise *café*, i.e. la règle $thé \rightarrow café$ est pertinente. Cependant, $\Delta_{thé,café} = 0.8 - 0.9 < 0 \Rightarrow M_{GK}(thé \rightarrow café) < 0$ signifie que *thé* défavorise *café*, ce qui exprime que la règle $thé \rightarrow café$ est loin d'être intéressante, donc non valide selon M_{GK}. Cela garantit encore le pouvoir discriminant de la mesure M_{GK} en termes de sélection des règles d'association valides dans une base de données. Ce n'est pas le cas pour la confiance, l'intensité d'implication et l'implifiance, elles restent encore significatives même si la prémisse et le conséquent sont négativement dépendants. Une situation particulièrement alarmante pour l'intensité d'implication, elle reste encore très collée à 1 (φ=0.99), alors que thé est négativement dépendant du café.

Très souvent, le temps de réponse de construction d'un graphe implicatif n'étant pas toujours meilleur lorsque les données sont denses. Ainsi, nous proposons une stratégie efficace pour l'espace de recherche des règles plus significatives (Section. 3.2.3).

3.2.3 Élagage de l'espace de recherche des règles valides

A notre connaissance, aucune technique pour limiter l'espace de recherche n'est utilisée dans les approches (Bruno *et al.*, 2007 ; Cheikh *et al.,* 2008 ; Gras *et al.,* 2013 ; Gras *et al.,* 2015 ; Hahsler, 2017 ; Carlos *et al.,* 2019). Alors qu'il est possible de restreindre cet espace de recherche en partitionnant en 2 l'ensemble ($X \rightarrow Y, Y \rightarrow X, \bar{X} \rightarrow \bar{Y}, \bar{Y} \rightarrow \bar{X}, X \rightarrow \bar{Y}, \bar{X} \rightarrow Y, Y \rightarrow \bar{X}$ et $\bar{Y} \rightarrow X$) des 8 règles candidates.

Nous expliquons ensuite cette restriction. Pour tous $X, Y \subseteq I$, on obtient $M_{GK}(X \rightarrow Y) = -M_{GK}(X \rightarrow \bar{Y})$ (Feno et Totohasina, 2008). Il en résulte que si la règle $X \rightarrow Y$ est intéressante, alors la règle négative $X \rightarrow \bar{Y}$ ne pourra pas être intéressante. Cette dichotomie s'explique par le fait que deux règles contraires sont contra-variantes, puisque les exemples de l'une sont les contre-exemples de l'autre, et réciproquement. Si l'une possède plus d'exemples et donc moins de contre-exemples qu'à l'indépendance, alors l'autre possède moins d'exemples et plus de contre-exemples qu'à l'indépendance, et

réciproquement. Dans les deux cas, on aura toujours à considérer une dépendance positive entre deux motifs de la base comme signalé dans la section ci-dessus. Dans (Bemarisika et Totohasina, 2020), nous avons démontré que si $\Delta_{X,Y} > 0$, ce sont alors les règles $X \rightarrow Y$, $Y \rightarrow X$, $\bar{X} \rightarrow \bar{Y}$ et $\bar{Y} \rightarrow \bar{X}$ qui seront étudiées. Tandis que si $\Delta_{X,Y} < 0$, ce sont alors les quatre règles contraires $X \rightarrow \bar{Y}$, $\bar{X} \rightarrow Y$, $Y \rightarrow \bar{X}$ et $\bar{Y} \rightarrow X$ qui seront étudiées. Ainsi, nous obtenons deux classes, à savoir classe des règles $X \rightarrow Y$, $Y \rightarrow X$, $\bar{X} \rightarrow \bar{Y}$ et $\bar{Y} \rightarrow \bar{X}$, notée ζ_1, et classe des $X \rightarrow \bar{Y}$, $\bar{X} \rightarrow Y$, $Y \rightarrow \bar{X}$ et $\bar{Y} \rightarrow X$, notée ζ_2.

Dans (Bemarisika et Totohasina, 2020), nous avons démontré que toutes les règles de classe ζ_1 peuvent être dérivées de $X \rightarrow Y$, alors que toutes les règles de classe ζ_2 sont dérivables de la règle $X \rightarrow \bar{Y}$. Grâce à ces résultats, nous n'étudions que deux types de règles d'association (soit une règle par classe), à savoir $X \rightarrow Y$ et $X \rightarrow \bar{Y}$, ce qui donne 100(8-2)/8 = 75% de réduction de l'espace de recherche des règles valides. Grâce à ces résultats, nous ne considérons dans la suite que ces deux types $X \rightarrow Y$ et $X \rightarrow \bar{Y}$.

Comme dans les règles inintéressantes, plusieurs règles redondantes peuvent être aussi enregistrées. Ainsi, nous proposons une stratégie efficace (Section 3.2.4) permettant d'élaguer ces règles redondantes dans les résultats finaux.

3.2.4 Élagage des règles d'association redondantes

A notre connaissance, aucune technique d'élagage des règles redondantes n'est trouvée dans (Bruno *et al.*, 2007 ; Cheikh *et al.,* 2008 ; Mouakher et Ben Yahia, 2011 ; Gras *et al.,* 2013 ; Gras *et al.,* 2015 ; Hahsler, 2017 ; Carlos *et al.,* 2019). Ainsi, nous proposons deux critères (cf. Définition 10) permettant de définir une règle redondante.

Définition 10 (Règle d'association redondante). Une règle d'association $r_1: X_1 \rightarrow Y_1$ est dite redondante s'il existe une autre règle d'association $r_2: X_2 \rightarrow Y_2$ telle que :

i. $X_2 \subseteq X_1$ et $Y_2 \supseteq Y_1$;

ii. $supp(r_1) = supp(r_2)$ et $M_{GK}(r_1) = M_{GK}(r_2)$.

Le premier critère permet de réduire la redondance, car il ne sélectionne que des règles informatives (règles n'ayant pas de règles génératrices). Une règle $X \rightarrow Y$ est dite informative si $X \in \mathcal{G}_{\gamma(X)}$ et $Y \in \mathcal{FC}$. Elle est génératrice s'il n'existe pas de $\tilde{X} \rightarrow \tilde{Y}$ tels que $\tilde{X} \subseteq X$ et $Y \subseteq \tilde{Y}$. Le deuxième critère permet de garantir la qualité de la règle.

Afin d'élaguer les redondances, nous exploitons le concept de base des règles. Une base des règles est un ensemble minimal qui ne contient pas les redondances auxquelles on peut dériver les autres règles valides. Le concept de base de règles a été développé par plusieurs travaux (Pasquier *et al.,* 2005 ; Feno *et al.,* 2006 ; Pasquier, 2009 ; Xu *et al.,* 2011). En fait, Feno *et al.* (2006) étendent Pasquier *et al.* (2005), et proposent quatre bases : *Base pour les règles d'association positives exactes, Base pour les règles d'association positives approximatives, Base pour les règles d'association négatives exactes (basée sur le concept de bordure positive de Mannila et Toivonen, 1997), et Base pour les règles négatives approximatives*. Toutefois, cette approche n'est pas très appropriée à la modélisation des règles d'association informatives, car elle sélectionne les prémisses à partir d'un ensemble non-minimal qui n'est pas conforme à la prémisse minimale. Pasquier (2009) quant à lui propose deux bases : *Base générique pour les règles exactes (BGE),* et *Base générique pour les règles approximatives (BGA).* Cependant, cette approche n'est pas suffisante et non optimale, car elle n'extrait que des

règles d'association positives dont plusieurs sont inintéressantes à cause de l'utilisation de la mesure confiance (Section 3.2.1). Xu *et al.* (2011) étendent aussi Pasquier (2005), et proposent deux bases : *Reliable Approximate Basis* et *Reliable Exact Basis*, en utilisant la mesure *CF (C*ertainty factor). Comme dans (Pasquier, 2009), l'approche de Xu *et al.* (2011) est aussi incomplète, elle ne contient pas aussi les règles négatives.

Ainsi, nous proposons trois nouvelles bases : CBA^+ (Déf. 12), CBE^- (Déf. 13), et CBA^- (cf. Déf. 14). Par ailleurs, nous définissons une base pour les règles positives exactes, appelée CBE^+ (cf. Déf. 11), qui étend Pasquier (2009). Nous montrons ensuite que ces bases concises sont non redondantes (Théorèmes 5, 6, 7 et 8).

Définition 11 (*CBE*$^+$). Soit $\mathcal{FC}$ l'ensemble des fermés. Pour chaque fermé $\mathcal{C}$ de $\mathcal{FC}$, $\mathcal{G}_\mathcal{C}$ désigne l'ensemble de ses générateurs, la base CBE^+est formellement définie par :

$$CBE^+ = \{G \to \mathcal{C}\backslash G | G \in \mathcal{G}_\mathcal{C}\ , \mathcal{C} \in \mathcal{FC}\ , G \neq \mathcal{C}\} \quad (15)$$

De le Tableau 2, on a trouvé que $[AC] = \{A, AC\}$ conduit à étudier la règle $A \to C$. En effet, $P(C|A) = \frac{|\phi(A \cup C)|}{|\phi(A)|} = \frac{|\phi(A)|}{|\phi(A)|} = 1 \Leftrightarrow M_{GK}(A \to C) = 1$, d'où $A \to C \in CBE^+$.

Théorème 5. (i) Toutes règles positives exactes sont dérivables de la base CBE^+. (ii) Toutes règles de la base CBE^+ sont des règles positives exactes non-redondantes.

Preuve. (i) Soit $r_1 : X_1 \to Y_1 \backslash X_1$ une règle positive exacte entre deux motifs fréquents X_1 et Y_1 tels que $X_1 \subset Y_1 \subseteq \gamma(X_1)$. Puisque $M_{GK}(r_1) = 1$ (par définition d'une règle exacte), on a $supp(X_1) = supp(Y_1)$. D'après la propriété de la connexion de Galois, nous avons $supp(\gamma(X_1)) = supp(\gamma(Y_1)) \Rightarrow \gamma(X_1) = \gamma(Y_1) = \mathcal{C}$, où $\mathcal{C}$ est un fermé fréquent (i.e. $\mathcal{C} \in \mathcal{FC}$), et évidemment, $\exists\ r_2 : G \to \mathcal{C}\backslash G \in CBE^+$, où G est générateur de $\mathcal{C}$ tel que $G \subseteq X_1$ et $G \subseteq Y_1$. Montrons maintenant que r_1 peut être dérivée de r_2. De $\gamma(X_1) = \gamma(Y_1) = \mathcal{C}$, on a $supp(r_1) = supp\big(\gamma(X_1)\big) = supp(\gamma(Y_1)) = supp(\mathcal{C}) = supp(r_2)$. Comme $supp(r_1) = supp(r_2)$, on a, $M_{GK}(r_1) = M_{GK}(r_2)$ (Cf. Lemme 2). Cela prouve que r_1 peut être dérivée de r_2, et elle est redondante par rapport à r_2.

(ii) Soit $r_2 : G \to \mathcal{C}\backslash G \in CBE^+$. Selon (15), on a $G \in \mathcal{G}_\mathcal{C}$ *et* $\mathcal{C} \in \mathcal{FC}$. Démontrons qu'il n'existe aucune règle $r_3 : X_3 \to Y_3 \backslash X_3$ telle que $supp(r_3) = supp(r_2)$, $M_{GK}(r_3) = M_{GK}(r_2)$, $X_3 \subseteq G$ et $\mathcal{C} \subseteq Y_3$. Si $X_3 \subseteq G$, on a alors (cf. Définition 3) $\gamma(X_3) \subseteq \gamma(G) = \mathcal{C}$ $\Rightarrow X_3 \notin \mathcal{G}_\mathcal{C}$ et $r_3 \notin CBE^+$. Si $\mathcal{C} \subseteq X_3$, on a $\mathcal{C} = \gamma(X_3) = \gamma(G) \subseteq Y_3$. On en déduit que $G \notin \mathcal{G}_{Y3}$ (D'après Définition 3), et concluons que $r_3 \notin CBE^+$, autrement dit, r_2 est non redondante. Cela prouve que CBE^+ est une base de règles non-redondantes.

Définition 12 (C*BA*$^+$). Soit $\mathcal{FC}$ un ensemble des fermés. Pour tout $\mathcal{C} \in \mathcal{FC}$, soit $\mathcal{G}_\mathcal{C}$ l'ensemble de générateurs du fermé $\mathcal{C}$. Prenons un réel α tel que $0 < \alpha \leq 1$, on a :

$$CBA^+ = \left\{G \to \mathcal{C}\backslash G \middle| (G, \mathcal{C}) \in \mathcal{G}_{\gamma(G)} \times \mathcal{FC}, \gamma(G) \subset \mathcal{C}, \Delta_{\gamma(G),\mathcal{C}} > 0, \pi_{\boldsymbol{G},\mathcal{C}} \geq 1 - \alpha\right\} \quad (16)$$

Dans le Tableau 2, on a $AC \subset ABCE$. On obtient, si *minconf=0.6,* $P(\{BCE\}'|A') = \frac{2}{3} = 0.7 > 0.6$ exprime que $A \to BCE$ est valide selon Pasquier (2009). Toutefois, $\Delta_{A,BCE} = \frac{2}{3} - \frac{5}{6} < 0$ exprime que $A \to BCE$ n'est pas non plus intéressante, donc $A \to BCE \notin CBA^+$. Cela prouve que CBA^+ est plus concise que *BGA* de (Pasquier, 2009).

Lemme 2 (Bemarisika et Totohasina, 2020). Pour tous $X, Y, T, Z \subseteq I$ tels que X favorise Y, Z favorise T, $X \cap Y = \emptyset$ et $Z \cap T = \emptyset$. Si $X \subset Z \subseteq \gamma(X)$ et $T \subset Y \subseteq \gamma(T)$, on obtient alors $supp(X \cup Y) = supp(Z \cup T)$ et $M_{GK}(X \to Y) = M_{GK}(Z \to T)$.

Le Lemme 2 est très central dans la construction du graphe implicatif, il signifie que lorsque l'arc (X, Y) est significatif, alors l'arc (Z, T) est redondant du graphe implicatif.

Théorème 6. (i) Toutes règles positives approximatives peuvent être dérivés de CBA^+. (ii) Toutes règles de CBA^+sont des règles positives approximatives non-redondantes.

Preuve. (i) Soit $r_1: X_1 \rightarrow Y_1 \backslash X_1$ une règle positive approximative entre deux fréquents X_1 et Y_1 tels que $X_1 \subset Y_1$. Puisque $M_{GK}(r_1) < 1$ (par définition d'une règle approximative), on a $\gamma(X_1) \subset \gamma(Y_1)$. Pour tous deux fréquents $X_1 et\, Y_1$, considérons deux générateurs G_1et G_2 tels que $G_1 \subset X_1 \subseteq \gamma(X_1) = \gamma(G_1)$ et $G_2 \subset Y_1 \subseteq \gamma(Y_1) = \gamma(G_2)$. Puisque $X_1 \subset Y_1$, nous avons $X_1 \subseteq \gamma(G_1) \subset Y_1 \subseteq \gamma(G_2)$; et évidemment il existe une règle $r_2: G_1 \rightarrow \gamma(G_2) \backslash G_1 \in CBA^+$. Montrons ensuite que r_1 peut être dérivée r_2. Comme $G_1 \subset X_1 \subseteq \gamma(X_1) = \gamma(G_1)$ et $G_2 \subset Y_1 \subseteq \gamma(Y_1) = \gamma(G_2)$, on a $supp(G_1) = supp(X_1)$ et $supp(G_2) = supp(Y_1) = supp(\gamma(G_2)) \Rightarrow supp(r_1) = supp(r_2)$ et $M_{GK}(r_1) = M_{GK}(r_2)$ (Cf. Lemme 2). Cela prouve que r_1 peut être dérivée de r_2, et elle est redondante par rapport à r_2.

(ii) Soit $r_2: G \rightarrow C \backslash G \in CBA^+$. Selon (16), on a $G \in \mathcal{G}_{\tilde{C}}\, et\, C \in \mathcal{FC}$ tels que $C \subset \tilde{C}$. Montrons qu'il n'existe aucune règle $r_3: X_3 \rightarrow Y_3 \backslash X_3$ telle que $supp(r_3) = supp(r_2)$, $M_{GK}(r_3) = M_{GK}(r_2)$, $X_3 \subseteq G$ et $C \subseteq Y_3$. Si $X_3 \subseteq G$, on a alors, par définition d'un générateur, $\gamma(X_3) \subseteq \gamma(G) = \tilde{C} \Rightarrow X_3 \notin \mathcal{G}_{\tilde{C}}$ et conclut que $r_3 \notin CBA^+$. Autrement dit, r_2 est non redondante. Cela prouve que CBA^+ est une base de règles non redondantes.

Définition 13 (CBE^-). Soit $\mathcal{MFC}$ l'ensemble des maximaux fréquents et, pour chaque maximal $M \in \mathcal{MFC}$, $\mathcal{G}_M$ désigne l'ensemble des générateurs du maximal M, on a :

$$CBE^- = \{G \rightarrow \bar{Y} \backslash G | G \in \mathcal{G}_M, M \in \mathcal{MFC}, Y \in \bar{\mathcal{F}}_{\min}\} \qquad (17)$$

Dans le Tableau 2, on a trouvé que $ABCE$ est maximal fréquent dans la base de données, i.e. $[ABCE] = \{AB, AE, ABCE\}$ et le minimal non fréquent associé est $I \backslash ABCE = D$. Cela conduit à étudier les règles d'association candidates $AB \rightarrow \overline{D}$ et $AE \rightarrow \overline{D}$. En effet,

$$P\left(\overline{D'} | \{AB\}'\right) = P\left(\overline{D'} | \{AE\}'\right) = \frac{2/6}{2/6} = 1 \Rightarrow \{AB \rightarrow \overline{D}, AE \rightarrow \overline{D}\} \in CBE^-.$$

Théorème 7. (i) Toutes règles négatives exactes peuvent être dérivées de la base CBE^-. (ii) Toutes règles de cette base CBE^-sont des règles négatives exactes non-redondantes.

Preuve. (i) Soit $r_1: X_1 \rightarrow \bar{Y}_1 \backslash X_1$ une règle négative exacte entre deux motifs fréquents X_1 et $\bar{Y}_1$ tels que $X_1 \subset \bar{Y}_1 \subseteq M$ $(i.e. Y_1 \in \bar{\mathcal{F}}_{min})$, où M est un maximal fréquent. Puisque $M_{GK}(r_1) = 1$ (définition d'une règle exacte), on a d'après Corollaire 1 $supp(X_1 \cup Y_1) = 0$ implique $supp(X_1 \cup \bar{Y}_1) = supp(X_1) = supp(\bar{Y}_1)$. D'après la propriété de Galois, on obtient $supp(\gamma(X_1 \cup \bar{Y}_1)) = supp(\gamma(X_1)) = supp(\gamma(\bar{Y}_1)) \Rightarrow \gamma(X_1 \cup \bar{Y}_1) = \gamma(X_1) = \gamma(\bar{Y}_1) = M$ (a), et évidemment, $\exists\, r_2: G \rightarrow \bar{Y} \backslash G \in CBE^-$, où G est générateur de M tel que $G \subseteq X_1$, $G \subseteq \bar{Y}_1$, et $G \subseteq \bar{Y}$ (selon l'équation 17). Montrons maintenant que r_1 peut être dérivée de r_2. Comme $r_2: G \rightarrow \bar{Y} \backslash G \in CBE^-$, on a (cf. Corollaire 1) $supp(G \cup \bar{Y}) = supp(G) \Rightarrow supp(\gamma(G \cup \bar{Y})) = supp\left(\gamma(G)\right) = supp\left(\gamma(\bar{Y})\right) \Rightarrow \gamma(G \cup \bar{Y}) = \gamma(G) = \gamma(\bar{Y}) = M$ (a'). En identifiant les relations (a) et (a'), nous avons $\gamma(G \cup \bar{Y}) = \gamma(X_1 \cup \bar{Y}_1) \Rightarrow supp(G \cup \bar{Y}) = supp(X_1 \cup \bar{Y}_1)$, i.e. $supp(r_1) = supp(r_2)$. Puisque $G \subseteq$

$X_1 \subset \overline{Y_1} \subset \overline{Y} \subseteq \gamma(G) = M$, on a $supp(G) = supp(X_1) = supp(\overline{Y_1}) = supp(\overline{Y}) = supp(M) \Rightarrow M_{GK}(r_1) = M_{GK}(r_2)$. Cela prouve que r_1 peut être dérivée de r_2, et elle est une règle redondante par rapport à r_2.

(ii) Soit $r_2: G \to \bar{y} \backslash G \in IBE^-$. Selon (17), on a $G \in \mathcal{G}_M$ *et* $Y \in \overline{\mathcal{F}}_{min}$. Montrons qu'il n'existe aucune règle $r_3: X_3 \to \overline{Y_3} \backslash X_3$ telle que $supp(r_3) = supp(r_2)$, $M_{GK}(r_3) = M_{GK}(r_2)$, $X_3 \subseteq G$ et $\overline{Y} \subseteq \overline{Y_3}$. Si $X_3 \subseteq G$, on obtient $\gamma(X_3) \subseteq \gamma(G) = \overline{Y} \Rightarrow X_3 \notin \mathcal{G}_M$ et $r_3 \notin CBE^-$. Si $\overline{Y} \subseteq \overline{Y_3}$, on obtient $\gamma(G) \subset \gamma(\bar{y}) \subseteq \gamma(\overline{Y_3}) = M$. D'après la définition 3, on a $G \notin \mathcal{G}_{\overline{Y_3}} \Rightarrow r_3 \notin CBE^-$, autrement dit r_2 est une règle non redondante. Cela prouve que cette base CBE^-est une base de règles d'association non redondantes.

Définition 14 (CBA^-). Soit $\mathcal{FC}$ un ensemble des fermés. Soit $\mathcal{G}_{\mathcal{C}}$ l'ensemble de générateurs du fermé $\mathcal{C}$ de $\mathcal{FC}$. Considérons un seuil α tel que $0 < \alpha \leq 1$, on a :

$$CBA^- = \left\{ G \to \bar{g} \backslash G \middle| (G, g) \in \mathcal{G}_{\gamma(G)} \times \mathcal{G}_{\gamma(g)}, \gamma(G) \subsetneq \gamma(g), \Delta_{\gamma(G),\gamma(g)} < 0, \pi_{\boldsymbol{G},\bar{g}} \geq 1 - \alpha \right\} \quad (18)$$

Dans le Tableau 2, on a trouvé que $[AC] = \{A, AC\}$ et $[BE] = \{B, E, BE\}$. Comme $AC \subsetneq BE$ et $\Delta_{AC,BE} < 0$ implique AC défavorise BE (i.e. AC favorise $\overline{BE}$). Cela conduit à étudier $A \to \bar{B}$ et $A \to \bar{E}$ en prenant α=0.05. Ici, $n=6$, $n_A = 3, n_B = n_E = 5, n_{AB} = n_{AE} = 2$ conduisent à $\pi_{A\bar{B}} = \pi_{A\bar{E}} = 42\% < 95\% \Rightarrow \{A \to \bar{B}, A \to \bar{E}\} \notin CBA^-$ à α=5%.

Lemme 3 (Bemarisika et Totohasina, 2020). Pour tous $X, Y, T, Z \subseteq I$ tels que X défavorise Y, Z défavorise T, $X \cap Y = \emptyset$ et $Z \cap T = \emptyset$. Si $X \subset Z \subseteq \gamma(X)$ et $Y \subset T \subseteq \gamma(Y)$, on obtient alors $supp(X \cup Y) = supp(Z \cup T)$ et $M_{GK}(X \to \bar{Y}) = M_{GK}(Z \to \bar{T})$.

Le Lemme 3 est aussi central dans la construction du graphe implicatif, il signifie que lorsque l'arc $(X, \bar{Y})$ est significatif, alors l'arc $(Z, \bar{T})$ est redondant du graphe implicatif.

Théorème 8. (i) Toutes règles négatives approximatives peuvent être dérivés de CBA^-. (ii) Toutes règles de cette base CBA^-sont des règles non-redondantes.

Preuve. (**i**) Soit $r_1: X_1 \to \overline{Y_1} \backslash X_1$ une règle négative approximative entre deux motifs fréquents X_1 et $\overline{Y_1}$ tels que $X_1 \subset \overline{Y_1}$. Puisque $M_{GK}(r_1) < 1$, on a $\gamma(X_1) \subset \gamma(\overline{Y_1})$. Pour deux fréquents X_1 et $\overline{Y_1}$, il existe deux générateurs G_1et G_2 tels que $G_1 \subset X_1 \subseteq \gamma(G_1) = \gamma(X_1)$ et $G_2 \subset Y_1 \subseteq \gamma(G_2) = \gamma(Y_1)$. Comme $X_1 \subset \overline{Y_1}$, on a $X_1 \subseteq \gamma(G_1) \subset \overline{Y_1} \subset \overline{G_2} \subseteq \gamma(\overline{Y_1}) = \gamma(\overline{G_2})$, et évidemment il existe une règle $r_2: G_1 \to \overline{G_2} \backslash G_1 \in CBE^-$. Montrons maintenant que r_1 est dérivable de r_2. Puisque $G_1 \subset X_1 \subseteq \gamma(G_1) \subset \overline{Y_1} \subset \overline{G_2} \subseteq \gamma(\overline{Y_1})$, on obtient $supp(G_1) = supp(X_1) = supp(\gamma(G_1)) = supp(\overline{Y_1}) = supp(\overline{G_2}) \Rightarrow |\phi(X_1 \cup \overline{Y_1})| = |\phi(G_1 \cup \overline{G_2})| \Rightarrow supp(r_1) = supp(r_2)$ et $M_{GK}(r_1) = M_{GK}(r_2)$ (Cf. Lemme 3). Cela prouve que r_1 peut être dérivée de r_2, et elle est redondante par rapport à r_2.

(**ii**) Soit $r_2: G \to \bar{g} \backslash G \in CBA^-$, où $G \in \mathcal{G}_{\mathcal{C}}$ *et* $g \in \mathcal{G}_{\tilde{\mathcal{C}}}$ et $\mathcal{C} \subsetneq \tilde{\mathcal{C}}$. Montrons qu'il n'existe aucune règle $r_3: X_3 \to \overline{Y_3} \backslash X_3$ telle que $supp(r_3) = supp(r_2)$, $M_{GK}(r_3) = M_{GK}(r_2)$, $X_3 \subseteq G$ et $Y_3 \subseteq g$ (i.e. $\bar{g} \subset \overline{Y_3}$). Si $X_3 \subseteq G$, on obtient $\gamma(X_3) \subseteq \gamma(G) = \tilde{\mathcal{C}} \Rightarrow X_3 \notin \mathcal{G}_{\tilde{\mathcal{C}}} \Rightarrow r_3 \notin CBA^-$. Si $X_3 \subseteq G$, nous avons $supp(X_3) > supp(G) \Rightarrow M_{GK}(r_3) < M_{GK}(r_2)$. Si $\bar{g} \subset \overline{Y_3}$, on a $\gamma(\bar{g}) \subset \gamma(\overline{Y_3}) \Rightarrow supp(\bar{g}) > supp(\overline{Y_3})$ et concluons que $M_{GK}(r_3) < M_{GK}(r_2)$. Comme $r_3 \notin CBA^-$, alors r_2 est une règle non redondante. Cela prouve que CBA^- est une base des règles négatives approximatives non redondantes.

Ces optimisations seront synthétisées dans l'algorithme CONCISE. Il est composé de deux algorithmes (Algorithme 3 et Algorithme 4). L'Algorithme 3 prend en entrée les ensembles $\mathcal{G}_C$, $\mathcal{FC}$, $\mathcal{MFC}$ et $\overline{\mathcal{F}}_{\min}$ ainsi que les seuils *minsup* et α. Il retourne les bases concises pour les règles d'association non-redondantes d'une base de données. En effet,

la liste $\mathcal{CB}$ des bases concises pour les règles non redondantes est initialisée à vide (ligne 1). L'algorithme 3 procède de deux parties. Tout d'abord, il génère les bases concises CBE^+ et CBA^+ (lignes 2-14). En effet, pour tout fermé $\mathcal{C}$ (ligne 2) et tout générateur G de ce fermé $\mathcal{C}$ (ligne 3) qui représentent respectivement la prémisse et le conséquent de la règle, l'algorithme 3 vérifie si ces deux motifs, à l'aide de la métrique

Algorithm 3 GENERATING CONCISE BASE OF NON-REDUNDANT RULES

Require: $\mathcal{G}_{\gamma(.)}$, $\mathcal{FC}$, $\mathcal{MFC}$, $\overline{\mathcal{F}}_{min}$, $minsup$ and α.
Ensure: $\mathcal{CB}$, A Concise Base of Non-Redundant PNARs.

1: $\mathcal{CB} = \emptyset$;
2: **for all** $(\mathcal{C} \in \mathcal{FC})$ **do**
3: **for all** $(G \in \mathcal{G}_{\mathcal{C}})$ **do**
4: **if** $(P(\mathcal{C}'|G') > P(\mathcal{C}'))$ **then**
5: **if** $(\gamma(G) = \mathcal{C})$ **then**
6: **if** $(G \neq \gamma(G)$ & $supp(G \cup \mathcal{C}) \geq minsup)$ **then**
7: $\mathcal{CB} \leftarrow \mathcal{CB} \cup \{G \rightarrow \mathcal{C} \backslash G\}$; /* CBE^+ Basis */
8: **end if**
9: **else if** $(\gamma(G) \subset \mathcal{C})$ **then**
10: **if** $(supp(G \cup \mathcal{C}) \geq minsup$ & $p_{GC} \geqslant 1 - \alpha)$ **then**
11: $\mathcal{CB} \leftarrow \mathcal{CB} \cup \{G \rightarrow \mathcal{C} \backslash G\}$; /* CBA^+ Basis */
12: **end if**
13: **end if**
14: **else if** $(P(\mathcal{C}'|G') < P(\mathcal{C}'))$ **then**
15: **for all** $(\mathcal{M} \in \mathcal{MFC})$ **do**
16: **for all** $(G \in \mathcal{G}_{\mathcal{M}})$ **do**
17: **for all** $(y \in \overline{\mathcal{F}}_{min})$ **do**
18: **if** $(supp(G \cup \overline{y}) \geq minsup)$ **then**
19: $\mathcal{CB} \leftarrow \mathcal{CB} \cup \{G \rightarrow \overline{y} \backslash G\}$; /* CBE^- Basis */
20: **end if**
21: **end for**
22: **end for**
23: **end for**
24: **for all** $(g \in \mathcal{G}_{\gamma(g)} | \gamma(G) \subsetneq \gamma(g)$ & $P(g'|G') < P(g'))$ **do**
25: **if** $(supp(G \cup \overline{g}) \geq minsup$ & $p_{G\overline{g}} \geqslant 1 - \alpha)$ **then**
26: $\mathcal{CB} \leftarrow \mathcal{CB} \cup \{G \rightarrow \overline{g} \backslash G\}$; /* Base CBA^- */
27: **end if**
28: **end for**
29: **end if**
30: **end for**
31: **end for**
32: **return** $\mathcal{CB}$

Figure 3 : Schéma de l'algorithme 3

$\Delta_{G,\mathcal{C}}$, sont positivement dépendants (ligne 4), et vérifie s'ils appartiennent à une même classe d'équivalence (ligne 5). De la ligne 6, il vérifie si le générateur G n'est pas un unique élément dans sa classe d'équivalence (i.e. $\gamma(G) \neq G$), et vérifie également si le motif union $G \cup \mathcal{C}$ est fréquent. Si c'est le cas, la règle $G \rightarrow \mathcal{C}$ est une règle exacte valide et ajoutée dans la liste $\mathcal{CB}$ (ligne 7). Si G et $\mathcal{C}$ n'appartiennent pas à une même classe d'équivalence (ligne 9), l'Algorithme 3 génère la règle d'association approximative $G \rightarrow \mathcal{C}$ telle que $\gamma(G) \subseteq \mathcal{C}$ (lignes 9-14). En effet, il vérifie si les motifs G et $\mathcal{C}$ sont fréquents et vérifie également si le poids (ou taux) $p_{G,\mathcal{C}}$ est significatif (ligne 10). Si c'est le cas, la règle $G \rightarrow \mathcal{C}$ est valide et ajouté dans la liste $\mathcal{CB}$ (ligne 11). Ceci termine la première partie de l'Algorithme 3. Dans la seconde partie, l'Algorithme 3 génère tout d'abord les règles négatives exactes (lignes 15-23). Après avoir collecté un maximal M, un générateur G du maximal M, ainsi qu'un minimal non fréquent y, l'Algorithme 3 vérifie si l'item $G \cup \bar{y}$ est fréquent (ligne 18). Si c'est le cas, la règle exacte $G \rightarrow \bar{y}$ est valide et ajoutée dans $\mathcal{CB}$

(ligne 19). Enfin, l'Algorithme 3 génère les règles négatives approximatives (lignes 24-28). Après avoir collecté un générateur $g \in \mathcal{G}_{\gamma(g)}$ tels que $\gamma(g)$ et $\gamma(G)$ sont incomparables et négativement dépendants (ligne 24), l'Algorithme 3 vérifie si le motif $G \cup \bar{g}$ est fréquent et si $\pi_{G\mathcal{C}} \geq 1 - \alpha$ (ligne 25). Si c'est le cas, la règle approximative $G \rightarrow \bar{g}$ est valide et ajoutée dans $\mathcal{CB}$ (ligne 26).

L'Algorithme 4 dérive toutes les règles valides. Il appelle la liste $\mathcal{CB}$ de l'Algo. 3.

Algorithm 4 DERIVING ALL VALID RULES

Require: $\mathcal{CB}$ *Concise Base*, and *minsup*.
Ensure: All_{RUL}, *All Valid Non-redundant PNARs*.
1: $All_{RUL} = \emptyset$;
2: **for all** $(X \rightarrow Y \backslash X \in \mathcal{CB})$ **do**
3: **if** $(supp(\overline{X} \cup \overline{Y}) \geqslant minsup)$ **then**
4: $All_{RUL} \leftarrow All_{RUL} \cup \{\overline{X} \rightarrow \overline{Y} \backslash \overline{X}\}$
5: **end if**
6: **end for**
7: **for all** $(X \rightarrow \overline{Y} \backslash Y \in \mathcal{CB})$ **do**
8: **if** $(supp(\overline{X} \cup Y) \geqslant minsup)$ **then**
9: $All_{RUL} \leftarrow All_{RUL} \cup \{\overline{X} \rightarrow Y \backslash \overline{X}\}$
10: **end if**
11: **end for**
12: **return** All_{RUL}

Figure 4 : Schéma de l'algorithme 4

La liste All_{RUL} des règles dérivées est initialisée à vide (ligne 1). L'algorithme 4 dérive tout d'abord les règles du type $\bar{X} \rightarrow \bar{Y}$ (lignes 2-6). En effet, après avoir collecté dans $\mathcal{CB}$ la règle $X \rightarrow Y$, il vérifie si le motif $\bar{X} \cup \bar{Y}$ est fréquent (lignes 2-3). Si c'est le cas, la règle $\bar{X} \rightarrow \bar{Y}$ est valide et ajoutée dans la liste All_{RUL}, car la règle $\bar{X} \rightarrow \bar{Y}$ est dérivable de $X \rightarrow Y$ (i.e. $M_{GK}(X \rightarrow Y) \leq M_{GK}(\bar{X} \rightarrow \bar{Y})$, c'est-ce que nous avons signalé dans la section 3.2.3). Enfin, l'algorithme 4 génère les règles du type $\bar{X} \rightarrow Y$ (lignes 7-11). En effet, après avoir collecté dans $\mathcal{CB}$ la règle $X \rightarrow \bar{Y}$, il vérifie si le motif $\bar{X} \cup Y$ est fréquent (lignes 7-8). Si c'est le cas, la règle $\bar{X} \rightarrow Y$ est valide et ajoutée dans All_{RUL}, car la règle $\bar{X} \rightarrow Y$ est dérivable de $X \rightarrow \bar{Y}$ (i.e. $M_{GK}(X \rightarrow \bar{Y}) \leq M_{GK}(\bar{X} \rightarrow Y)$).

3.3 Configuration topologique d'un graphe implicatif

Nous commençons par modéliser un graphe implicatif d'une base de données.

3.3.1 Un modèle de graphe implicatif

Le modèle que nous allons privilégier est classique. Pour tous motifs $X, Y \subseteq I$ d'une base de données, une règle d'association $X \rightarrow Y$ est modélisée par un arc ayant pour origine (resp. extrémité) le sommet X (resp. Y). Un arc allant du nœud X au nœud Y, est notée (X, Y). Dans ce cas, X est le prédécesseur et Y successeur qui représentent à cet effet les sommets du graphe implicatif. Notons par S l'ensemble des sommets du graphe, et par $\mathcal{E}$ l'ensemble d'arcs. Cet ensemble $\mathcal{E}$ d'arcs est alors défini par :

$$\mathcal{E} = \{(X, Y) | X, Y \subseteq S\} \qquad (19)$$

Pour chaque sommet $X \in S$, on définit $d^-(X)$ (resp. $d^+(X)$) le degré sortant de X comme étant le nombre d'arcs partant de X (resp. le degré entrant de X comme étant le

nombre d'arcs arrivant sur X). Un chemin entre deux sommets est une séquence de sommets et d'arêtes distincts reliant ces deux sommets. La longueur d'un chemin est donnée par son nombre d'arcs, i.e. $|S| - 1$. Un chemin constitué d'un seul sommet est de longueur nulle. Pour chaque paire X et Y de S, notons par $\pi_{X,Y}$, le poids de l'arc (X, Y), qui est un nombre positif qui mesure la dépendance entre les sommets X et Y du graphe. Un chemin est dit pondéré s'il est attribué à un poids sur chaque arc représentant le nombre d'occurrences de cet arc parmi toutes les occurrences dans le chemin. Ainsi, le chemin $X \to Y \to Z$ est un chemin pondéré des $\pi_{X,Y}$ et $\pi_{Y,Z}$, noté par $X \xrightarrow{\pi_{X,Y}} Y \xrightarrow{\pi_{Y,Z}} Z$, où $\pi_{X,Y}$ et $\pi_{Y,Z}$ sont respectivement poids des arcs $X \to Y$ et $Y \to Z$.

Une des difficultés liées à la configuration d'un graphe repose sur le parcours des sommets et/ou arêtes du graphe. Pour cela, nous définissons une matrice d'implication.

Définition 15 (Matrice d'implication). Soit S un ensemble de sommets de taille r (i.e. $|S| = r$). La matrice d'implication M est une matrice carrée d'ordre $r \times r$, définie par :

$$M_{ij} = \begin{cases} 1, si\ \pi_{i,j} \geq 1 - \alpha \\ 0, \qquad sinon \end{cases} \qquad (20)$$

où $\pi_{i,j}$ est le poids de l'arc (i, j), et α un seuil choisi dans l'intervalle $]0; 1]$.

En fait, la matrice M est une matrice de nombres dans $[0; 1]$. Pour tous sommets i et j de S, M_{ij} contient un tuple $\pi_{i,j}$ de taille r, où $M_{ij} = 1$ si $\pi_{i,j} \geq 1 - \alpha$, et $\pi_{i,j} = 0$ sinon pour $\alpha \in]0,1]$. Le but est d'opérer des regroupements sur les sommets du graphe afin de proposer une simplification possible (sans perte d'information) qui soit représentative des combinaisons de composantes les plus fréquentes au sein des chemins. Ces combinaisons étant révélées par un poids $\pi_{\cdot,\cdot}$ de l'arc (Eq. 14). Il s'agit d'un regroupement de sommet à sommet. Pour ce faire, nous adoptons la procédure suivante: 1) Partant d'un sommet source i, on sélectionne les successeurs immédiats dans lequel ces successeurs étant supposés ordonnés dans l'ordre croissant de leur coût. On définit, pour un $n^{\text{ème}}$ itération, $\nabla_{i,j}^{(n)}$, le coût minimum d'un chemin de i à j, donné par :

$$\nabla_{i,j}^{(n)} = \min\left\{\nabla_{i,j}^{(n-1)}, \nabla_{p(i),i}^{(n-1)} + \pi_{i,j}^{(n)}\right\} \qquad (21)$$

où $p(i)$ dénote le prédécesseur du sommet i. En particulier, $\nabla_{i,j}^{(0)} = \pi_{i,j}$. Ainsi, on choisit celui, j, dont le coût ∇_{ij} est minimum parmi les successeurs naturels de i. Rappelons qu'un sommet est dit source si son degré entrant (resp. sortant) est nul (resp. non nul), i.e. un sommet n'ayant pas des prédécesseurs mais admettent des successeurs :

$$\forall i \in S, \exists j \text{ tel que } j \neq i, \text{ on a } M_{ji} = 0 \text{ et } M_{ij} = 1. \qquad (22)$$

2) On arrête le chemin de ce sommet source lorsque les voisins (ou successeurs immédiats) sont appelés, de même que les propres successeurs de ceux-ci. 3) Partant d'un autre sommet source, on construit comme précédemment le chemin de ce nouveau sommet source. Il arrive que figurent, parmi les successeurs naturels ou de la liaison de ce nouveau sommet source, des mêmes successeurs du premier sommet d'origine.

Soit R_α une relation définie sur $S \times S$ par M_{GK} pour un seuil $\alpha \in]0; 1]$: $i\mathrm{R}_\alpha j$ si et seulement si $\pi_{i,j} \geq 1 - \alpha$. La relation définie par l'implication statistique, si elle est réflexive et acyclique, n'est pas transitive. Or, nous voulons qu'elle modélise la relation d'ordre partiel entre deux variables. Ainsi, nous étendons la relation R_α (Proposition 1).

Proposition 1 (Transitivité). Soient deux motifs disjoints i et j de I. Si $i\mathrm{R}_\alpha j$ et $j\mathrm{R}_\alpha k$, on a alors la fermeture transitive $i\mathrm{R}_\alpha k$ si et seulement si $\pi_{i,k} \geq 0{,}5$. Autrement dit, la valeur 0,5 est une valeur limite correspondant à l'indépendance entre les motifs i et k.

Cette Proposition 1 est très centrale : Si $i \to j$ et si j$\to k$, il y a *fermeture transitive* $i \to k$ si et seulement si $\pi_{i,k} \geq 0{,}5$, c'est-à-dire si la relation implicative de i sur k, qui traduit une certaine dépendance positive entre i et j, est meilleure que sa réfutation. Ainsi, nous introduisons une définition formelle (Déf. 16) d'un graphe implicatif.

Définition 16 (Graphe implicatif). Soit M une matrice d'implications, α un seuil fixé dans l'intervalle]0 ;1]. Un graphe implicatif associé à M et α est $\mathrm{G}_\alpha^M = (S, \mathcal{E})$, un couple constitué par deux ensembles de sommets S et d'arcs $\mathcal{E}$ tels que $(i,j) \in \mathcal{E}$ ssi $i\mathrm{R}_\alpha j$.

Il s'agit d'un graphe de la relation R_α définie par $i\mathrm{R}_\alpha j \Leftrightarrow \pi_{i,j} \geq 1-\alpha$, où $\alpha \in]0\,;1]$, un seuil fixé par l'analyste. En faisant varier α, il est évident que le nombre d'arcs du graphe varie dans le sens opposé, i.e. le nombre d'arc diminue (resp. augmente) si α augmente (resp. diminue). En particulier, le nombre d'arcs du graphe pour un seuil 0,05 est inférieur ou égal à ceux qui constitueraient le graphe au seuil 0,10. Relativement, des arcs du graphe peuvent apparaître ou disparaître selon les variations du seuil.

Ce modèle produit une partition en plusieurs sous-graphes, chacun étant polarisé par un sommet indépendant. Un sous-graphe implicatif est alors classique, il correspond à la réunion d'un sous-ensemble de sommets du graphe et des arcs existant entre eux.

Définition 17 (Sous-graphe implicatif). Soit M une matrice d'implications, $\alpha \in]0,1]$ un seuil. Un sous-graphe de $\mathrm{G}_\alpha^M = (S, \mathcal{E})$ est un graphe $\mathrm{Q}_\alpha^M = (\tilde{\mathrm{S}}, \mathcal{E}_{\tilde{\mathrm{S}}})$, tel que $\tilde{\mathrm{S}}$ est un sous-ensemble de S et $\mathcal{E}_{\tilde{\mathrm{S}}}$ est l'ensemble des arcs de $\mathcal{E}$ dont les sommets sont dans $\tilde{\mathrm{S}}$.

On dit que Q_α^M est le sous-graphe implicatif de G_α^M induit par l'ensemble S, et que l'ensemble $\mathcal{E}_{\tilde{\mathrm{S}}}$ est la restriction de l'ensemble $\mathcal{E}$ à l'ensemble $\tilde{\mathrm{S}}$. Un graphe partiel du graphe implicatif G_α^M est un graphe $\mathrm{H}_\alpha^M = (\mathrm{S}, \widetilde{\mathcal{E}})$ tel que $\widetilde{\mathcal{E}}$ est un sous-ensemble de S.

Une autre difficulté liée à la construction d'un graphe implicatif tient au fait que ce graphe n'est pas planaire. L'algorithme qui en permet la construction, doit le prendre en compte et, en particulier, se doit de décroiser les chemins afin d'en permettre une lisibilité acceptable. Ainsi, nous définissons le Théorème 9 permettant d'élaguer des branches de transitivité afin d'améliorer la lisibilité du graphe a priori dense.

Théorème 9. Un graphe implicatif G_α^M est (i) transitif et (ii) acyclique.

Preuve. (i) Soient $X, Y, Z \subseteq I$, t.q. X, Y et Z se favorisent 2 à 2, et $X \subseteq Y \subseteq Z$. Comme $X \subseteq Y \subseteq Z$, on a $\phi(Z) \subseteq \phi(Y) \subseteq \phi(X)$ (i.e. $P(Z') \leq P(Y') \leq P(X')$) conduit à :

$$P(Y'|X') = \frac{|\phi(X \cup Y)|}{|\phi(X)|} = \frac{|\phi(X) \cap \phi(Y)|}{|\phi(X)|} = \frac{|\phi(Y)|}{|\phi(X)|} = \frac{P(Y')}{P(X')}$$

$$et\ M_{GK}(X \to Y) = \frac{P(Y'|X') - P(Y')}{1 - P(Y')} = \frac{\frac{P(Y')}{P(X')} - P(Y')}{1 - P(Y')} = \frac{P(Y') - P(X')P(Y')}{P(X')(1 - P(Y'))} = \frac{P(Y')(1 - P(X'))}{P(X')(1 - P(Y'))}.$$

Par suite, $M_{GK}(X \to Y) \times M_{GK}(Y \to Z) = \frac{P(Y')\left(1-P(X')\right)}{P(X')\left(1-P(Y')\right)} \times \frac{P(Z')\left(1-P(Y')\right)}{P(Y')\left(1-P(Z')\right)}$

$$= \frac{P(Z')\left(1 - P(X')\right)}{P(X')\left(1 - P(Z')\right)}$$

$$= \frac{P(Z'|X') - P(Z')}{1 - P(Z')} = M_{GK}(X \to Z).$$

Plus généralement, pour tous $X_1, \ldots, X_{n-1}, X_n \subseteq I$ tels que $X_1 \subseteq \cdots \subseteq X_{n-1} \subseteq X_n$, on a :

$$M_{GK}(X_1 \to X_n) = \prod_{k=1}^{n} M_{GK}(X_{k-1} \to X_k),$$

i.e. $(X_1, X_2) \in \mathcal{E}, \ldots, (X_{n-1}, X_n) \in \mathcal{E} \Rightarrow (X_1, X_n) \in \mathcal{E}$ conclut que G_α^M est transitif.

(ii) Un graphe est acyclique si $(i,j) \in \mathcal{E}$, alors $(j,i) \notin \mathcal{E}$. Supposons qu'il existe un chemin entre i et j, alors il existe aussi un arc $(i,j) \in \mathcal{E}$. Or, $(j,i) \notin \mathcal{E}$ implique G_α^M est acyclique. Par exemple l'item i n'implique pas l'item i (ou $\bar{\imath}$).

3.3.2 Algorithme IGRAPH

IGRAPH (Algorithme 5) prend en argument une matrice d'implication M et un seuil de risque $\alpha \in]0;1]$, et retourne les graphes implicatifs G_α^M. D'un point vu technique, il parcourt de sommet en sommet l'espace de recherche. En effet, la liste $\mathcal{E}$ des arcs est

Algorithm 5 IGRAPH

Require: Une matrice de données $\mathcal{M}$, un seuil de risque α tel que $0 < \alpha \leq 1$
Ensure: Graphe implicatif $G_\alpha^{\mathcal{M}} = (S, \mathcal{E})$

```
1: E = ∅;
2: repeat
3:    for all (i ∈ M) do
4:       for all (j ∈ M) do
5:          if (M_ji = 0 ∧ M_ij = 1|∇_ij < ∇_i., ∀j ≠ i) then
6:             E ← E ∪ {(i, j)};
7:             for all (k ∈ M) do
8:                if (M_ik = M_jk = 1|∇_ik < ∇_i.||∇_jk < ∇_j.) then
9:                   E ← E ∪ {(i, j) ∪ (j, k)};
10:               end if
11:               if (M_ik = 1 ∧ M_jk = 0 | π_jk ≥ 1 − 2α) then
12:                  E ← E ∪ {(i, j) ∪ (i, k)};
13:               else if (M_ik = 0 ∧ M_jk = 1) then
14:                  if (π_ik ≥ 0.5) then
15:                     E ← E ∪ {(i, k) ∪ (j, k)};
16:                  else
17:                     E ← E ∪ {(i, j)}||{(j, k)};
18:                  end if
19:               end if
20:            end for
21:         end if
22:      end for
23:   end for
24: until (|E| = |S| − 1)
25: Return G_α^M;
```

Figure 5 : Schéma de l'algorithme 5

initialisée à vide (ligne 1). Pour chaque sommet i et j de la matrice M, l'algorithme IGRAPH vérifie si les sommets i et j sont respectivement source et successeur immédiat tel que le coût $\nabla_{i,j}$ est minimal à celui des autres sommets voisins de i (lignes 2-5). Si c'est le cas, il enchaîne l'arc (i,j) (ligne 6). Pour chaque k de M, IGRAPH vérifie si k est successeur naturel de i et j tel que k admet un poids le plus faible parmi les successeurs naturels de i ou de j (lignes 7-8). Si c'est le cas, IGRAPH enchaîne k à j, i.e. $i \to j \to k$ (ligne 9). Puis, il vérifie si l'arc (i,k) est significatif mais l'arc (j,k) ne l'est pas alors que

son poids est supérieur ou égal à $1 - 2\alpha, \forall\alpha \in]0; 1]$ (ligne 11). Si c'est le cas, il enchaîne les arcs (i, j) et (i, k) (ligne 12). Inversement, si l'arc (j, k) est significatif mais l'arc (i, k) ne l'est pas, l'algorithme IGRAPH vérifie si le poids π_{ik} est supérieur ou égal à 0.5. Si c'est le cas, il enchaîne les deux arcs (i, k) et (j, k) (ligne 15), et considère, sinon, l'un des arcs (i, j) et (j, k) (ligne 17). En répétant ce processus jusqu'à ce que les sommets (voire arcs) soient appelés (ligne 24). Autrement dit, l'algorithme IGRAPH s'arrête lorsqu'il n'y a plus des arcs qui puissent être établis.

4 Evaluations expérimentales

L'approche a été implémentée au moyen d'un nouveau prototype *rchicmgk* interface R. Il s'agit d'une extension du logiciel RCHIC (Couturier et Ghanem, 2015; Delacroix *et al.*, 2019 ; Anastasiadou, 2019, Bodin et Giovannini, 2019) dans l'environnement R en utilisant une mesure M_{GK}. L'idée est d'offrir un moyen visuel aux utilisateurs pour interagir et comprendre un système de relations entre les entités d'un ensemble des règles d'association valides. Comme dans RCHIC, il est possible de choisir une zone de travail par défaut et la faire évoluer au fil de l'utilisation. Communément, le problème de graphes implicatifs n'est pas trivial, car sa complexité croit linéairement avec le nombre des nœuds et des arcs. Ce nombre peut être conséquent, ce qui nous dispose d'un grand espace de travail pouvant être largement supérieur à la taille de l'écran. Pour cela, *rchicmgk* pourra sélectionner les nœuds qui semblent utiles. Il permet aussi désactiver temporairement les motifs désirés à l'aide d'une boite de dialogue prévu à cet effet. Ensuite, il met à jour à nouveau les graphes sans réimporter la base de données.

Il est possible de sauvegarder l'état d'un graphe restitué, c'est-à-dire la disposition des motifs, les seuils de risque (ou seuil de signification), la sélection ou non de chaque motif (voire arc). Ainsi, l'utilisateur peut reprendre un graphe qu'il avait organisé soigneusement lors d'une précédente session. De plus, il est possible de sauvegarder plusieurs états sur le même graphe et ainsi mettre en évidence différentes parties du graphe. On peut également exporter un graphe sous Word, pour le moment.

Les résultats expérimentaux rapportés dans cette section ont été obtenus sur un PC Core i3-2350M avec 2,3 GHz (4Go RAM). Nous utilisons pour cela une base de données réelles *iris*[4] qui est la donnée de l'UCI Machine Learning Repository, et décrit les typologies des plantes. Nous détaillons ci-dessous les différents graphes implicatifs obtenus de cette base *iris* en faisant varier le seuil de risque α de 0.01, 0.05, 0.10 puis 0.15. Cela nous permet de mettre en évidence une partition en 3 sous-graphes pour chacune des quatre figures obtenues. Chaque modalité conduira à la construction de sous-intervalles optimaux, c'est-à-dire la détermination de sous-intervalles optimisant, du moins localement sinon globalement, l'inertie interclasse. Ainsi, on établit les règles telles que si un sujet relève de l'intervalle $]x^i, x^j]$ de la modalité x de l'attribut *taille* alors, généralement, il relève de l'intervalle $]\pi^i, \pi^j]$ de la modalité π de l'attribut *poids*. Autrement dit, les sommets dont l'arc reliant un sommet de moindre poids sont *légers (dits indépendants),* et les sommets dont l'arc reliant un sommet de poids plus grand sont *dominants (dits aussi subordonnés)*. Les sommets subordonnés reliés à d'autres sommets subordonnés restent sous la domination du sommet indépendant en remontant la

[4] URL http ://www.ics.uci.edu/MLRepository.html

hiérarchie selon le principe de transitivité et les sommets de degré nul sont alors des *terminaux*. Les partitions en 4 sous-intervalles calculées par l'outil *rchicmgk* sont données dans le Tableau 3 ci-dessous. En effet, si par exemple, l'item dominant `Setosa`

Optimal Parameters	Optimal Parameters	Optimal Parameters	Optimal Parameters
From 4.3 To 5.3	From 2 To 2.8	From 1 To 1.9	From 0.1 To 0.6
From 5.4 To 6.2	From 2.9 To 3.3	From 3 To 4.9	From 1 To 1.6
From 6.3 To 7.9	From 3.4 To 4.4	From 5 To 6.9	From 1.7 To 2.5
[1] "Sepal.Width"	[1] "Petal.Length"	[1] "Petal.Width"	[1] "Setosa"
			[1] "Versicolor"
			[1] "Virginica"

Tableau 4 – Partitions optimales calculées par l'outil *rchicmgk*

admet généralement de poids grand sur l'item `Petal.Length`, et donc n'est pas considéré comme léger. En revanche, l'item récessif (i.e. de petite taille) `Petal.Width` admet généralement d'un moindre poids sur l'item `Sepal.Width`, et donc n'est pas considéré comme dominant. La Figure 2 ci-après rapporte les graphes implicatifs au seuil de 0.01.

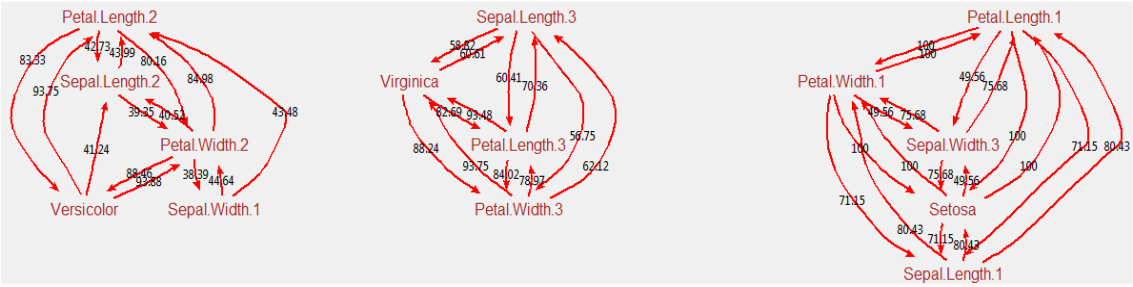


Figure 6 – Graphe implicatif au seuil α=0.01

Sur cette Figure 6 apparaissent clairement l'ébauche de trois sous-graphes respectivement composés des items (`Petal.Lenth.2`, `Sepal.Lenth.2`, `Petal.Width.2`, `Versicolor` et `Petal.Lenth.1`) noté sous-graphe 1, (`Sepal.Lenth.3`, `Viginica`, `Petal.Length.3` et `Petal.Width.3`) noté sous-graphe 2, et (`Petal.Lenth.1`, `Petal.Width.1`, `Sepal.Width.3`, `Setosa` et `Sepal.Lenth.1`) noté sous-graphe 3. Nous remarquons que, malgré la base iris n'est pas trop dense, les graphes sont très lisibles. Les premier et deuxième sous-graphes mettent en évidence la conjonction des items de différentes classes d'équivalence (c'est-ce qu'on a développé dans l'Algorithme GC2M, sous-section 3.1). Sur le sous-graphe 1, par exemple, les items `Versicolo` et `Sepal.Length.2` font partie de différentes classes d'équivalence, qui forment une règle approximative `Versicolo` → `Sepal.Length.2`, ce qui confirme les résultats du tableau 3 ci-dessus. A son tour, le sous-graphe 3 met en évidence à la fois la conjonction des motifs de mêmes classes et ceux de différentes classes d'équivalence (Algorithme 3, section 3.2.1), qui vont apparaître les règles d'association approximatives et exactes. Par exemple, les arcs `Petal.Width.1` → `Petal.Length.1` et `Setosa` → `Petal.Length.1` obtiennent respectivement des poids 75.15% et 100%, ce qui montre que `Petal.Width.1` → `Petal.Length.1` est une règle approximative, alors que `Setosa` → `Petal.Length.1` est une règle exacte. Autrement dit, les items `Petal.Width.1` et `Petal.Length.1` n'appartiennent pas à une même classe d'équivalence, tandis que les items `Setosa` et `Petal.Length.1` appartiennent à une même classe d'équivalence, ce qui confirme encore les résultats du Tableau 4 ci-dessus.

Nous continuons à augmenter le seuil α. L'objectif est de trouver un seuil de risque (ou de confiance 1-α) optimal auquel les graphes soient stables. La Figure 3 ci-après restitue l'ensemble des graphes aux seuils de 0.01 (arcs rouges) et de 0.05 (arcs verts).

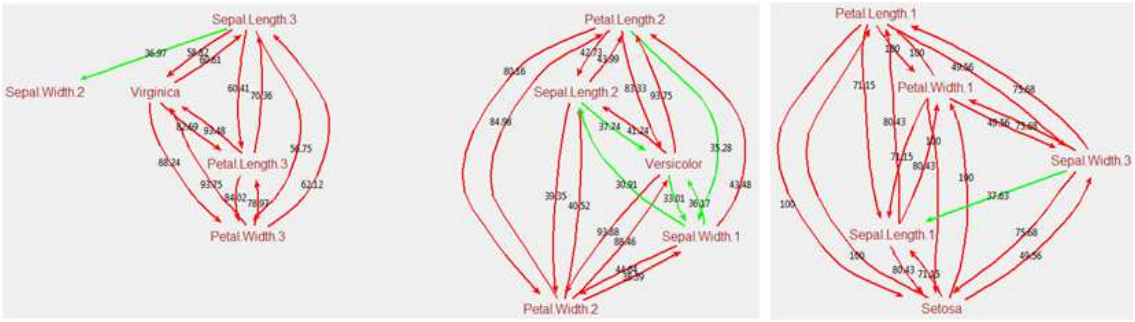


Figure 7 – Graphes implicatifs aux seuils α=0.01 et 0.05

Un peu spectaculaire, les premier et deuxième graphes de la Figure 6 se changent de place dans cette Figure 7: le 1er graphe de la Figure 6 devient le 2ème graphe de la Figure 7. Dans la suite, le sous-graphe 1 s'enrichit d'un item `Sepal.Width.2` depuis l'item `Sepal.Length.3`, et donne naissance à une nouvelle règle `Sepal.Width.2` → `Sepal.Length.3` (arc vert) du premier graphe de la Figure 7. Le 3ème graphe reste fixe, il ne change pas de place. Nous y remarquons une nouvelle liaison entre `Sepal.Width.3` et `Sepal.length.1`, et donnent naissance à un arc `Sepal.Width.3` → `Sepal.Length.1` (arc vert). Le 2ème sous-graphe a vu sa structure se complexifier, sans que des nouveaux nœuds ne viennent se greffer, des nouveaux arcs (voire chemins) s'apparaissent. En effet, nous observons l'apparition d'une équivalente statistique entre les nœuds `Sepal.Width.1` et `Verscicolor`, repérée par l'arc double `Sepal.Width.1` ⇄ `Verscicolor` de ce sous-graphe 2. Par ailleurs, nous avons marqué une nouvelle liaison équivalente statistique `Sepal.Length.2` → `Verscicolor` qui va doubler l'arc `Verscicolor`→`Sepal.Length.2`. A cet effet, deux nouveaux chemins `Sepal.Width.1`→ `Verscicolor`→ `Sepal.Length.2` et `Sepal.Length.2`→ `Verscicolor`→ `Sepal.Width.1` s'apparaissent auxquels s'ajoute à ce dernier le nouvel arc `Sepal.Width.1`→ `Sepal.Length.1` qui forme une nouvelle fermeture transitive de ce sous-graphe 2.

Augmentons encore le seuil de risque α à 0.10, les résultats seront donnés dans les graphes de la Figure 8 ci-après et centrons notre attention sur le deuxième sous-graphe.

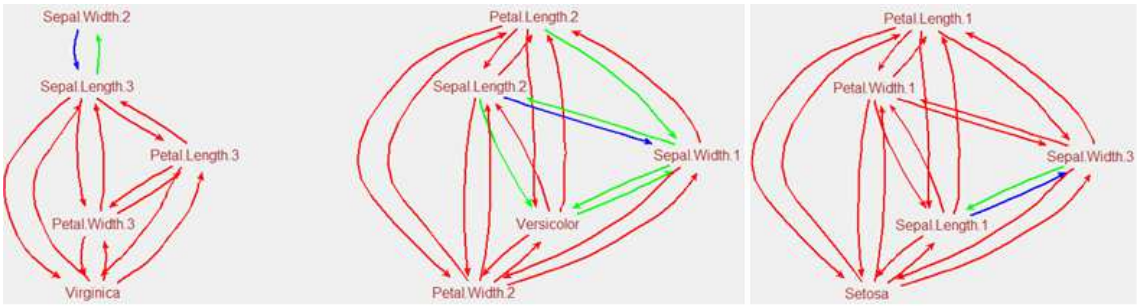


Figure 8 – Graphes implicatifs aux seuils α=0.01, 0.05 et 0.10

Les 3 sous-graphes 1, 2 et 3 n'ont pas pu tous trop bouger. Ils supportent pour chacun une seule modification, alors que cette modification n'apporte aucune nouvelle connaissance pour l'interprétation statistique. En effet, nous observons sur le sous-graphe 1 d'une équivalente statistique, repérée par l'arc `Sepal.Width.2`→ `Sepal.Length.3` en bleu. Le sous-graphe 2 s'enrichit d'une équivalente statistique, repérée par l'arc `Sepal.Length.2`→

Sepal.Width.1, en bleu. Le troisième sous-graphe s'enrichit aussi à son tour d'une équivalente statistique, repérée par l'arc Sepal.Length.1→ Sepal.Width.3.

Nous poursuivons notre analyse en faisant encore augmenter le risque α à 0.15. La Figure 9 ci-après représente les résultats obtenus. Ainsi, nous voyons une stabilité des

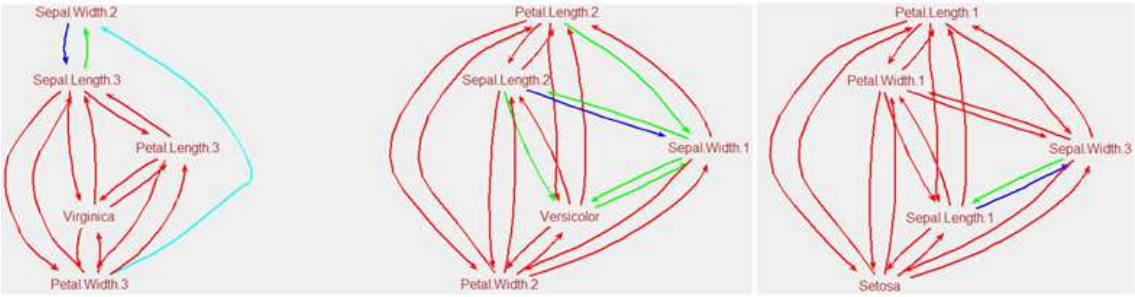


Figure 9 – Graphes implicatifs aux seuils α=0.01, 0.05, 0.10 et 0.15

sous-graphes 2 et 3 jusqu'au seuil α=0.15, alors qu'au seuil α=0.10 apparaissent respectivement les arcs Sepal.Length.2→ Sepal.Width.1 et Sepal.Width.1→ Sepal.Width.3. Malheureusement, ce sont des arcs équivalents statistiquement, qui n'apportent aucune nouvelle connaissance pour l'utilisateur, ce qui illustre juste la complexité limite admissible de l'analyse. Le sous-graphe 1 continue à son tour son instabilité. En effet, il s'enrichit d'un nouvel arc Petal.Width.3→ Sepal.Width.2 qui forme une fermeture transitive du chemin Petal.Width.3→ Petal.Length.3→Sepal.Length.3→ Sepal.Width.2. Les deux fermetures transitives entre Sepal.Length.3 et Viginica se désactivent de Viginica, et viennent former une double fermeture transitive entre Sepal.Length.3 et Petal.Width.3. Alors que nous avons marqué qu'à α=0.10 apparaît une équivalente statistique entre Sepal.Width.2 et Sepal.Length.3, repérée par l'arc double Sepal.Width.2 $\rightleftarrows$ Sepal.Length.3, qui traduit aussi la complexité limite admissible de l'analyse. A signaler que si nous poursuivons l'augmentation de α à 0.20, nous voyons une stabilité de ce sous-graphe 1. On s'aperçoit que α=0.05 est le seuil le plus optimal auquel nous allons arrêter l'analyse graphique, soumise alors à l'interprétation des résultats les plus significatifs.

5 Conclusion et perspectives

Nous avons présenté et évalué une approche de construction condensée et interactive des graphes implicatifs. Il ressort que cette approche est capitale en termes de lisibilité. Les résultats expérimentaux ont permis de juger que notre approche est à l'échelle sur les problèmes des graphes implicatifs. On peut noter qu'en dépit de sa simplicité, cette approche rend aux experts un outil très puissant pour la fouille de données massives, et permet d'aider ces experts dans leur prise de décision.

Malgré son intérêt incontestable, notre modèle pose quelques limites. Nous n'avons pas pu comparer son efficacité sur d'autres approches. Il serait alors intéressant de mener de telle comparaison sur d'autres méthodes existantes telles que celles citées dans l'introduction (Section 1). Actuellement, l'identifiant des règles positives et négatives retenues n'est pas encore réalisé mais pourrait très bien s'automatiser. Nous souhaitons aussi poursuivre ce travail en nous penchant sur le problème de classification.

Références

[1] Agrawal, R., and R. Srikant (1994), Fast Algorithms for Mining Association Rules. *Proceedings of 20th VLDB Conference*, 487-499.

[2] Anastasiadou, S. (2019), Comparison of contemporary advanced Statistical methods regarding construct validity Evaluation of teique-sf instrument: statistical Implicative analysis vs. Principal components Analysis. *In J-C. Régnier et al.(Eds), Actes du 10ème Colloque International de l'ASI,* 148-163.

[3] Bemarisika, P., and A. Totohasina (2020), An Efficient Method for Mining Informative Association Rules in Knowledge Extraction. *In A. Holzinger et al. (Eds.), CD-MAKE,* LNCS 12279, 227-247.

[4] Bruno P., J. Blanchard, P. Kunt, et F. Guillet (2007), Garviz : Une plate-forme de fouille visuelle 2D-3D de règles d'association. *In Actes de l'atelier visualisation et extraction de connaissances, Conférence EGC'07.*

[5] Bodin, A., et M-L. Giovannini (2019), Lecture didactique des résultats d'une recherche évaluative à la lumière de l'ASI. *In J-C. Régnier et al.(Eds), Actes du 10ème Colloque International de l'ASI,* pp. 416-436.

[6] Cao, L., X. Dong, and Z. Zheng (2016), E-NSP: Efficient negative sequential pattern mining. *In Artificial Intelligence*, 156-182.

[7] Carlos F-B., M. Dolores Ruiz, M. Delgado, and J. M-B. Maria (2019), A comparative analysis of tools for visualizing association rules: A proposal for visualising fuzzy association rules. *In 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT).*

[8] Cheikh, T.D., A. Giacometti, P. Marcel, et M. Ndiaye (2008), Visualisation interactive de grands ensembles de règles d'association. *In Revue des nouvelles technologies de l'information* (*RNTI).*

[9] Couturier, R., et S. Ghanem (2015), Ajout de la confiance au graphe implicatif. *In Analyse Statistique Implicative (ASI'2015)*, 117-129.

[10] Delacroix, T., P. Lenca, et S. Lallich (2017), Du local au global : Un nouveau défi pour l'analyse statistique implicative. *In J-C. Régnier et al.(Eds), Actes du 9ème Colloque International de l'ASI,* 102-115.

[11] Dong, X., G. Yongshun, and L. Cao (2018), F-NSP+: A Fast Negative Sequential Patterns Mining Method with Self-adaption Data Storage Strategy. In Pattern Recognit, 13-27.

[12] Feno, D., J. Diatta, and A. Totohasina (2006), Galois Lattices and Based for M_{GK}-valid Association Rules. *In Ben Yahia et al. (Eds.),* CLA 2006, 186-197.

[13] Feno, D.R., et A. Totohasina (2008). De la qualité des règles d'association : Etude comparative des mesures M_{GK} et confiance. In *CARI'2008*, 561-568.

[14] Ganter, B., and R. Wille (1999), Formal concept analysis: Mathematical foundations. In Springer Verlag.

[15] Gras, R. (1979), Contribution étude expérimental et l'analyse de certaines acquisitions cognitives et de certains objectifs en didactique des mathématiques, *Thèse d'Etat, Université de Rennes 1.*

[16] 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*, 11-40.

[17] Gras, R., R. Couturier, et P. Gregori (2015), Un mariage arrange entre l'implication et la confiance ? *In Colloque International de l'ASI 2015.*

[18] Guillaume, S (2000), Traitement des données volumineuses. Mesures et algorithmes d'extraction des règles d'association et règles ordinales. *Thèse de doctorat, Universté de Nantes, France.*

[19] Guyet, T., and R. Quiniou (2018), NegPSpan: Efficient extraction of negative sequential patterns with embedding constraints. *In https://hal.inria.fr/hal-01743975v2.*

[20] Hahsler, M. (2017), arulesViz: Interactive Visualization of Association Rules with R. *In R Journal Vol. 9/2*, ISSN 2073-4859, 163-175.

[21] Kuntz, P., R. Lehn, F. Guillet, et B. Pinaud (2008), Découverte interactive de règles d'association via une interface visuelle. *In Revue des nouvelles technologies de l'information* (*RNTI).*

[22] Mannila, M., H. Toivonen (1997), Levelwise Search and Borders of Theories in Knowledge Discovery. *In Data Mining Knowledge Discovery*, 241-258.

[23] Mouakher, A. et S. Ben Yahia (2011), Une nouvelle approche anthropocentrée de visualisation d'une couverture optimale de règles d'association guidée par la pertinence cognitive. *In Atelier fouille visuelle de données* (EGC'2011), 4-15.

[24] Pasquier, N., R. Taouil, Y. Bastide, G. Stumme, and L. Lakhal (2005), Generating a condensed representation for association rules. *J. of Intell. Info. Syst.,* 29-60.

[25] Pasquier, N. (2009), Frequent Closed Itemsets Based Condensed Representations for Association Rules. *Techniques for Effective Knowledge Extraction, Information Science Reference*, pp. 248-273.

[26] Rodriguez-Jimenez, J.M., P. Cordero, M. Encisoa, and A. Mora (2014), Negative attributes and implications in Formal Concept Analysis. *In Procedia Computer Science,* 758-765.

[27] Totohasina, A., and H. Ralambondrainy (2005). ION : A pertinent new measure for mining information from many types of data. *In Proceedings of IEEE, SITIS,* pp. 202-207.

[28] Xu, T., T. Li, and X. Dong (2018), Efficient High Utility Negative Sequential Patterns Mining in Smart Campusy. *In IEEE Access*, 23839-23846.

[29] Xu, Y., Y. Li, and G. Shaw (2011), Reliable representations for association rules. *In Data and Knowledge Engineering*, 555-575.

TENDÊNCIAS DE PESQUISAS SOBRE RESILIÊNCIA EDUCACIONAL EM TESES E DISSERTAÇÕES NO BRASIL: ANÁLISE À LUZ DA ASI.

Paula Virgínia Chaves Cabral Andrade[1], Aurino Lima Ferreira[2], Nadja Maria Acioly-Régnier[3], Vladimir Lira Véras Xavier De Andrade[4]

TENDANCES DES RECHERCHES SUR LA RÉSILIENCE ÉDUCATIVE DANS LES THÈSES ET LES MÉMOIRES AU BRÉSIL : ANALYSE À LA LUMIÈRE DE L'ASI

TRENDS OF RESEARCH ON EDUCATIONAL RESILIENCE IN THESIS AND DISSERTATIONS IN BRAZIL: ANALYSIS IN THE LIGHT OF ASI.

RESUMO

Apresentamos neste artigo os resultados de uma pesquisa bibliográfica do mapeamento das tendências sobre a resiliência educacional em teses e dissertações no Brasil no período de 2008 a 2021. Os dados foram tratados utilizando a estatística descritiva e a Análise Estatística Implicativa (ASI). Na ASI, utilizamos a árvore coesitiva e o grafo implicativo. Focamos as análises em quatro grandes grupos de variáveis relativas: as instituições onde foram defendidas as teses/dissertações; aos sujeitos e ambientes pesquisados; a perspectiva de resiliência e a metodologia. Os resultados indicam um crescimento nas pesquisas na área educacional sobre resiliência e uma concentração na Região Sudeste, sendo seguida pelas Regiões Sul e Nordeste. Apesar desse crescimento, ainda foram observadas muitas pesquisas limitadas a uma perspectiva individual de resiliência, indicando a necessidade de se ampliar o debate na área face às limitações dessa abordagem.

Palavras-chave: *Resiliência Educacional, pesquisa bibliográfica, Análise Estatística Implicativa.*

RÉSUMÉ

Dans cet article, nous présentons les résultats d'une revue de la littérature cartographiant les tendances de la résilience éducative dans les thèses et mémoires au Brésil de 2008 à 2021. Les données ont été traitées à l'aide de statistiques descriptives et d'une analyse statistique implicative (ASI). En ASI, nous utilisons l'arbre cohésif et le graphe implicatif. Nous avons concentré les analyses sur quatre grands groupes de variables relatives : les établissements où les thèses/mémoires ont été soutenus ; aux sujets et environnements recherchés; la perspective de la résilience et de la méthodologie. Les résultats indiquent une croissance de la recherche dans le domaine éducatif sur la résilience et une concentration dans la région du Sud-Est, suivie par les régions du Sud et du Nord-Est. Malgré cette croissance, il existait encore de nombreuses études limitées à une perspective individuelle de la résilience,

[1] Doctorante en Sciences de l'Éducation en régime de cotutelle entre l'Université Lumière Lyon 2 – Laboratoire Éducation Cultures Politiques, École Doctorale 485 EPIC et l'Université Fédéral de Pernambuco – Brasil, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife – PE, Brasil, CEP: 50670-901, paulavchaves@gmail.com

[2] Departamento de Psicologia e Orientação Educacionais - Programa de Pós-Graduação em Educação da Universidade Federal de Pernambuco, Av. Prof. Moraes Rego, 1235 - Cidade Universitária, Recife – PE, Brasil, CEP: 50670-901. aurinolima@gmail.com.

[3] Nadja Acioly-Régnier, Université Claude Bernard Lyon 1 – INSPE – Lyon, France EA 4571 - Laboratoire Éducation Cultures Politiques, École Doctorale 485 EPIC, nadja.acioly-regnier@univ-lyon1.fr

[4] 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 – Recife-PE, Brasil, CEP: 52171-900, vladiandrade@gmail.com

indiquant la nécessité d'élargir le débat dans le domaine étant donné les limites de cette approche.

Mots-clés : *Résilience éducative, Analyse statistique implicative, recherche bibliographique.*

ABSTRACT

In this article, we present the results of a literature review mapping trends in educational resilience in theses and dissertations in Brazil from 2008 to 2021. Data were treated using descriptive statistics and Statistical Implicative Analysis (ASI). In ASI, we use the cohesive tree and the implicative graph. We focused the analyzes on four large groups of relative variables: the institutions where the theses/dissertations were defended; to the researched subjects and environments; the perspective of resilience and methodology. The results indicate a growth in research in the educational area on resilience and a concentration in the Southeast region, followed by the South and Northeast regions. Despite this growth, many researches limited to an individual perspective of resilience were observed, indicating the need to broaden the debate in the area given the limitations of this approach.

Keywords : *Educational Resilience, Statistical Implicative Analysis, bibliographic research.*

1 Introdução

Neste artigo, fazemos um levantamento das pesquisas realizadas sobre resiliência educacional em teses e dissertações no campo acadêmico brasileiro. Designamos o termo resiliência educacional para as pesquisas específicas sobre resiliência relacionadas à área de educação e formação.

O termo resiliência é utilizado em várias áreas do conhecimento, tais como na Psicologia, na Sociologia, na Educação, na Arquitetura, na Física, na Ecologia etc. Tendo nos últimos anos adquirido uma complexidade "multissistêmica" (Ungar, 2018a, 2020), requerendo dos pesquisadores um esforço para delimitar qual fenômeno estão investigando. Brandão, Mahfoud e Gianordoli-Nascimento (2011) indicam que não existe um consenso para o surgimento da palavra resiliência e que na literatura brasileira é quase que uma tendência autores associarem esse construto como vindo da física ou importado dessa disciplina, contudo:

> [...] se o termo/conceito de resiliência usado pela Psicologia foi originado na Física, na resistência dos materiais, ele foi transposto de maneira imprecisa, pois se relaciona mais com o conceito de elasticidade do que de resiliência dos materiais (Brandão, Mahfoud, Gianordoli-Nascimento, 2011, p. 264).

Yanaga (2017) corrobora com este pensamento e indica que termos como invulnerabilidade e invencibilidade não deveriam ser utilizados para se referirem diretamente à resiliência, haja vista que mesmo em seu caráter polissêmico e sem um consenso sobre sua definição, a resiliência apresenta:

> três componentes essenciais [...]: 1. a noção de adversidade, trauma, risco ou ameaça ao desenvolvimento humano; 2. a adaptação positiva ou superação da adversidade; 3. o processo que considera a dinâmica entre mecanismos emocionais, cognitivos e socioculturais que influem no desenvolvimento humano. (Infante, 2005, p. 26)

De acordo com Infante (2005), a identificação e organização de fatores protetivos e de risco em pessoas com psicopatologias na família ou que sofriam situações de estresse psicológico e adversidades por um grande período de tempo foi uma grande contribuição

para o avanço dos estudos na área da resiliência, configurando-se como característica central da primeira geração de estudos sobre resiliência.

Os estudos da primeira geração trazem uma perspectiva da resiliência como característica pessoal e segundo Brandão (2009) influenciou o campo de pesquisas no Brasil, sendo ainda encontrados em estudos mais recentes. Dessa forma, classificamos os levantamentos das pesquisas que fizemos sobre resiliência como característica pessoal, aquelas nas quais prevaleciam o foco nos fatores mais internos dos indivíduos. Vamos chamar essa primeira perspectiva como **resiliência individual**. Consideramos essa perspectiva limitada, pois coloca sobre os indivíduos como únicos responsáveis pelo processo de superação das adversidades. Brandão e Nascimento (2019) chamam essa primeira fase ou primeira geração sobre estudos da resiliência enquanto traço.

Uma outra perspectiva de resiliência apontada por Brandão (2009) e Brandão e Nascimento (2019) é a da resiliência como um processo. Libório, Castro e Coelho (2006, p. 92) apresentam uma comparação entre a resiliência individual e a resiliência enquanto processo afirmando que : "[...] sua compreensão como um traço individual, ou como um fenômeno ou processo, que decorre de inter-relações do indivíduo e o meio". Nessa segunda perspectiva, temos a resiliência como resultante do relacionamento dos indivíduos com o meio.

> Uma segunda geração de pesquisadores expandiu o tema da resiliência em dois aspectos: a noção de processo, que implica a dinâmica entre fatores de risco e de resiliência, que permite ao indivíduo superar a adversidade, e a busca de modelos para promover resiliência de forma efetiva em termos de programas sociais (Infante, 2005, p. 24).

Dentro dessa perspectiva, temos pesquisas que apontam o relacionamento dos sujeitos com a família, com a comunidade ou ainda com a escola, como elemento promotor da resiliência. (Rutter, 1987, 1993; Grotberg, 2005, Ferreira, 2007)

Brandão e Nascimento (2019, p. 4-5) destacam que, embora ainda tenhamos pesquisadores que estão limitados a uma perspectiva individual da resiliência, atualmente:

> [...] caminha-se para um consenso no entendimento da resiliência como um fenômeno processual complexo que envolve um sistema composto de indivíduos (e suas características intrínsecas e relacionais, neurológicas, genéticas e epigenéticas), famílias, comunidades, cultura, sociedade, apoio social, crenças, não podendo, portanto, ser reduzido ou determinado por nenhum desses âmbitos isoladamente [...].

O nosso entendimento é no sentido de que sozinhos não conseguimos ultrapassar as adversidades impostas pela vida. Em especial quando estamos tratando da relação eminentemente dialógica entre resiliência e educação. Em algum tempo, em algum lugar, precisamos do apoio da família, ou da avó que sempre nos ajudou nos estudos, de um tio, de um padrinho, de um amigo, de uma professora, enfim, precisamos de sistemas (pessoas, família, escola, etc.) que nos impulsionem a termos uma rede maior de ajuda, seja ela de forma psicológica ou material. Ferreira (2007) ressalta o papel das relações e apoios para promover aspectos resilientes que podem potencializar processos de formação humana.

No início de suas pesquisas, Ungar (2018b) acreditava que o sucesso das crianças era devido a algumas qualidades especiais que elas tinham como indivíduos. Somente quando Ungar (2018b, p. 6, tradução livre) começou a pesquisar a resiliência é que viu que "[...]

as qualidades individuais das crianças contam muito menos do que bons ambientes de apoio ricos em recursos. Esses ambientes incluem ótimos programas projetados para criar as condições que as crianças precisam para serem melhores, não importando os desafios que enfrentam".

Acioly-Régnier, Chaves e Araújo (2011) ressaltam que a concepção de que a resiliência se desenvolve ao longo da vida, não se limitando aos primeiros anos da vida, é coerente com a teoria dos Campos Conceituais (Vergnaud, 1994) que investiga também o desenvolvimento cognitivo adulto, podendo ser usada dessa forma na investigação de pesquisa com professores no tocante à resiliência.

Com base nos aspectos citados, vamos classificar as pesquisas que não se limitam à resiliência enquanto atributo do indivíduo, mas a resiliência enquanto processo. Para comparar com a perspectiva anterior, vamos chamar essa perspectiva de **resiliência processual**.

Em sua pesquisa de mestrado, Silva (2019) investigou processos de educação não formal realizada pelo Núcleo Educacional Irmãos Menores de Francisco de Assis (Neimfa), organização da sociedade civil, destacando o constructo de **resiliência integral** para se referir a:

> [...] um processo integral que contempla as dimensões subjetiva, intersubjetiva, objetiva e interobjetiva do ser, possibilitando a superação de eventos traumáticos vividos ao longo de suas vidas [...]. A noção de resiliência que assumimos neste trabalho, toma o humano como um ser integral, multidimensional, processual, histórico, imerso em uma rede relacional complexa de fatores socioculturais que o habilita a possibilidades de modificações constantes (Silva, 2019, p. 89).

Silva (2019) ressalta que essa definição está alinhada com a perspectiva ecológica e social do pesquisador canadense Michael Ungar, na qual a resiliência seria a capacidade dos sujeitos e, também, da coletividade de mobilizarem "[...] recursos culturalmente relevantes de que precisam para se saírem bem diante das adversidades, bem como sua capacidade de negociar para que esses recursos sejam fornecidos de forma significativa" (Ungar, 2015, p. 89, tradução livre).

A introdução dos estudos sociológicos e culturais propostos por Ungar (2015) aliada à inclusão de perspectivas integrais (Brasil, Ferreira, 2020; Ferreira et al. 2018, Silva, 2019) na investigação sobre a temática, apresenta um caminho crítico e desnaturalizante do essencialismo da resiliência, enquanto condição preestabelecida e da noção de pessoa resiliente, de modo que a resiliência passa a ser compreendida como um fenômeno humano que surge a partir da interação consigo, com as outras pessoas e com o mundo, sendo a percepção e a qualidade das relações vividas e estabelecidas um grande sinalizador de possibilidades no enfrentamento das adversidades.

Dada a complexidade das perspectivas assumidas no campo acadêmico brasileiro acerca da resiliência educacional, com repercussões diretas nos modos de encaminhamentos das pesquisas, indicamos a importância de mapeá-las. Neste sentido, esta pesquisa tem como objetivo apresentar as tendências sobre a resiliência educacional em teses e dissertações no Brasil no período de 2008 a 2021, indicando a distribuição das pesquisas por instituições, sujeitos e ambientes investigados, possíveis perspectivas de resiliência assumidas e elementos da metodologia.

2 Modelagem das variáveis e construção dos dados

Esta pesquisa objetiva a observação de tendências sobre diferentes pesquisas na área de resiliência educacional. Apresentamos a seguir a descrição de como os dados foram modelados e como as escolhas foram feitas.

2.1 Construção da amostra

Para realização deste estudo, fizemos uma busca em teses e dissertações no Brasil sobre resiliência em pesquisas na área de Educação com foco no ensino. Para isso, utilizamos dois sistemas de buscas. O primeiro foi a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD). Ela é ligada ao Instituto Brasileiro de Informação em Ciência e Tecnologia e teve início em 2002. Buscamos pelas palavras descritoras "resiliência educacional". Selecionamos as teses e dissertações de 2008 a 2021 (dos últimos 13 anos). Foram achados, entre teses e dissertações, 46 publicações. Em um segundo momento, se olhou o título, o resumo e quando necessário, o texto na íntegra (tese/dissertação). Procurou-se observar pesquisas em que a resiliência era um elemento que envolvia o ensino e/ou aprendizagem tanto na educação básica como na educação superior. Ampliamos também para a educação não formal, como por exemplo, formações em ONGs, museus etc. Das 46 teses/dissertações, achamos 14 que contemplam esses critérios. Além desses critérios, tivemos uma que foi excluída por não conter resumo e meios de acesso ao documento digital, nem pelo sistema da BDTD, nem por pesquisa no buscador do Google.

O segundo sistema de busca foi o catálogo de teses e dissertações da CAPES. Buscamos o termo "resiliência educacional" e não foi encontrado nenhuma pesquisa sobre o assunto. Mudamos para o termo "resiliência na educação", logo foram achadas 7 pesquisas. Dessas, depois de uma análise dos títulos e resumos, tivemos 5 selecionadas que atendessem ao critério desta pesquisa. Uma foi excluída por não apresentar nenhuma informação além do título. Buscamos pela ferramenta de pesquisa do Google e não achamos a referida dissertação. Por último, para ampliar a busca, procuramos apenas o termo "resiliência". Esse termo é utilizado em diversas áreas: Educação, Psicologia, Sociologia etc. O resultado foi encontrado em 3432 trabalhos. Como o foco é em Educação, utilizamos a ferramenta "refinar meus resultados" e selecionamos como área de concentração: Psicologia e Educação. Logo foram encontrados 96 trabalhos. Com base no título, fizemos uma triagem de trabalhos que pelo título indicavam que tratavam da resiliência educacional e quando surgiam dúvidas, íamos ao resumo. Encontramos pesquisas na escola nas quais no resumo não tínhamos nenhuma referência ao termo resiliência, o que fez com que fosse descartada da amostra. Também observamos em resumos a palavra resiliência, mas que o foco não era no processo educacional, mas envolvia outros aspectos. Também foram excluídos os que se repetiam em outras buscas. Nessa nova triagem, foram encontradas mais 13 pesquisas. Com isso, o universo desta pesquisa é formado por 32 teses/dissertações.

2.2 Quadro teórico do tratamento e da análise dos dados

Os dados foram construídos e organizados utilizando-se para isso a estatística descritiva e a Análise Estatística Implicativa.

Na estatística descritiva temos as técnicas necessárias para organização e sistematização dos dados de forma que possibilite a sua análise e interpretação. Contudo, para observar as relações entre as variáveis, consideramos necessário o uso da Análise Estatística Implicativa (ASI). O uso da ASI possibilita um enriquecimento nas análises realizadas. Segundo Andrade (2020, p.12), a ASI é "um método classificatório não simétrico de análise de dependências orientadas que se apoia em bases probabilísticas". Trata-se de uma abordagem teórica que possibilita a descoberta de relações não apresentadas por outros métodos estatísticos que se apoiam na probabilidade condicional. A abordagem da ASI considerando a utilização de metarregra (regras de regra) em uma hierarquia orientada, é no momento única no mundo (Gras, Régnier, 2015). Utilizaremos em nossa pesquisa o tratamento com a árvore coesitiva que possibilita a observação das metarregras em relação aos dados que construímos. Também utilizaremos o uso do grafo implicativo que faz parte do quadro teórico da ASI e possibilita selecionar as variáveis no modo cone, indicando relações sobre determinadas variáveis que se pretendem focar.

2.3 Construção das variáveis

Com base nos levantamentos feitos, organizamos as variáveis em 4 grandes categorias:

a) Variáveis relativas à instituição onde foi defendida a tese/dissertação (Quadro 1);
b) Variáveis relativas aos sujeitos e ambientes da pesquisa (Quadro 2);
c) Elementos que pudessem indicar a perspectiva da resiliência (Quadro 3);
d) Elementos da metodologia da pesquisa (Quadro 4).

No primeiro critério, os dados foram organizados segundo o ano de publicação, indicados no Quadro 1. Nem todos os anos indicados no Quadro 1 foram observados, deixamos apenas para indicar que todos esses anos foram contemplados na busca, embora nem sempre encontradas pesquisas. O objetivo da indicação do ano era de termos uma indicação da evolução e das tendências por ano. Também consideramos o tipo de publicação (tese ou dissertação) e a instituição onde estava sendo realizada a pesquisa: pública ou privada e a área do programa (Educação ou Psicologia). Também procuramos observar onde estavam sendo realizados os estudos (estado e região).

No segundo critério, foram levantados: o ambiente onde foi realizada a pesquisa, o sujeito da pesquisa e o nível de ensino. Na categoria ambiente, poderíamos ter uma pesquisa em uma instituição pública de ensino, em uma instituição privada de ensino ou em espaços não formais. Para os casos não identificados, colocamos a categoria "outros".

Em relação ao sujeito da pesquisa, foram observadas pesquisas que envolvem: professor, gestor/coordenação escolar, estudantes, professores em formação inicial (fizemos esse recorte por considerar relevante essa categoria na área de Educação), documentos/livros e teses (pesquisas bibliográficas e/ou documentais).

Observamos também pesquisas que abordam grupos específicos, tais como: pesquisas com indígenas, negros, mulheres, pessoas em situação de rua e com necessidades especiais.

Na categoria nível escolar, foram consideradas: ensino fundamental anos iniciais, ensino fundamental anos finais, ensino médio, EJA e ensino superior. Não identificamos pesquisas com a educação infantil e por isso não incluímos essa no quadro 2. Observamos pesquisas que não estão associadas a espaços formais. Como exemplos de pesquisas em

Educação em espaços não formais, temos as promovidas em museus e ONGs. Uma última categoria criada foi: não se aplica. Nessa categoria, temos pesquisas em que não se podem encaixar em um nível específico, como uma pesquisa com um afrodescendente em que o mesmo sofreu traumas na infância e na adolescência (ao longo da sua trajetória de vida e não em um nível escolar específico) e na pesquisa relata o apoio dos professores como parte de uma rede de apoio de promoção de processos de resiliência. Também temos pesquisas com professores, que não classificamos em um nível escolar específico e por isso foi inserida nessa categoria. Foram também incluídas nessa última categoria, levantamentos não limitados a um nível escolar.

No Quadro 3, os dados foram organizados segundo as perspectivas de resiliência (descritas na introdução desta pesquisa).

No Quadro 4, temos as variáveis relativas à metodologia. Nos apoiamos nessa categoria para os elementos indicados no resumo das teses/dissertações levantadas, e em alguns não apresentados, complementamos as informações no corpo das teses/dissertações.

Apresentamos nos Quadros 1, 2, 3 e 4 as variáveis agrupadas segundo as classificações definidas nesta pesquisa e os seus respectivos códigos. Os códigos foram criados para o tratamento estatístico com o software CHIC.

Descrição		**Variável**
Ano	Publicações de 2008	A_2008
	Publicações de 2009	A_2009
	Publicações de 2010	A_2010
	Publicações de 2011	A_2011
	Publicações de 2012	A_2012
	Publicações de 2013	A_2013
	Publicações de 2014	A_2014
	Publicações de 2015	A_2015
	Publicações de 2016	A_2016
	Publicações de 2017	A_2017
	Publicações de 2018	A_2018
	Publicações de 2019	A_2019
	Publicações de 2020	A_2020
Tipo	Dissertação	TiDI
	Tese	TiTE
Instituição de realização da tese/dissertação	Instituição pública	IES_Pb
	Instituição privada	IES_Priv
Área do Programa de Pós-Graduação	Educação	ArED
	Psicologia	ArPS
	Outras	ArOut

Quadro 1 – Variáveis relativas às instituições onde foram defendidas as teses/dissertações.

Descrição		Variável
Estado	Distrito Federal	EsDF
	Goiás	EsGO
	Minas Gerais	EsMG
	Paraná	EsPR
	Pernambuco	EsPE
	Rio de Janeiro	EsRJ
	Rio Grande do Norte	EsRN
	Rio Grande do Sul	EsRS
	São Paulo	EsSP
Região	Centro-Oeste	RgCO
	Nordeste	RgNE
	Norte	RgNT
	Sudeste	RgSUD
	Sul	RgSUL

Quadro 1 (*continuação*) – Variáveis relativas às instituições onde foram defendidas as teses/dissertações.

Ambiente onde foi realizada a pesquisa	Instituição de ensino pública	A_Epb
	Instituição de ensino privada	A_Epv
	Espaços não formais	A_ENF
	Ambiente não identificado se é público, privado ou não se aplica.	A_Outro
Sujeito da pesquisa	Professor	SuPro
	Gestor/coordenação escolar	SuGest
	Estudante	SuEst
	Professor em formação inicial	SuPFI
	Documentos/livros/teses	SuDoc
Sujeito da pesquisa que pertence a grupos específicos	Negros	G_NG
	Índios	G_IN
	Mulheres	G_M
	Situação de rua	G_SR
	Necessidades especiais	G_NESP
Nível de ensino	Ensino fundamental (anos iniciais)	NiEF01
	Ensino fundamental (anos finais)	NiEF02
	Ensino médio	NiEM
	EJA	NiEJA
	Ensino superior	NiES
	Formação em espaços não formais como ONGs, museus etc	NiOut
	Não se aplica, por exemplo pesquisa com professores ou com gestores	NiNao

Quadro 2 – Sujeito e ambiente da pesquisa.

Perspectiva da resiliência	Individual	PR_Ind
	Processual	PR_Proc
	Integral	PR_Int
	Levantamento de concepções/posicionamento	PR_Lev

Quadro 3 – Perspectiva da resiliência.

Abordagem	Qualitativa	AB_Qual
	Quantitativa	AB_Quant
	Mista	AB_Mist
	Não descrita pelo pesquisador(a)	AB_Nao
Método de pesquisa/ Instrumento	Documental	MP_Doc
	Bibliográfica	MP_BIB
	Análise de narrativas	MP_AN
	Análise de narrativas de vida/autobiográficas	MP_NV
	Análise de conteúdo	MP_AC
	Análise textual discursiva	MP_ATD
	Métodos visuais (vídeos, fotos)	MP_MV
	Estudo de caso	MP_EC
	Pesquisa participante	MP_PP
	Entrevista	MP_Ent
	Entrevista com grupo focal	MP_EGF
	Entrevista semiestruturada	MP_ESE
	Pesquisa por questionário	MP_Qt
	Outros	MP_Out
Instrumentos de medição	Escala de bem-estar espiritual (EBE)	EscBEE
	Escala de Resiliência para Crianças e Adolescentes - ERCA (Sandra Prince-Embury, 2007, traduzida e validada por Barbosa, 2008).	EscERCA
	Medida de Resiliência em Crianças e Jovens (Child and Youth Resilience Measure – CYRM)	EscCYRM
	Escala de Avaliação das Atribuições de Causalidade para Sucesso e Fracasso Escolar (Boruchovitch & Santos, 2011)	EscASF
	ACE (Adverse Childhood Events)	EscACE
	ScanBullying	EscScanB
	ESCALA DE RESILIÊNCIA: Wagnild e Young (1993)	EscWY

Quadro 4 - Metodologia

2.4 Tratamento das variáveis por meio da A.S.I.

Para o tratamento dos dados por meio da ASI, se faz necessário o uso de um software específico chamado de CHIC. O termo CHIC é uma abreviatura de Classificação Hierárquica Implicativa e Coesitiva. Couturier, Bodin e Gras (2003, p. 1) informam que esse software: "[...] 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 obtidas por meio dessas regras". Nesta pesquisa, utilizamos a versão 7.0 do software CHIC.

3 Resultados e análises

Organizamos os resultados em duas partes. Na primeira parte, fizemos um tratamento com a estatística descritiva das pesquisas levantadas. E na segunda parte, organizamos o tratamento dos dados utilizando a ASI.

3.1 Pesquisas levantadas

Um primeiro elemento que levantamos foi o ano das teses e dissertações que envolvem o tema sobre a resiliência educacional. O período dos anos foi de 2008 a 2021. Nos anos de 2010, 2020 e 2021 não encontramos nenhuma dissertação ou tese sobre o tema da resiliência na educação. Nos demais anos, observamos uma tendência de crescimento até 2014, depois uma queda nas produções e depois novamente um crescimento nas pesquisas, chegando a 7 produções em 2019. Nos anos de 2020 e 2021, não observamos produções ligadas à temática desta pesquisa. Procuramos observar em outras áreas apenas para ver se o problema se restringiu às pesquisas que envolvem a resiliência educacional. Observamos o catálogo de teses e dissertações da CAPES com a palavra educação que foi de 6.832 publicações que devem ampliar, pois ainda estamos em meados de 2021 e outras teses/dissertações devem ser publicadas. Na Biblioteca Digital Brasileira de Teses e Dissertações, buscando o termo "educação" e selecionando o ano de 2021, foram encontradas 2.562 publicações de diversas áreas com o termo educação. Isso indica, no que tange às pesquisas-objeto desse trabalho, uma ausência de produções em 2020 (ano que se iniciou a pandemia covid-19) e em 2021, na área de resiliência educacional. Destacamos que com a pandemia covid-19, houve redirecionamentos de muitas pesquisas de campo que ficaram inviáveis para o modo presencial, tendo que ser readaptadas e inovadas, atrasando com isso a conclusão ou ainda mudando o direcionamento de muitas pesquisas.

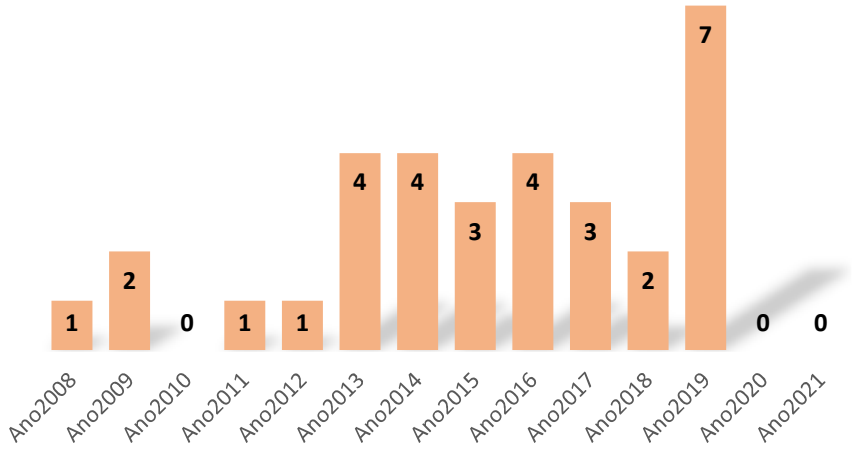


Figura 3 – Ano das teses e/ou dissertações selecionadas

Um outro elemento levantado nesta pesquisa se tratava de teses ou dissertações. Das 32 pesquisas levantadas, observamos que a maioria delas era dissertação. Com um total de 22 dissertações e 10 teses. Também procuramos observar se essas produções eram de instituições públicas ou privadas. A maioria das pesquisas levantadas era de instituições públicas, com um total de 25 pesquisas. Nas instituições privadas foram levantadas apenas 7 pesquisas. Os critérios de busca eram pesquisas ligadas à Educação. Dessa forma, a maioria dos programas era na área da Educação, levando a que 30 pesquisas fossem na área de Educação. Contudo, foi observado uma pesquisa na área da Psicologia e uma pesquisa em outra área. Na área da Psicologia, observamos uma dissertação de 2008 (Rachman, 2008), cujo objetivo era fazer um levantamento de pesquisas sobre

resiliência, tanto na área de Educação, como na área da Psicologia. Dessa forma, a Educação fazia parte da temática desta pesquisa. A pesquisa identificada em outra, era uma pesquisa de mestrado em "Desenvolvimento Humano: Formação, Políticas e Práticas Sociais". Nessa pesquisa (Costa, 2013) procurou observar em 71 professores fatores estressores da sua atividade, logo foi aplicada uma escala de resiliência de Wagnild e Young e observado fatores promotores de resiliência, como também fatores que dificultam o desenvolvimento da resiliência.

Um outro ponto que consideramos relevante é se as produções são realizadas na maioria dos estados ou se concentram em alguns estados. Observamos que as produções se limitam a 9 estados, conforme o gráfico 3. A maior produção está no estado de São Paulo (que tem um grande número de universidades e também é o mais populoso do Brasil) com 10 teses/dissertações. Sendo 3 da Universidade Estadual de Campinas - UNICAMP (Silva 2012; Garcia, 2013); 2 da Universidade de Taubaté (Salles, 2014; Costa, 2013), 2 da Universidade Estadual de Paulista – UNESP (Pereira, 2009; Yanaga, 2017), 2 da Universidade Estadual Paulista Júlio de Mesquita Filho (Sakotani, 2016; Francisco, 2013), 1 da Pontifícia Universidade Católica de São Paulo - PUC-SP (Rachman, 2008) e 1 do Centro Universitário Moura Lacerda - Ribeirão Preto (Junior, 2019). Em segundo lugar, temos o Rio Grande do Sul com 7 produções, sendo 3 da Universidade Federal de Santa Maria (Rodrigues, 2015; Trevisan, 2014; Bulhões, 2016), 2 da Universidade La Salle (Bady, 2019; Weschenfelder, 2017), 1 da Universidade Federal do Rio Grande do Sul (Muller, 2019) e 1 da Universidade Católica do Rio Grande do Sul (Almeida, 2019). Em terceiro lugar, temos Pernambuco com 5 dissertações/teses (Brasil, 2015; Brasil 2019; Silva, 2019; Silva, 2016; Rezende, 2016). As 5 pesquisas levantadas em Pernambuco estão associadas ao Programa de Pós-Graduação em Educação da Universidade Federal de Pernambuco. Dos três estados, Pernambuco foi o que obteve o maior número de pesquisas por instituição. Em quarto lugar, temos Minas Gerais com três produções (Assis, 2015; Porcaro, 2011; Corrêa, 2019). Em quinto lugar com duas publicações temos o Rio Grande do Norte (Silva, 2019; Cruz, 2018) e o Paraná (Tatiane, 2014; Marochi, 2017). Por último, temos com uma pesquisa o Distrito Federal (Ferreira, 2018); Goiás (Machado, 2014) e o Rio de Janeiro (Pereira, 2013).

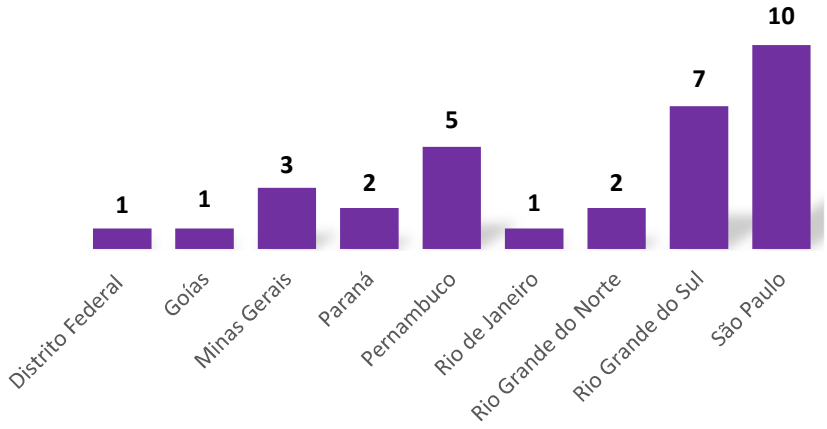


Figura 4 – Pesquisas por estado.

Procuramos, além de observar por estado, levantar as produções por regiões. A Região Sudeste lidera as pesquisas levantadas com 14 produções (a maioria é de São Paulo: 10)

que corresponde a 44% das pesquisas levantadas no Brasil. Em segundo lugar, temos a Região Sul com 9 dissertações/teses (a maioria do Rio Grande do Sul: 7) que corresponde a 28% das produções. E em terceiro lugar são as do Nordeste com 7, que correspondem a 22% das pesquisas nacionais na seleção realizada. A concentração das produções do Nordeste é no estado de Pernambuco. No Centro-Oeste, foram observadas duas produções que correspondem a 6%. Na região Norte, não foi observada nenhuma produção nas instituições dessa região, o que não significa dizer que não foram realizadas pesquisas nessa região. Observa-se em Correa (2019) uma pesquisa de uma universidade de Minas Gerais cujos sujeitos são docentes indígenas em Roraima. Contudo, eles não aparecem neste primeiro momento, no qual queríamos identificar as instituições onde foram realizados os mestrados/doutorados.

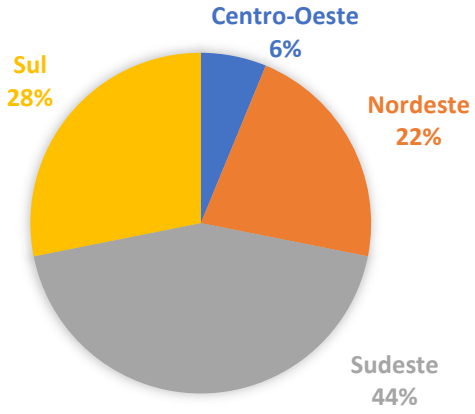


Figura 5 – Pesquisas por região.

Em um segundo momento, procuramos avaliar o ambiente onde foi realizada a pesquisa. Na figura 4, apresentamos os resultados e as porcentagens. Com base nos dados construídos, organizamos essa categoria em 4 tipos. Os dois primeiros foram em instituições públicas ou privadas de ensino. Observamos que uma pesquisa foi realizada em um espaço não formal. Trata-se da ONG Núcleo Educacional Irmãos Menores de Francisco de Assis (NEIMFA) (Silva, 2019). Essa organização é situada em uma comunidade muito carente no Recife, o Coque. Ela desempenha diversas atividades de assistência à comunidade carente, entre elas a de educação não formal e de formação humana integral. Dessa forma, criamos uma categoria de pesquisa em espaços não formais. Em terceiro lugar, temos pesquisas que não estão associadas a uma determinada instituição, como pesquisas em banco de teses e dissertações ou ainda que não tinha uma identificação da instituição em que foi realizada. A maioria das instituições onde ocorreram as pesquisas são instituições públicas com um total de 22 pesquisas nessas instituições, o que correspondem a 67% do total. Em segundo lugar, temos instituições privadas com 4 pesquisas realizadas nessas instituições, que correspondem a 18% do total. Em terceiro lugar com 6 pesquisas (18% do total), temos pesquisas na categoria outros. Os resultados ressaltam o papel social das instituições públicas de ensino.

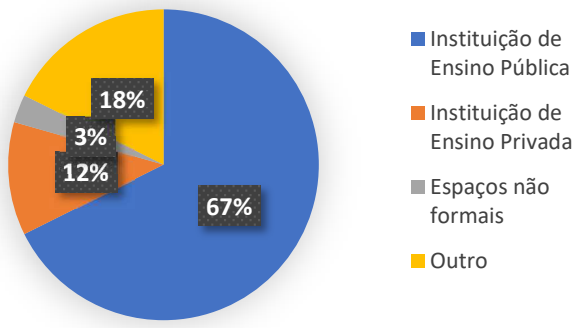


Figura 6 – Instituições de ensino onde foram realizadas as pesquisas.

Procuramos observar nas 32 teses/dissertações levantadas quais eram os sujeitos das pesquisas. Apresentamos o resultado na Tabela 1. O total apresentado é 33, uma vez que temos uma pesquisa (Machado, 2014) que aborda a resiliência ligada ao bem-estar do professor em teses e dissertações. Dessa forma, consideramos ela como uma pesquisa relacionada ao professor, ao mesmo tempo que tem como sujeito as teses/dissertações. O maior número de pesquisas tem como o sujeito o estudante, observando a resiliência ligada à essa fase da vida. Temos um total de 16 pesquisas com estudantes, o que corresponde a 48,49% do total. Esse número na verdade é maior, uma vez que os professores em formação inicial são estudantes da licenciatura. Incluímos em separado os professores em formação inicial, uma vez que os futuros professores têm uma dupla função: lidar com os problemas estressantes da sua atuação profissional e por outro lado contribuir no desenvolvimento dos estudantes.

Observamos duas pesquisas que envolvem o gestor ou diretor ou coordenador das escolas. Esse profissional também pode contribuir ou não na promoção da resiliência em função da forma como ele atua, considerando necessária uma formação do gestor, tendo em vista a promoção de fatores protetivos que contribuíam para os fatores estressantes que envolvem os professores e estudantes.

Sujeito da pesquisa	**N**	**%**
Professor	8	24,24
Professor em formação inicial	4	12,12
Estudante	16	48,49
Gestor	2	6,06
Pesquisa bibliográfica e/ou documental	3	9,09
Total	33	100,0

Tabela 1 – Sujeito da pesquisa

As pesquisas em sua maioria são com estudantes, muitos vindos de escolas públicas com problemas sociais. Contudo, observamos pesquisas com certas categorias específicas como pesquisas com negros, indígenas, com a educação especial, com mulheres e pessoas em situação de rua. A educação especial (conforme consta no decreto 10.502 de 30 de setembro de 2020) é vista como sendo "modalidade de educação escolar oferecida, preferencialmente, na rede regular de ensino aos educandos com deficiência, transtornos globais do desenvolvimento e altas habilidades ou superdotação". Em nosso levantamento, observamos pesquisas específicas com sujeitos que tinham necessidades especiais. Podemos observar em Yanaga (2017) que investiga a inclusão escolar e processo de resiliência. Nessa pesquisa, temos doze alunos como sujeitos. Sendo, quatro

surdos, dois com Síndrome de Asperger, um cego, um deficiente visual, um com Síndrome de Down, um com amputação do braço, um com raquitismo, um com mielomeningocele. Já na pesquisa de Bulhões (2016), temos um estudante superdotado/altas habilidades. Em Cruz (2018), temos uma pesquisa que investigou narrativas autobiográficas de quatro adultos cegos. Foram consideradas três categorias: relações com a escola regular e especial, estratégias de enfrentamento e tutores de resiliência.

Na pesquisa com mulheres em situação de rua (Bady, 2019) foi incluída tanto em pesquisas que envolvem mulheres, como pesquisas com pessoas em situações de rua. Esse estudo aborda a violência de gêneros dessas mulheres que participam da Educação de Jovens e Adultos (EJA). Na figura 5, apresentamos um gráfico que mostra as frequências dessas variáveis.

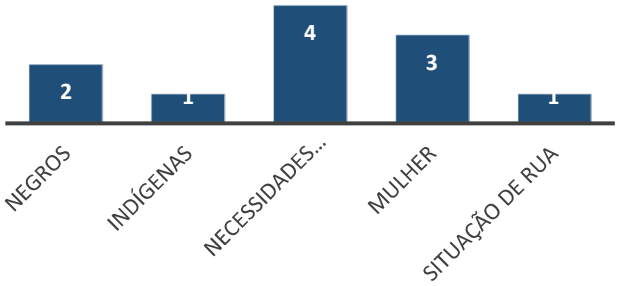


Figura 7 – Sujeitos pertencentes a grupos específicos.

Procuramos levantar o nível de ensino em que as pesquisas estavam inseridas. Não foram observadas pesquisas na educação infantil. Na figura 6, apresentamos os resultados. Foi observada uma pesquisa nos anos iniciais do ensino fundamental. Seis estudos nos anos finais do ensino fundamental, assim como sete estudos no ensino médio e sete no EJA. Isso indica uma preocupação maior das pesquisas nesses três níveis de ensino. Já no ensino superior foram observadas três pesquisas. Observamos uma pesquisa em espaço não formal realizada na ONG NEIMFA. Tivemos trezes casos que não se aplicam, por exemplo, pesquisas documentais, pesquisas bibliográficas, pesquisas com professores e com gestores.

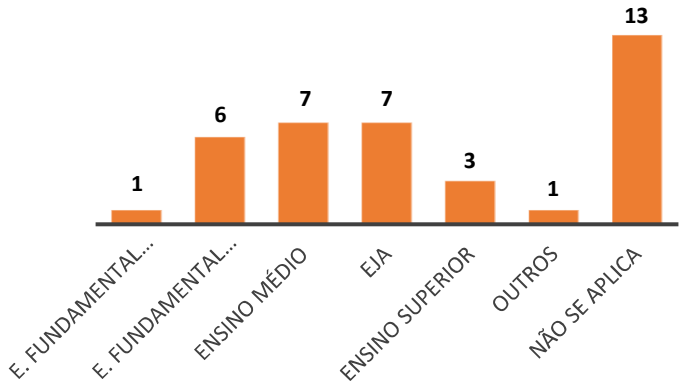


Figura 8 – Nível de ensino das pesquisas.

Na categoria “Perspectiva de resiliência” nos apoiamos em 4 situações: resiliência individual, processual, integral e pesquisas que abordam o levantamento de concepções/posicionamento. Na Tabela 2, temos uma síntese dessas perspectivas. Observamos que a maioria das pesquisas sobre a resiliência na Educação ainda foca na perspectiva individual. Temos na pesquisa o trabalho de Silva (2016) que faz um levantamento em estudantes universitários sobre a espiritualidade e bem-estar espiritual (levantamento), ao mesmo tempo que se apoia em uma abordagem de formação humana integral que associamos a um posicionamento da resiliência integral. Dessa forma, essa dissertação foi pontuada como resiliência integral e ao mesmo tempo como levantamento de concepções/posicionamento.

Perspectivas de resiliência	N	%
Individual	14	42,43
Processual	8	24,24
Integral	4	12,12
Levantamento de concepções/posicionamento	7	21,21
Total	33	100,0

Tabela 2 – Perspectivas da resiliência.

Os resultados que apresentamos sobre diferentes perspectivas de resiliência indicam que ainda temos muitas pesquisas limitadas a uma abordagem individual sobre a resiliência educacional. O nosso fez um levantamento das pesquisas entre os anos de 2008 a 2021. Na pesquisa realizada por Rachman (2008), essa pesquisadora levantou os documentos no Banco de Teses e Dissertações da CAPES, além de bibliotecas digitais na área de resiliência no Brasil, no período de 2002 a 2007, no conceito nas áreas de Educação e Psicologia. Ela verificou como os pesquisadores brasileiros vêm tratando o fenômeno da resiliência nas áreas de Educação e Psicologia da Educação. A pesquisadora constatou que a maior parte das pesquisas produzidas na área de Psicologia com foco na educação entende o construto da resiliência como processo, enquanto as da Educação oscilam entre a resiliência enquanto processo e enquanto perspectiva individual. Comparando os resultados que apresentamos com a pesquisa de Rachman (2008) nos anos anteriores, temos uma indicação que as perspectivas individualistas da resiliência continuam fortes na área educacional, apesar das limitações dessa visão.

Um outro aspecto que levantamos foi como os autores e autoras classificaram as pesquisas. Muitos classificam as pesquisas em qualitativas, quantitativas e mistas, embora consideremos que muitas pesquisas podem ter o predomínio de uma abordagem sobre outra, mas sempre podemos ter elementos das duas, através da quantificação de elementos existentes. Dessa forma, consideramos a forma como cada autor classificou a sua pesquisa e também quando isso não ocorreu, optamos por considerar como abordagem não indicada. Os resultados indicam o predomínio das pesquisas qualitativas em pesquisas sobre resiliência educacional.

Tipo de abordagem	N	%
Abordagem qualitativa	16	50,00
Abordagem quantitativa	1	3,13
Abordagem mista	5	15,63
Abordagem não indicada	11	31,25
Total	32	100%

Tabela 3 – Abordagem de pesquisa indicada.

Observamos uma grande variedade de metodologias utilizadas nas pesquisas que abordam a resiliência educacional que indicamos na figura 7. Sendo que predomina a pesquisa por questionários e bibliográfica.

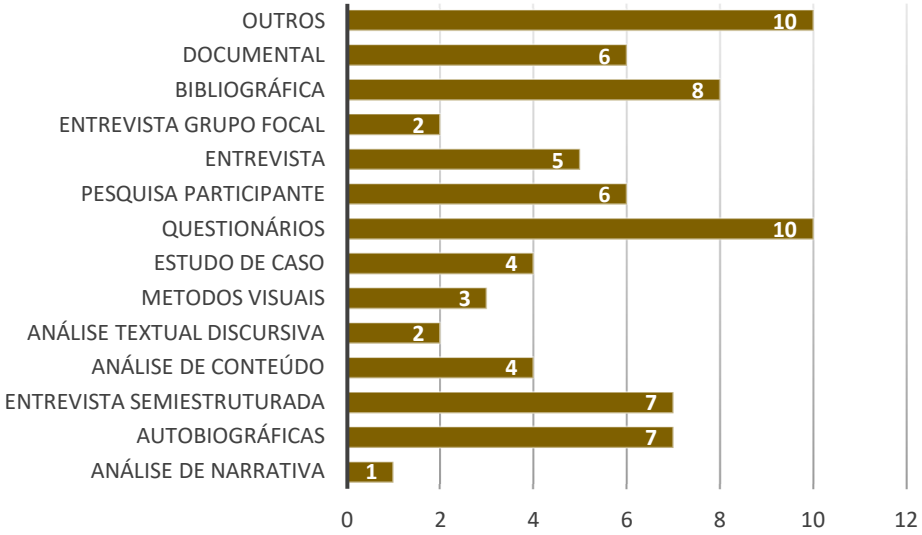


Figura 9 – Metodologias/instrumentos utilizados em pesquisas sobre resiliência educacional.

Algumas pesquisas sobre resiliências adotam diferentes escalas de medida de resiliência, além de outros instrumentos. Procuramos levantar nessas pesquisas esses instrumentos. Os resultados que apresentamos na tabela 4 indicam a variedade dos instrumentos, bem como que não existem a predominância de um em especial. Das 33 pesquisas levantadas, temos que 1/3 usam algum tipo de instrumento de medição das características dos participantes.

	N	%
Escala de Bem-estar Espiritual (EBE)	1	11,11
Escala de Resiliência para Crianças e Adolescentes - ERCA (Sandra Prince-Embury, 2007, traduzida e validada por Barbosa, 2008).	1	11,11
Escala de Medida de Resiliência em Crianças e Jovens (Child and Youth Resilience Measure – CYRM)	1	11,11
Escala de Avaliação das Atribuições de Causalidade para Sucesso e Fracasso Escolar (Boruchovitch & Santos, 2011)	2	22,22
ACE (Adverse Childhood Events)	1	11,11
ScanBullying	1	11,11
Escala de Resiliência: Wagnild e Young (1993)	2	22,22
TOTAL	9	100,00

Tabela 4- Escalas/questionários

3.2 Tratamento com a Análise Estatística Implicativa por meio do software CHIC em sua versão 7.0

No tratamento utilizando o software CHIC em sua sétima versão, foram escolhidas as opções: nós significativos, cálculo longo e cálculo dos intervalos. Foi utilizada a implicação segundo a teoria clássica. Uma outra definição relevante no tratamento é o tipo de lei. Na versão 7.0 do software CHIC temos duas opções: lei binomial e a lei de

Poisson. A escolha dessas leis depende do tamanho da amostra. Como se trata de uma amostra pequena, o recomendado é usar a lei binominal. Nesse sentido temos que:

> a modelagem probabilística da variável aleatória pode ser uma distribuição binomial ou uma de Poisson. Na modelagem com a distribuição de Poisson, supõe-se que o universo de estudo seja infinito numerável, senão tão grande (muitos sujeitos) que ele pode ser considerado como tal, enquanto na modelagem com a distribuição binomial o universo de estudo é finito (Régnier, Andrade, 2020, p. 62).

Para o tratamento com a ASI, utilizamos a árvore coesitiva e o grafo implicativo. No primeiro, usamos com o objetivo de observar a forma como as variáveis se apresentam indicando regras e metarregras. No segundo, usamos o modo cone para focar em algumas variáveis.

3.2.1 Representação pela Árvore coesitiva

Na Figura 8, apresentamos a árvore coesitiva. Consideramos os valores acima de 0,70. Dessa forma, temos um total de 17 classes formadas. Indicamos neste primeiro gráfico com uma linha tracejada o ponto de corte (acima de 0,70) e as classes formadas. Depois fazemos as análises de cada classe formada. O número das classes que apresentamos corresponde à ordem dessas classes apresentadas na árvore coesitiva da esquerda para à direita.

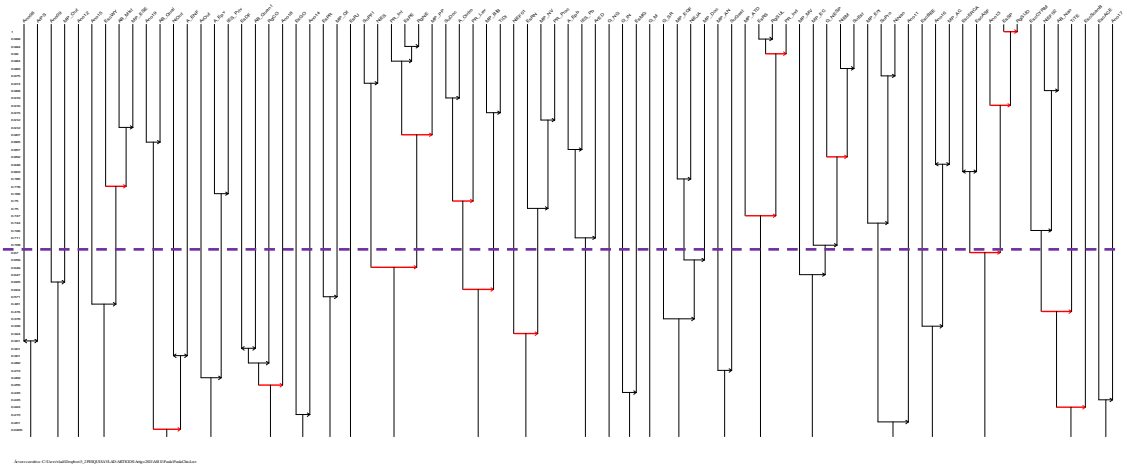


Figura 10 – Árvore coesitiva.

Classe 01: (EscWY ⇒ (AB_Mist⇒MP_ESE)).

Observamos a regra: quem classificou sua pesquisa como de abordagem mista também utilizou entrevistas semiestruturadas (AB_Mist⇒MP_ESE). Por sua vez, temos uma metarregra, indicando que quem utilizou a escala de resiliência desenvolvida por Wagnild e Young, tem uma tendência a utilizar uma abordagem mista e entrevistas semiestruturadas.

Classe 02: (Ano19 ⇒ AB_Qual).

As dissertações/teses no ano de 2019 têm uma tendência de indicar que usaram a abordagem qualitativa.

Destacamos que em função da complexidade de elementos que envolvem a resiliência (sociais, psicológicos, entre outros), existe a necessidade de pesquisas mais robustas e longitudinais, o que poderia indicar a necessidade de uma abordagem dita "mista".

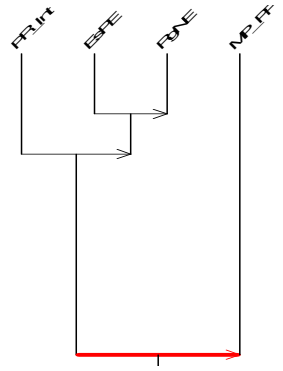
Classe 03: (A_Epv ⇒ IES_Priv)

Esse resultado indica que pesquisas realizadas com sujeitos de uma de instituição de ensino privada (A_Epv) têm uma inclinação a serem realizadas por pesquisadores de uma instituição privada.

Classe 04: (SuPFI ⇒ NiES)

Essa classe apresenta uma relação óbvia e que por isso não traz nenhuma contribuição: Todo professor em formação inicial tem uma tendência de estar em uma instituição de nível superior (o que não poderia ser diferente).

Classe 05 : ((PR_Int ⇒ (EsPE RgNE)) ⇒ MP_PP



Na 5ª classe apresentamos também outras tautologias. Pesquisas no estado de Pernambuco têm uma inclinação de serem realizadas no Nordeste. Por sua vez, temos uma metarregra (regra de regra) que indica que a perspectiva de resiliência integral tem uma tendência de ser realizada em Pernambuco/Nordeste, indicando a força das pesquisas realizadas em Pernambuco e que focam nessa perspectiva. Observa-se também que quem realiza pesquisas na perspectiva da resiliência integral tem uma tendência de utilizar como metodologia a pesquisa participante.

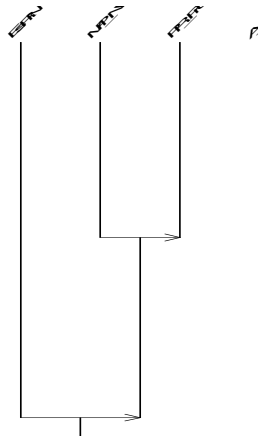
Classe 06: ((SuDoc ⇒ A_Outro) ⇒ PR_Lev

As pesquisas documentais/bibliográficas levantadas não indicaram se a abordagem era qualitativa, quantitativa ou mista. Essa regra aponta para uma metarregra ((SuDoc ⇒ A_Outro) ⇒PR_Lev), indicando que pesquisas documentais com abordagem não definidas não apresentam uma perspectiva de resiliência. Isso se justifica uma vez que se trata de levantamento de concepções/posicionamentos (PR_Lev) e nesse levantamento, podemos ter diferentes perspectivas (no caso de teses e dissertações) e no caso de outros tipos de materiais bibliográficos e/ou documentos, dependendo da natureza dos mesmos, não faz sentido uma classificação da perspectiva de resiliência.

Classe 07: (MP_BIB ⇒ TiDI)

Em uma 7ª classe temos que pesquisas bibliográficas tendem as ser realizadas em dissertações de mestrado (MP_BIB ⇒ TiDI). Essa inclinação é natural, uma vez que as pesquisas sobre resiliência demandam muito tempo, e no caso de mestrado, as pesquisas de natureza bibliográfica podem ser realizadas em um tempo menor. Isso não significa dizer que as pesquisas bibliográficas não são desenvolvidas em pesquisas de doutorado. Uma vez que se pode fazer pesquisas bibliográficas com um nível de aprofundamento que se traduza em pesquisas de doutorado.

Classe 08: (EsRN ⇒ (MP_NV ⇒ PR_Proc).



Numa 8ª classe verificamos uma regra na qual pesquisas que utilizam análises de narrativas de vida e suas trajetórias têm uma tendência a terem uma perspectiva da resiliência enquanto processo. Consideramos bastante pertinente essa relação, uma vez que as características da resiliência das pessoas mudam ao longo do tempo em função dos contextos, relações e ambientes aos quais as pessoas passam. Isso pode ser observado em pesquisas que procuram observar as histórias de vida. A essa regra, temos uma metarregra, indicando as pesquisas no Rio Grande do Norte que tendem a utilizar como metodologia as histórias de vida e uma abordagem da resiliência enquanto processo, (EsRN ⇒ (MP_NV ⇒ PR_Proc).

Classe 09: ((A_Epb ⇒ IES_Pb) ⇒ ArED).

Nessa classe, temos que as pesquisas nas quais os dados são construídos com sujeitos de instituições públicas (A_Epb) tendem a ser realizadas por pesquisadores, vinculados no momento de realização das pesquisas de doutorado/mestrado a instituições públicas (IES_PB). Essa regra tende a ser aplicada em programas da área de Educação. Considerando os dados da estatística descritiva nos dados construídos, a maioria das pesquisas era de programas na área de educação (31 das 33). E a maioria das pesquisas sobre resiliência educacional (na nossa amostra) são realizadas por instituições de ensino pública (26 das 33). Esses dados podem ter influenciado essa tendência.

Classe 10: (MP_EGF ⇒ NiEJA).

Pesquisas que utilizam entrevista com grupo focal têm uma inclinação de serem realizadas pela Educação de Jovens e Adultos (Ni EJA).

Classe 11: (MP_ATD ⇒ ((EsRS ⇒ RgSUL) ⇒ PR_Ind))

Na 11ª classe, observamos uma regra óbvia: pesquisas do estado do Rio Grande do Sul tendem a ser realizadas na Região Sul. Por sua vez, observa-se uma metarregra associando as pesquisas nessa região a pesquisas em uma perspectiva individual de resiliência. Destacamos que, apesar do pioneirismo do Rio Grande do Sul nas pesquisas sobre resiliência no campo da Psicologia, foi identificado nessas últimas pesquisas essa tendência associada a uma perspectiva mais limitada da resiliência no campo educacional. Temos também uma outra metarregra que indica que as pesquisas que utilizam a análise textual discursiva tendem a focar em uma abordagem de resiliência individual e serem realizadas no Rio Grande do Sul (RgSul).

Classe 12: (MP_EC ⇒ (G_NESP ⇒ (NiEM ⇒ SuEst))

Na 12ª classe, constatamos que as pesquisas que utilizam como abordagem metodológica o estudo de caso, têm uma inclinação de serem realizadas com estudantes do ensino médio e esses estudantes têm necessidades especiais.

Classe 13 : (MP_Ent ⇒ (SupPro ⇒ NiNao)

Na 13ª classe, observamos que pesquisas que utilizam a entrevista (MP_Ent) tendem a ser realizadas em pesquisas com professores (SupPro). Quando se trata de pesquisas com professores (SupPro), não relacionamos a um nível de ensino específico (NiNao),

uma vez que os estudantes muitas vezes estão inseridos em um determinado nível escolar (por exemplo, alunos do ensino médio), enquanto o professor já concluiu sua formação inicial, por isso não associamos a um nível de ensino.

Classe 14: (Ano16 ⇔ MP_AC)

Na 14ª classe, temos uma quase equivalência, indicando que as pesquisas concluídas em 2016 tendem a utilizar como ferramenta metodológica a análise de conteúdo. A recíproca é verdadeira, indicando que as pesquisas que utilizaram a análise de conteúdo tiveram uma inclinação a ser realizadas no ano de 2016.

Classe 15: (EscERCA ⇔ EscASF).

Na 15ª classe, temos que as pesquisas que utilizaram a Escala de Resiliência para Crianças e Adolescentes – ERCA (EscERCA) têm uma tendência a utilizar também a escala de Avaliação das Atribuições de Causalidade para Sucesso e Fracasso escolar (EscASF) sendo a recíproca verdadeira.

Classe 16: (Ano13 ⇒ (EsSP ⇒RgSUD)

Na 16ª classe, temos que as pesquisas do ano de 2013 têm uma inclinação a serem realizadas no estado de São Paulo (Região Sudeste).

Classe 17: (EscCYRM ⇒ (NiEF02 ⇒ AB_Não)

Na 17ª classe, observamos que as pesquisas que utilizam a Medida de Resiliência em Crianças e Jovens (CYRM) têm uma tendência a serem realizadas nos anos finais do ensino fundamental e não indicam se a abordagem é qualitativa, quantitativa ou mista. Consideramos que essa indicação é questionável, uma vez que é normal mesmo em pesquisas qualitativas termos elementos que são contáveis e considerados na análise, como por exemplo, a frequência que um determinado sujeito realiza determinada atividade (qualitativa e ao mesmo tempo levando em conta uma frequência – quantitativa). Régnier e Crouzier (2013) fazem uma crítica a essa postura dicotômica assinalando a complementariedade das mesmas.

3.2.2 Representação pelo Grafo implicativo

No uso do grafo implicativo, tal como na árvore coesitiva, selecionamos como valor mínimo 0,70. As variáveis podem ser organizadas no grafo implicativo em até 4 níveis por cores. As cores padrão são: o vermelho, o azul, o verde e o cinza. Organizamos as cores do grafo da seguinte forma:

Vermelha ≥0,94	0,94> azul ≥0,86	0,86> verde ≥0,78	0,78> cinza ≥0,70

Dessa forma, os grafos na cor vermelha indicam uma maior intensidade de implicação com valores de 0,94 e acima (até 1). Na cor cinza, temos as menores intensidades com valores de 0,70 até abaixo de 0,78. Nos níveis intermediários, temos as cores verde e azul.

Selecionamos algumas variáveis para estudar as relações relativas às mesmas. Para tanto, usamos o modo cone no grafo implicativo. No modo cone, podemos selecionar uma ou mais variáveis e em seguida o software gera um grafo indicando apenas as variáveis, que nos níveis de intensidade definidos apresentam relação com as variáveis selecionadas. Na Figura 9, temos o grafo gerado em função da seleção das variáveis relativas aos anos das pesquisas e as variáveis associadas a elas.

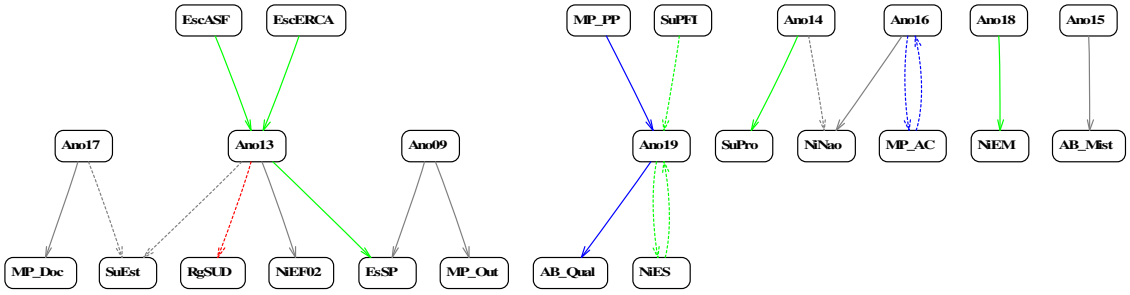


Figura 11 – Grafo Implicativo: modo cone com a seleção dos anos.

No ano de 2017, tivemos uma tendência das teses ou dissertações defendidas nesse ano utilizarem como metodologia a pesquisa documental. Tanto o ano de 2013 como o ano de 2017, observamos o foco em pesquisas realizadas com estudantes. O uso da Escala de Resiliência para Crianças e Adolescentes – ERCA (EscERCA) e da Escala de Avaliação das Atribuições de Causalidade para Sucesso e Fracasso escolar (EscASF) tem uma propensão de se apresentarem em pesquisas defendidas em 2013. Também se observou que as pesquisas concluídas em 2009 e 2013 têm uma tendência de serem realizadas em São Paulo. O grafo também indica uma relação entre o ano de 2013 e as pesquisas nos anos finais do ensino fundamental, indicando nesse ano uma atenção especial para esse nível escolar.

No ano de 2009, observamos uma relação com pesquisas que utilizaram outros métodos de pesquisas, diferente das que classificamos no Quadro 4. No Quadro 4, foram apresentadas 13 ferramentas/métodos de pesquisa. Contudo, foram encontrados outros, indicando a grande variedade empregada em pesquisas sobre resiliência no quadro educacional. Isso pode ser um indicador da complexidade dos fenômenos e o uso de diferentes instrumentos nas pesquisas, procurando diferentes estratégias de investigação.

As pesquisas que utilizaram como metodologia a pesquisa participante e que envolvem professores em formação inicial têm uma tendência de serem apresentadas nas pesquisas concluídas mais recentemente em 2019. No ano de 2019, observamos uma propensão da utilização da abordagem qualitativa. Também observamos uma quase equivalência indicando que as pesquisas em 2019 tendem a ser realizadas no ensino superior e as pesquisas realizadas com estudantes do ensino superior têm uma tendência de serem realizadas em 2019.

As teses/dissertações defendidas no ano de 2014 têm uma tendência de serem realizadas com professores e não entrando dessa forma na classificação de níveis de ensino.

No ano de 2016, temos uma tendência de termos pesquisas não relacionadas a um determinado nível de ensino (que podem ser pesquisas com professores, bibliográficas ou documentais). Também nesse ano, temos uma quase equivalência indicando que pesquisas no ano de 2016 têm uma inclinação de serem realizadas sobre análise de conteúdo e pesquisas com análise de conteúdo tendem a ser realizadas no ano de 2016.

No ano de 2018, temos uma predisposição para pesquisas relativas a estudantes do ensino médio.

No ano de 2015, as teses/dissertações têm uma tendência para indicar abordagens mistas em suas pesquisas.

Usando o cone implicativo, selecionamos as variáveis relacionadas às perspectivas de resiliência. Na figura 10, apresentamos os resultados.

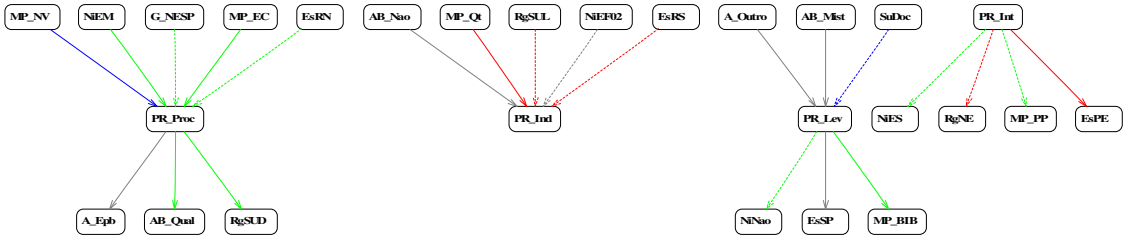


Figura 12 – Grafo Implicativo: perspectiva de resiliência.

Na figura 11, apresentamos as variáveis relativas à perspectiva individual de resiliência. Na abordagem não descrita pelo pesquisador (AB_Nao), o método de pesquisa por questionário (MP_Qt) e o nível de ensino nos anos finais do ensino fundamental (NIEF02). Verificamos nesses casos uma inclinação das pesquisas que apresentaram uma perspectiva individual de resiliência.

Na Região Sul (RgSul), como também no estado do Rio Grande do Sul (EsRS), percebemos uma predisposição nas pesquisas na área de Educação com a resiliência numa perspectiva individual, apesar desse estado ter sido pioneiro nos estudos sobre resiliência no Brasil.

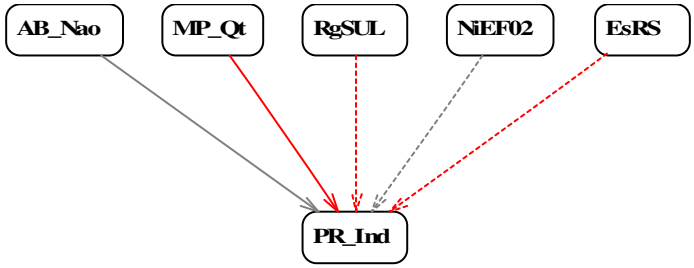


Figura 13 – Variáveis relacionadas à perspectiva individual de resiliência.

Na figura 12, temos as variáveis relacionadas à perspectiva da resiliência enquanto processo. Observamos que as pesquisas que utilizam as análises de narrativas (MP_AN) têm uma tendência a apresentarem uma perspectiva da resiliência enquanto processo. Observamos essa mesma relação na árvore coesitiva e é bastante coerente com a perspectiva de resiliência enquanto processo. Além de ser uma forma de investigação bastante pertinente para essa perspectiva, uma vez que ao longo da vida, em função das teias de apoios, das relações sociais e ambientes por onde se "navegam", as pessoas podem se apresentar mais ou menos resilientes.

Na figura 12, temos também outras relações importantes. As pesquisas no ensino médio (NiEM) com pessoas com necessidades especiais (G_NESP: cego, superdotado etc) que utilizam o estudo de caso como ferramenta metodológica (MP_EC) e no Rio Grande do Norte (EsRN) tendem a utilizar a perspectiva da resiliência enquanto processo. Consideramos esclarecer que tivemos duas pesquisas no Rio Grande do Norte e todas elas classificadas como adotando a abordagem da resiliência enquanto processo. Por isso, temos uma tendência das pesquisas no Rio Grande do Norte serem sobre resiliência enquanto processo (o valor da intensidade não foi mais alta, pois na amostra foram apenas dois casos). Destacamos que a recíproca não é verdadeira, uma vez que tivemos 8

pesquisas classificadas como resiliência enquanto processo e apenas 2 são do Rio Grande do Norte. Das 8 pesquisas com abordagem de resiliência enquanto processo, tivemos 6 da Região Sudeste, levando a que as pesquisas que abordam a resiliência enquanto processo, têm uma tendência de serem produzidas na região Sudeste. Na Região Sudeste, tivemos 15 pesquisas, sendo 6 que classificamos a resiliência enquanto processo, 5 na perspectiva individual, 4 com levantamentos sobre concepções e posicionamentos, que dessa forma não se aplica a uma perspectiva em especial.

Observamos ainda na figura 12 que quem utiliza a perspectiva de resiliência, enquanto processo, tem uma tendência a fazer parte de uma instituição de ensino pública e a informar que utiliza uma abordagem qualitativa.

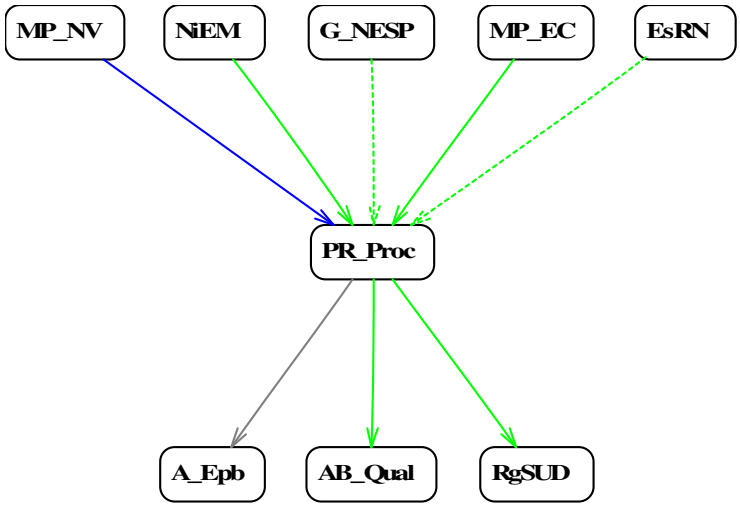


Figura 14 – Variáveis relacionadas à perspectiva de resiliência processual.

Na Figura 13, temos pesquisas que adotam uma perspectiva de resiliência integral e têm uma propensão de serem realizadas no ensino superior, de utilizarem como metodologia a pesquisa participante e serem realizadas em Pernambuco (Nordeste). Destacamos que foram ao todo 4 pesquisas que utilizaram a perspectiva de resiliência integral e todas essas foram realizadas no estado de Pernambuco, o que vem a reforçar essa tendência. Tivemos 5 pesquisas em Pernambuco em que 4 delas têm a perspectiva integral.

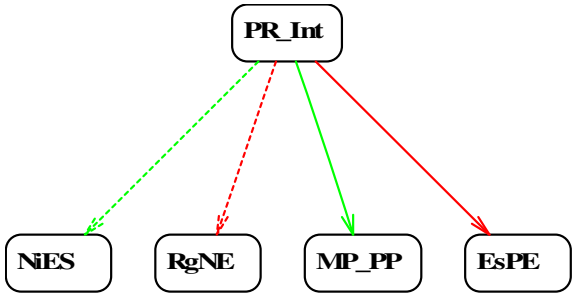


Figura 15 – Variáveis relacionadas à perspectiva de resiliência integral.

Na figura 14, temos as variáveis relativas a situações em que não são pertinentes falar de uma perspectiva de resiliência, uma vez que se trata de levantamento de

concepções/posicionamento. Observamos que nas pesquisas em que classificamos como em um ambiente não identificado (A_Outro), ou seja, não se identificou se é público, privado ou espaços não formais, elas tendem a ser construídas em um levantamento (PR_Lev). Esse resultado é natural, uma vez que as pesquisas com levantamentos não têm uma definição de um ambiente específico de realização. As abordagens mistas (AB_Mist) e que os sujeitos da pesquisa são documentos e/ou livros e/ou teses (SuDoc) tendem a se apresentar em levantamentos de concepções/posicionamentos (PR_Lev).

Verificamos que as situações em que foram classificadas como levantamento de concepções/posicionamentos têm uma tendência de serem em um nível de ensino que não se aplica (NiNao), além de serem realizadas no estado de São Paulo e serem utilizadas a pesquisa bibliográfica como metodologia.

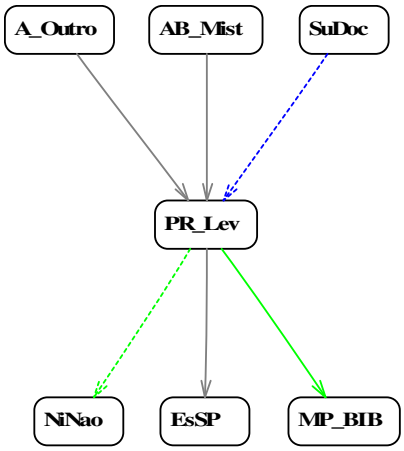


Figura 16 – Variáveis relacionadas à perspectiva de resiliência integral.

4 Considerações finais

O campo de pesquisas na área da resiliência educacional vem crescendo no Brasil. Esta pesquisa apresenta como indicadores as teses e dissertações publicadas na área entre os anos de 2008 e 2021. Observamos uma concentração das pesquisas na região Sudeste, com grande produção em São Paulo, sendo seguida pela região Sul e Nordeste. A região Centro-Oeste apresentou um número menor de pesquisas. O Norte não apresentou pesquisas na amostra levantada, isso indica que essa área do conhecimento tem muito que se expandir. Através do uso da ASI, pudemos observar tendências em relação aos aspectos metodológicos, da abordagem da perspectiva de resiliência, das instituições onde as teses e dissertações foram desenvolvidas, bem como do ambiente onde elas foram realizadas e os seus campos de investigação. Esses resultados indicam a variedade dos métodos, dos sujeitos e das possibilidades de investigação, mas também a necessidade de ampliação dos enfoques de resiliência para perspectivas mais atuais que consideram não apenas o sujeito como único responsável pelo seu desenvolvimento, mas a necessidade de criação de suportes, de redes de apoio, ambientes propícios para o bem-estar que vão muito além de "culpar" os indivíduos pelo seu fracasso, mas de dar o apoio necessário oferecendo os meios para que eles possam se desenvolver plenamente. O bem-estar social é resultante de escolhas individuais e sociais, mas precisa-se do auxílio de políticas públicas adequadas que possam auxiliar os seres humanos a viverem de forma solidária e humana.

Referências

[1] Acioly-Régnier, N. M., Chaves, A. L. L., Araújo, M. N. M. (2011), *Souffrance au travail et résilience de l'enseignant* : analyse de pratiques pédagogiques à l'école primaire au Brésil, *Les Collectifs du Cirp*, 2, 248-257.

[2] Almeida, L. H. D. de (2019), *Autoconhecimento Emocional do Professor: a preocupação com a pessoa, antes do profissional*, Dissertação (Mestrado), Pontifícia Universidade Católica do Rio Grande do Sul.

[3] Andrade, V. L. V. X. de (2020). Apresentação. Dans J.-C Régnier et V. L. V. X. de Andrade (dir.), *Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC*. (p. 41-84). Recife: EDUFRPE, 2020. Récupéré le 18 février 2021 du site de l'Université Fédérale Rural de Pernambuco (Brésil) : http://www.editora.ufrpe.br/ASI.

[4] Assis, N. P. de (2015), *Jovens Negros Trabalhadores*: um estudo sobre trajetórias de escolarização e resiliência na Educação de Jovens e Adultos de Ribeirão das Neves, Dissertação (Mestrado), Centro Federal de Educação Tecnológica de Minas Gerais.

[5] Bady, J. B. (2019), *Violências de Gênero nas Trajetórias de Aprendizagens: alunas em vulnerabilidade social e em situação de rua em uma escola de educação de jovens e adultos*, Ensino fundamental, Porto Alegre, Sul do Brasil, Dissertação (Mestrado), Universidade la Salle.

[6] Brandão, J. M., Nascimento, E. do (2019), Resiliência psicológica: da primeira fase às abordagens baseadas em trajetória, *Memorandum*, 36.

[7] Brandão, J. M., Mahfoud, M., Gianordoli-Nascimento, I. F. (2011). A construção do conceito de resiliência em psicologia: discutindo as origens, *Paidéia*, maio-ago. 21(49), 263-271.

[8] Brandão, J. M. (2009). *Resiliência: de que se trata?* O conceito e suas imprecisões, Dissertação (Mestrado), Universidade Federal de Minas.

[9] Brasil, T. L. (2015), *Promoção de Resiliência em Alunos do Curso de Pedagogia da UFPE*: uma análise do Programa Adulto: resiliência ao longo da vida, Dissertação (Mestrado), Universidade Federal de Pernambuco.

[10] Brasil, T. L. (2019), *Resiliência integral*: um caminho de possibilidades para formação humana de futuros docentes, Tese (doutorado), Universidade Federal de Pernambuco.

[11] Brasil, T. L.; Ferreira, A. L. (2020), Resiliência integral: uma experiência na formação humana de futuros docentes. Dans A. S. G. Pessoa; S. H. Koller (dir.). *Resiliência & educação : perspectivas teóricas e práticas* (p. 275-294). 1ed. São Paulo: Vetor Editora.

[12] Bulhões, P. F. (2016), *O Cinema e a História de Vida: representações de um estudante com altas habilidades/superdotação acerca da identidade resiliente*, Dissertação (Mestrado), Universidade Federal de Santa Maria.

[13] Corrêa, Á. L. M. O. (2019), *Percursos de resiliência e identidade em histórias, memórias e experiências de alfabetizadores (as) indígenas em Roraima*, Tese (doutorado), Universidade Federal de Juiz de Fora.

[14] Costa, F. R. C. P. (2013), *Fatores estressores na Atividade Docente*, Dissertação (Mestrado), Universidade de Taubaté.

[15] Cruz, E. D. G. da (2018), *Experiências da Escola: além dos limites da cegueira,* Dissertação (Mestrado), Universidade Federal do Rio Grande do Norte.

[16] Ferreira, A. L. (2007), *Do entre-deux de merleau-ponty à atenção consciente do budismo e da abordagem transpessoal: análise de uma experiência de formação integral*, Tese, Universidade Federal de Pernambuco.

[17] Ferreira, A. L.; Acioly-Régnier, N. M.; Cunha, D. P. ; Oliveira, V. P. B; Silva, S. C. R. Da; Silva, S. C. R. ; Eskinazi, P. R. V. (2018), *As adolescências periféricas: uma pesquisa-ação integral/transpessoal*. 1. ed. São Paulo: APRIS.

[18] Ferreira, C. M. do C. (2018), *Jovens expostos à violência intrafamiliar e percepções sobre a escola*: *um estudo comparativo*, Dissertação (Mestrado), Universidade Católica de Brasília.

[19] Francisco, M. V. (2013), *A Construção Social da Personalidade de Adolescentes Expostos ao Bullying Escolar e os Processos de "Resiliência Em-Si": Uma Análise Histórico-Cultural*, Tese (doutorado), Universidade Estadual Paulista Júlio de Mesquita Filho.

[20] Garcia, N. R. (2013), *Atribuições de causalidade para sucesso e fracasso escolar e resiliência em estudantes do ensino fundament*al, Dissertação (mestrado), Universidade Estadual de Campinas.

[21] Gras, R.; Régnier, J.-C. (2015) Origem e Desenvolvimento da Análise Estatística Implicativa (A.S.I.). In: VALENTE, J. A., ALMEIDA, M. E. B. (Org.). *Uso do CHIC na Formação de Educadores*: à guisa de apresentação dos fundamentos e das pesquisas e foco. Rio de Janeiro: Letra Capital, 2015.

[22] Grotberg, E. H. (2005), Introdução: novas tendências em resiliência. Dans A. Melilo, E. N. S. Ojeda e cols. *Resiliência: descobrindo as próprias fortalezas*. Porto Alegre: Artmed.

[23] Infante, F. (2005), A resiliência como processo: uma revisão da literatura recente. Dans A. Melillo, E. N. S. Ojeda e cols. *Resiliência: descobrindo as próprias fortalezas*. V. Campos, trad. p. 23-38. Porto Alegre: Artmed.

[24] Junior, J. F. B. (2009), *Resiliência e sucesso escolar*: um estudo de caso com afrodescendente no ensino público, Dissertação (Mestrado), Centro Universitário Moura Lacerda (Ribeirão Preto).

[25] Machado, L. A. B. (2014), *Mal-estar/bem-estar e profissionalização docente: um estudo de produções acadêmicas brasileiras*, Dissertação (Mestrado), Universidade Federal de Goiás.

[26] Marochi, A. C. A. (2017), *Escolarização e o Tempo na Vida das Trabalhadoras/Estudantes do Ceebja de Irati-PR*, Dissertação (Mestrado), Universidade Estadual do Centro-Oeste.

[27] Muller, J. L. (2019), *Formação de Professores na Unemat: um estudo do curso de pedagogia para educadores do campo*, Tese (doutorado), Universidade Federal do Rio Grande do Sul.

[28] Pereira, H. R. S. (2013), História de Vida de uma Aluna com Baixa Visão: Construção de Itinerários de Aprendizagem em Cursos Técnicos Agrícolas, Dissertação (Mestrado), Universidade Federal Rural do Rio de Janeiro.

[29] Pereira, K. A. F. (2009), *Protagonismo juvenil e educação da juventude no ensino médio brasileiro*, Dissertação (Mestrado), Universidade Estadual Paulista.

[30] Porcaro, R. C. (2011), *Caminhos e desafios da formação de educadores de jovens e adultos*, Tese (doutorado), Universidade Federal de Minas Gerais.

[31] Rachman, V. C. B. (2008), *Resiliênci*a: o emprego desse conceito nas áreas de educação e psicologia da educação no Brasil (2002 a 2007), Dissertação (mestrado), Pontifícia Universidade Católica de São Paulo.

[32] Régnier, J.-C. et Andrade, V. L. V. X. de (2020). A Análise Estatística Implicativa e Análise de Similaridade. Dans J.-C Régnier et V. L. V. X. de Andrade (dir.), Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC. (p. 41-84). Recife: EDUFRPE, 2020. Récupéré le 12 février 2021 du site de l'Université Federal Rural de Pernambuco (Brésil) : http://www.editora.ufrpe.br/ASI.

[33] Régnier, J.-C; Crouzier, M.-F. (2013), Comprendre les liens professionnels entre le Rased et les équipes éducatives: intérêt des traitements statistiques fondés sur la complémentarité du quantitatif et du qualitatif. *Conjectura: Filos. Educ*., 18 (1), 179-199.

[34] Rezende, E. V. dos S. (2016), *Resiliência no Contexto da Prática Pedagógica*: um estudo com professoras de escolas públicas municipais no Recife-PE, Tese (doutorado), Universidade Federal de Pernambuco.

[35] Rodrigues, M. de O. (2015), *Ser Docente na Gestão e Ser Gestor na Docência:* um olhar sobre a iniciação à carreira no contexto universitário, Universidade Federal de Santa Maria.

[36] Rutter, M. (1987), Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, 57 (3), 316-331.

[37] Rutter, M. (1993), Resilience: some conceptual considerations. *Journal of Adolescent Health*, 14, 626-631.

[38] Sakotani, V. H. (2016), *Resiliência e Educação: um balanço das produções científicas de 2000 a 2013*. Dissertação (Mestrado), Universidade Estadual Paulista Júlio de Mesquita Filho.

[39] Salles, R. R. (2014), *Resiliência em docentes*: sentido e significado na prática profissional no contexto da educação básica), Dissertação (mestrado), Universidade de Taubaté.

[40] Silva, L. A. C. da (2016), *Espiritualidades e Bem-Estar Espiritual no Processo Formativo de Estudantes de Psicologia do Recife – PE à Luz da Abordagem*

*Integral/Transpessoa*l, Dissertação (Mestrado), Universidade Federal de Pernambuco

[41] Silva, A. C. da. (2012), *Gestão educacional resiliente: uma proposta para um contexto de mudanças paradigmáticas*. Tese (doutorado), Universidade Estadual de Campinas, Faculdade de Educação, Campinas, SP. Disponível em: <http://www.repositorio.unicamp.br/handle/REPOSIP/250781>. Acesso em: 21 ago. 2018.

[42] Silva, M. L. F. da. (2019), *Resiliência integral e juventudes* periféricas: análise de uma experiência formativa no campo educacional. Dissertação (mestrado), Universidade Federal de Pernambuco.

[43] Silva, S. C. do C. (2019), *Narrativas Individuais… Memórias Coletivas: Diálogo, Resiliência e Transformação*, Dissertação (Mestrado), Universidade do Estado do Rio Grande do Norte.

[44] Tatiane S. R. R. (2014), *Interesse por matemática, engajamento e a experiência de fracasso esc*olar: um estudo com adolescentes, Dissertação (Mestrado), Universidade Estadual de Londrina.

[45] Trevisan, N. V. (2014), Ambiência [Trans] Formativa na Educação Superior: Processos de Resiliência no Início da Carreira Docente, Dissertação (Mestrado), Universidade Federal de Santa Maria.

[46] Ungar M. (dir.) (2020), *Multisystemic Resilience*: *Adaptation and Transformation in Contexts of Change*. New York, NY: Oxford University Press.

[47] Ungar, M. (2018a), What Works *A Manual for Designing Programs that Build Resilience*, Resilience Research Centre, Halifax.

[48] Ungar, M. (2018b), Systemic resilience: principles and processes for a science of change in contexts of adversity. Ecology and Society 23(4):34. https://doi.org/10.5751/ES-10385-230434.

[49] Vergnaud, G. (1994). Le rôle de l'enseignant à la lumière des concepts de schème et de champ conceptuel. In R. Artigue, R. Gras, C. Laborde & P. Tavignot (éd.), *Vingt ans de didactique des mathématiques en France*. Paris : La pensée sauvage.

[50] Weschenfelder, G. V. (2017), Os Super-Heróis das Histórias em Quadrinhos como Recursos para a Promoção de Resiliência em Crianças e Adolescentes em Situação de Risco, Tese (doutorado), Universidade la Salle.

[51] Yanaga, T. W. (2017), *Inclusão escolar e processos de resiliência em adolescentes e jovens da Educação Especial*, Tese (doutorado), UNESP.

COMPETÊNCIAS SOCIOEMOCIONAIS (CSE) NA EDUCAÇÃO: TENDÊNCIAS DA PRODUÇÃO CIENTÍFICA NAS EDIÇÕES DO CONEDU 2014-2020 À LUZ DA A.S.I.

Cristiana Marinho Da Costa[1], Quércia Carvalho Eloi[2], Vladimir Lira Véras Xavier De Andrade[3], Jean-Claude Régnier[4].

COMPÉTENCES SOCIO-ÉMOTIONNELLES (CSE) EN ÉDUCATION : TENDANCES DE LA PRODUCTION SCIENTIFIQUE DANS LES ÉDITIONS CONEDU 2014-2020 À LUMIÈRE DE L'A.S.I.

SOCIO-EMOTIONAL SKILLS (CSE) IN EDUCATION: SCIENTIFIC PRODUCTION TRENDS IN CONEDU 2014-2020 EDITIONS THE LIGHT OF A.S.I.

RESUMO

Diversos autores, a exemplo de Goleman (2011), Bisqueira (2000), Casassus (2009) e Gonsalves (2015) têm destacado o papel das competências socioemocionais e a sua importância no desenvolvimento de diferentes habilidades. A importância dessas habilidades vem norteando orientações e pesquisas na área de Educação. Neste artigo, apresentamos o resultado de uma pesquisa sobre as tendências da educação socioemocional sinalizadas nos trabalhos do evento CONEDU no recorte temporal de 2014 a 2020. A metodologia foi baseada em pesquisas bibliográficas com recorte para os trabalhos que envolvem a temática da educação socioemocional. Os dados obtidos foram analisados utilizando a abordagem da Análise Estatística Implicativa – A.S.I. utilizando o software CHIC 7.0. Essas análises foram modelizadas com esse software e articuladas à hermenêutica interpretativa dos sujeitos pesquisados que nos levam à compreensão da produção teórica do fenômeno em estudo, tendo como base as relações de quase-implicação entre as variáveis na perspectiva de entender as tendências, lacunas e vieses sobre o tema abordado.

Palavras-chave: *Competências Socioemocionais, Tendências, Análise Estatística Implicativa.*

RÉSUMÉ

Plusieurs auteurs, tels que Goleman (2011), Bisqueira (2000), Casassus (2009) et Gonsalves (2015) ont mis en évidence le rôle des compétences socio-émotionnelles et leur importance dans le développement de différentes compétences. L'importance de ces compétences a guidé les lignes directrices et la recherche dans le domaine de l'éducation. Dans cet article, nous présentons le résultat d'une enquête sur les tendances de l'éducation socio-émotionnelle signalées dans les travaux de l'événement CONEDU sur la période 2014-2020. La méthodologie était basée sur une recherche bibliographique en mettant l'accent sur les travaux impliquant le thème de l'éducation socio-émotionnelle. Les données obtenues ont été analysées selon l'approche d'Analyse Implicative Statistique – A.S.I. en utilisant le logiciel

1 Universidade Federal Rural de Pernambuco, Recife (PE) - Brasil, cmcmarinhos@gmail.com

2 Secretaria de Educação de Pernambuco, Garanhuns (PE) - Brasil, querciacarvalho2012@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 Professeur des universités émérite. Membre du laboratoire UMR 5191 ICAR http://www.icar.cnrs.fr/membre/jcregnier/, Université Lumière Lyon 2. Professeur invité à l'Université d'Etat de Tomsk (Russie) – National Research Tomsk State University– UMR5191 – ICAR Université Lumiere Lyon 2 (FRA) jean-claude.regnier@univ-lyon2.fr

CHIC 7.0. Ces analyses ont été modélisées avec ce logiciel et articulées à l'herméneutique interprétative des sujets de recherche qui nous conduisent à une compréhension de la production théorique du phénomène étudié, basée sur les relations de quasi-implication entre les variables dans la perspective de compréhension des tendances, lacunes et biais sur le sujet traité.

Mots-clés : *Compétences socio-émotionnelles, Tendances, Analyse statistique implicative.*

ABSTRACT

Several authors, such as Goleman (2011), Bisqueira (2000), Casassus (2009) and Gonsalves (2015) have highlighted the role of socio-emotional skills and their importance in the development of different skills. The importance of these skills has been guiding guidelines and research in the field of Education. In this article, we present the results of a research on the trends in social-emotional education signaled in the works of the CONEDU event in the time frame from 2014 to 2020. The methodology was based on bibliographical research with a focus on works involving the theme of social-emotional education. The data obtained were analyzed using the Statistical Implicative Analysis approach – A.S.I. using CHIC 7.0 software. These analyzes were modeled with this software and articulated to the interpretative hermeneutics of the researched subjects that lead us to an understanding of the theoretical production of the phenomenon under study, based on the quasi-implication relationships between the variables in the perspective of understanding trends, gaps and biases on the topic covered.

Keywords: *Socioemotional Skills, Trends, Statistical Implicative Analysis.*

1 Introdução

A abordagem acerca das Competências Socioemocionais (CSE) vem se intensificando nos últimos anos e notadamente veio à tona no cenário atual de pandemia em que se fez presente nos mais variados contextos educacionais, profissionais e sociais.

As competências socioemocionais (CSE) caracterizam-se como um conjunto de habilidades que promovem autorregulação de sentimentos, espírito de equipe, resiliência, empatia para atuar na sua vida social e profissional levando o indivíduo a conviver com as diferenças, aberto a novas experiências, entre outras habilidades que unidas podem contribuir para uma sociedade muito melhor.

Essas competências envolvem processos que auxiliam o ensino-aprendizagem do aluno no âmbito do currículo e tem como objetivo abrir reflexão para aplicação de atitudes significativas, não só para a vida acadêmica, mas como a pessoal em que buscam competências como resiliência, atividades grupal e principalmente a gestão de sentimentos. Na contemporaneidade não é admissível reduzir a aprendizagem do aluno somente à avaliação e mensurações cognitivas, pois os alunos são seres humanos com emoções que se conectam com professores, amigos de sala, família. (Abed, 2016). As competências socioemocionais permitem desenvolver nos indivíduos habilidades para gerenciar, autorregular emoções numa proposta de empatia, espírito de equipe, solidariedade, resiliência, ética, cidadania. Valores esses tão desejáveis nas redes de relações e necessários atualmente. Isso é destacado por autores como Goleman (2011), Bisqueira (2000), Casassus (2009), Vallés (2003), Gonsalves (2015) que ressaltam a relevância da CSE na educação e a sua inserção curricular.

A Base Nacional Comum Curricular (BNCC) aprovada no final de 2017 veio para impulsionar ainda mais o compromisso educacional com o desenvolvimento integral dos estudantes baseando-se em pilares essenciais que vão desde a educação básica até o

ensino médio. Dessas, incluem dez competências gerais que têm por premissas a articulação da construção de conhecimentos, o desenvolvimento de competências na formação de valores e atitudes, e deverão ser ensinadas, praticadas e estimuladas nas instituições escolares num viés de formação integral e não meramente cognitiva.

Nessa perspectiva, o referido trabalho se justifica pela importância de saber o que emerge da produção científica do evento CONEDU no que diz respeito às CSE, em relação a seus vieses, lacunas e tendências, no recorte temporal de 2014-2020. Destacamos que o evento trouxe em sua edição de 2017 o primeiro GT (Grupo de Trabalho) sobre Educação Emocional, permitindo um espaço de diálogo e legitimação para avanços na temática, não excluindo e inclusive reforçando a possibilidade de trabalhos em outros GT's, haja vista que a temática apresenta transversalidade perpassando por diversas áreas educacionais.

Nesse contexto buscando compreender as tendências que emergiram, assim como, discussões e problematizações, temas abrangentes, origem e localizações, tipos de pesquisas e metodologias utilizadas, acessamos os anais do CONEDU existentes desde sua primeira versão em 2014 até a sua última versão na modalidade on-line adotada em 2020. Devido ao cenário pandêmico, buscamos por trabalhos aglutinados na categoria temática "Competências Socioemocionais/ Habilidades Socioemocionais" como ponto de partida. Foram localizados, no universo da amostragem, cerca de 27 trabalhos referentes a temática constituindo o corpus de análise documental da referida pesquisa.

A referida pesquisa apresenta por metodologia baseada em pesquisas bibliográficas, documentais, através do viés qualitativo com abordagem no "estado da arte" ou "estado do conhecimento" buscando encontrar lacunas, tendências, vieses acerca do contexto abordado.

Espera-se que este trabalho seja mais uma contribuição para o fortalecimento das discussões sobre as Competências Socioemocionais na esfera educacional. Não tendo pretensões de esgotar o tema, mas construir um percurso de investigação que acene para novos questionamentos, produções e indagações sobre essa abordagem tão relevante no cenário atual.

2 Pressupostos teóricos nas pesquisas acerca das Competências Socioemocionais na Educação

Ao pensar a práxis educacional, observa-se alguns elementos que embasam a educação socioemocional como a: carência de afeto, empatia e sensibilidade nas relações interpessoais no âmbito escolar e social. Além da desvalorização da educação e da instituição escolar somadas à crise social que levam o mundo a contextos de violência e competição. Ferreira e Acioly-Régnier (2010, p. 22) destacam que "a questão da relação entre cognição e afetividade, ao lado da violência, sem dúvida, constitui um ponto focal da agenda educativa no século XXI". Nesse sentido, esses pesquisadores ressaltam a relevância da afetividade e sua relação com cognição e como ela influencia na experiência educacional tomando por base casos de alunos em situação de risco. Essas questões nos conduzem a evidenciar a importância da educação socioemocional e a levantar alguns pressupostos teóricos nessa área.

Wallon traz três aspectos importantes para o desenvolvimento da criança: motor, afetivo e cognitivo. Esses aspectos agem de maneira integrada e complementar entre si (Dantas, 1992). Logo, não se pode falar em desenvolvimento cognitivo sem trabalhar as emoções. Além do desejo e curiosidade de aprendizado que leva o indivíduo a querer buscar o conhecimento. Nesse sentido, consideramos o afeto como a mola propulsora para provocar o desejo de aprender no educando.

Os pressupostos educativos de Wallon são endossados por Mahoney e Almeida (2009): o motor, o afetivo, o cognitivo, a pessoa, embora cada um desses aspectos tenha identidade estrutural e funcional diferenciada, eles estão tão integrados que cada um é parte constitutiva dos outros. Sua separação se faz necessária apenas para a descrição do processo. Uma das consequências dessa interpretação é de que qualquer atividade humana sempre interfere em todos eles. Portanto, a atividade motora

> tem ressonâncias afetivas e cognitivas; toda disposição afetiva tem ressonâncias motoras e cognitivas; toda operação mental tem ressonâncias afetivas e motoras. E todas essas ressonâncias têm um impacto no quarto conjunto: a pessoa, que, ao mesmo tempo em que garante essa integração, é resultado dela (Mahoney e Almeida, 2008, p. 15).

Freire (1992) destaca a importância da amorosidade na relação de ensino-aprendizagem. A amorosidade e o diálogo configuram-se como elementos indispensáveis para que ocorra, no processo educativo, "o encontro amoroso entre os homens que, mediatizados pelo mundo, o 'pronunciam', isto é, o transformam, e, transformando-o, o humanizam para a humanização de todos" (Freire, 1992, p. 43).

Piaget (1993) pontua os aspectos cognitivos e afetivos na conduta humana. Apesar de não trazer tal abordagem em suas investigações, sempre destacou a importância do desenvolvimento afetivo. Reforçando que a afetividade caberia então como fonte de energia da qual dependeria o funcionamento da inteligência (Piaget, 1993, p. 188).

Libâneo (1994) concebe a relação docente/discente como uma condição importante e favorável à aprendizagem para que essa aprendizagem seja significativa, sendo assim, o centro de todo processo de ensino e aprendizagem do aluno.

Para Vygotsky (2003), o contexto emocional está em conexão com outros processos psicológicos de forma geral, ao desenvolvimento da consciência alegando que quem separa o pensamento do afeto nega a possibilidade de estudar a influência inversa do pensamento no plano afetivo, demonstrando que o indivíduo é resultado do processo afetivo, mental, cognitivo físico, interno e externo caracterizando, como um sujeito sócio-histórico.

De fato, o ser humano é fruto de suas emoções, logo a busca do equilíbrio entre mente e corpo é necessário. Assim, se torna essencial o conhecimento das suas emoções e a descoberta de si mesmo (Casassus, 2009).

Na busca por si mesmo, os indivíduos são movidos por vivências e relações ao longo da vida e desfrutam das mais variadas emoções. Nesse sentido, faz-se necessário desenvolver a capacidade de regulação emocional em que o reconhecimento, a compreensão e a percepção dessas emoções serão determinantes para prevenir efeitos nocivos na vida pessoal e profissional (Alzina, 2005, 2003, 2017).

Segundo Goleman (2011), as emoções são impulsionadas pelos legados e evolução na ocorrência de uma ação imediata e instantânea que busca lidar com a vida. Logo, a

capacidade da autoconsciência, controle de impulsos, resiliência, empatia e habilidade social denominam-se inteligência emocional.

Nesse viés, temos a educação emocional (EE) que é caracterizada como uma área pautada numa perspectiva de quebra de engessamento e falta de afetividade presente nos espaços sociais. As emoções passam a ser vistas como primordiais para o desenvolvimento integral de todos os indivíduos, em especial os que são historicamente vítimas de exclusão e marginalização social (Bisqueira, 2000).

Nesse contexto, a educação emocional vem sendo compreendida como processo educativo contínuo e permanente em que um conjunto de necessidades sociais são desenvolvidas nos estudantes não como complemento da educação cognitiva, mas como uma face da formação humana formado organicamente com um de vários elementos que constitui a unidade do ser (Vallés, 2003).

A educação emocional traz consigo esse processo de construção humana que ocorre de forma integralizada, sendo complexo e amplo, pois tem propósito de potencializar o desenvolvimento emocional através do autoconhecimento das emoções.

Esse processo não permeia o controle emocional, mas educativo na perspectiva de desenvolver habilidades que proporcionem identificação de situações pessoais e subjetivas, buscando ampliar a automotivação diante da vida pessoal e profissional (Gonsalves, 2015).

A educação emocional auxilia no desenvolvimento das competências socioemocionais numa perspectiva de autoconsciência, regulação de emoções, inteligência emocional e autogestão de sentimentos com o objetivo de valorizar aspectos emocionais e sociais na vida do sujeito em sociedade, criando uma nova forma de estruturar e formar o indivíduo para as diversas mudanças que ocorrem em nossa sociedade, almejando o bem estar individual e coletivo (Alzina, 2003).

Assim, esse desejo de bem estar, sobretudo coletivo, é imprescindível e vital, pois somos sujeitos planetários. Logo, a ideia do processo auto-eco-organizacional leva à condição indispensável e abertura ao ecossistema do qual se nutre e também se transforma. Nesse cenário, a educação deve favorecer à aptidão para resolver os problemas e estimular o pleno emprego das inteligências, pois ocorre emergência vital de educar para a era planetária e atender as demandas do século XXI, entre essas estão: reforma do modo do conhecimento, do pensamento e do ensino levando a “repensar o pensamento” (Morin, 2003, 2014).

A importância e relevância da educação emocional é fato para o modelo de educação para o século XXI, mas refletimos sobre a formação docente, do preparo desses educadores para essas demandas, dos modelos de formação de professores atualmente pautados em sua maioria em metas, modelos tradicionais e cumprimentos burocráticos.

De acordo com Pimenta (2006), é de extrema urgência a reformulação da estrutura organizacional de formação de profissionais da educação, incluindo o papel das IES e a inserção curricular numa perspectiva diferenciada de ampliar conhecimentos, habilidades, procedimentos, crenças, atitudes e valores com objetivo de diversificação da ação pedagógica na sociedade (Libâneo, 2001).

3 Modelagem das variáveis e construção dos dados

Este estudo teve por objetivo analisar as tendências dos trabalhos que enfocam as competências socioemocionais na educação. A seguir, apresentamos os procedimentos metodológicos utilizados nessa investigação.

3.1 Construção da Amostra

Neste contexto e buscando compreender as discussões que estão sendo construídas na interface das Competências Socioemocionais na Educação como ponto de partida, foram localizados no universo de 21.513 trabalhos, cerca de 27 deles referentes aos descritores "Competências Socioemocionais" e "Habilidades Socioemocionais". Esse constituiu o corpus da nossa análise documental num recorte temporal de 2014-2020, porém a temática em questão da presente pesquisa foi contemplada no evento CONEDU, a partir de 2017-2020. O presente trabalho assumiu o desenho de uma pesquisa documental cujo corpus de análise foi composto pelos trabalhos completos disponíveis nos anais extraídos de endereço eletrônico do evento na internet.

Como se trata de abordagem qualitativa de cunho exploratório, através da qual se buscou a compreensão de realidades, seus significados não se limitaram ao quantitativo, pois se trabalha com o universo de motivos, demandas, aspirações, valores e atitudes (Minayo, 2011).

3.2 Quadro teórico do tratamento e da análise dos dados

Esta etapa consistiu na utilização do quadro teórico da análise estatística implicativa, sinalizando tendências por meio das relações de quase-implicação "causa-efeito" que nos permitem fazer inferências através dos descritores.

O conceito de implicação estatística ou, mais precisamente, o conceito de "quase-implicação" serve para diferenciar da implicação lógica dos domínios da lógica e da matemática. Esse conceito de quase-implicação encontra-se centralizado na construção de ferramentas teóricas que instrumentam um método de análise de dados.

3.3 Construção das variáveis

Temos como objetivo nesta seção descrever como foram construídas as variáveis e apresentar um quadro com as variáveis encontradas e uma descrição delas.

Descrição		Variável
Ano	Publicações de 2014	A_2014
	Publicações de 2015	A_2015
	Publicações de 2016	A_2016
	Publicações de 2017	A_2017
	Publicações de 2018	A_2018
	Publicações de 2019	A_2019
	Publicações de 2020	A_2020
Tema/Abrangência	Percepções/Sentidos_CSE	TM_PERC_SENT_CSE
	Currículo_CSE	TM_CURR_CSE
	Gestão Escolar_CSE	TM_GEST_ESC_CSE
	Violência Escolar_CSE	TM_VIOL_ESC_CSE
	Educação Profissionalizante_CSE	TM_EDUC_PROFIS_CSE
	Psicologia/Transtornos/Síndromes_CSE	TM_PSIC_TRANST_SIND_CSE
	Formação Docente_CSE	TM_FD_CSE
	Prática Docente_CSE	TM_PT_DC_CSE
Grupo de Trabalho	Educ. Emocional	GT_EE
	Educ.Especial	GT_EESP
	Didática/Currículo/Política Educacional	GT_DID_CUR_POL_CSE
	Educ. Infantil	GT_EI
	Formação de Professor	GT_FP
Natureza da Pesquisa (Metodologia)	Pesquisa Participante	METOD_PESQ_PARTIC
	Pesquisa Bibliográfica/ Documental	METOD_PESQ_BIBL_DOC
	Observação	METOD_OBS
Tipo IES	Pública	IES_PUB
	Particular	IES_PART
Regiões (Estado)	Pernambuco	EST_PE
	Alagoas	EST_AL
	Paraíba	EST_PB
	Rio Grande do Norte	EST_RN
	Sergipe	EST_SE
	Ceará	EST_CE
	Rio de Janeiro	EST_RJ
	Goiás	EST_GO
	Tocantins	EST_TO
	Paraná	EST_PR
	Minas Gerais	EST_MG

Quadro 1 – Descrição das variáveis construídas

3.4 Procedimentos de tratamento das relações entre variáveis por meio da A.S.I

Nesta pesquisa, faremos uso do software CHIC (versão 7) com o objetivo de tratar os dados construídos a partir das unidades da amostra. Essa ferramenta complementará a análise descritiva e qualitativa, contemplando conexões, aproximações, interligações e distanciamentos, através dos resultados gerados pela aplicação do software, principalmente às categorias.

O software (CHIC 7.0) tem a função de extrair implicações através de sujeitos e variáveis que se cruzam e que no contexto teórico da ASI são organizados em gráficos e hierarquias. Consequentemente, esses gráficos ajudam o pesquisador na articulação das relações entre variáveis e sujeitos que de acordo com o referencial utilizado, podem fazer emergir novos conhecimentos que às vezes estão escondidos num universo abrangente de dados (Valente; Almeida, 2015).

Como destacam Régnier e Andrade (2020a, p. 85), o software CHIC possibilita realizar "análises de dados multidimensionais, através da classificação hierárquica de similaridade, a análise implicativa e a hierarquia implicativa, além de outros tratamentos dos dados". Nesta pesquisa utilizaremos a Análise Estatística Implicativa que possibilita observar as relações de implicações e apresentar os dados em um grafo implicativo.

4 Análise dos resultados obtidos por tratamentos com o software CHIC (versão 7)

Para o tratamento utilizando o software CHIC, na sua versão 7.0, foi escolhido as opções "Nós significativos", cálculo longo e cálculo dos intervalos. Utilizou-se a implicação segundo a teoria clássica. O CHIC, em sua versão 7 apresenta como tipo de lei a Lei Binominal e a Lei de Poisson. Na escola entre esses duas leis, deve levar em conta que:

> a modelagem probabilística da variável aleatória pode ser uma distribuição binomial ou uma de Poisson. Na modelagem com a distribuição de Poisson, supõe-se que o universo de estudo seja infinito numerável, senão tão grande (muitos sujeitos) que ele pode ser considerado como tal, enquanto na modelagem com a distribuição binomial o universo de estudo é finito (Régnier e Andrade, 2020b, p. 62).

Em função do número limitados pequeno da amostra, foi escolhido para o tratamento a modelagem probabilística da variável aleatória utilizando uma distribuição binominal.

4.1 Análise e interpretações dos dados a partir dos grafos implicativos

Na ASI, os dados podem ser apresentados por meio de grafos. Na construção do grafo implicativo consideramos o valor mínimo do índice de implicação como sendo 0,7. As setas foram agrupadas em cores, de acordo com o valor da intensidade de implicação. O software possibilita organizar em até 4 cores por padrão. Dessa forma, temos as cores vermelho (para o maior valor), azul, verde e cinza (para o menor valor). Mas se pode escolher milhares de cores para substituir essas cores padrão. Adotamos como menor valor a intensidade de implicação de 0,70. Para os maiores valores usamos a cor padrão vermelho. Adotamos como o intervalo iniciado em 0,88 para vermelho (considerando que acima desse valor tínhamos duas relações de implicação). Entre 0,70 e 0,88 dividimos em intervalo de 0,6. Em função desses critérios, as cores dos grafos foram organizadas da forma a seguir:

Vermelha ≥0,88	0,88> azul ≥0,82	0,82> verde ≥0,76	0,76> cinza ≥0,70

Na figura 1, apresentamos o grafo implicativo gerado no CHIC em sua versão 7.0 com base nos critérios já indicados.

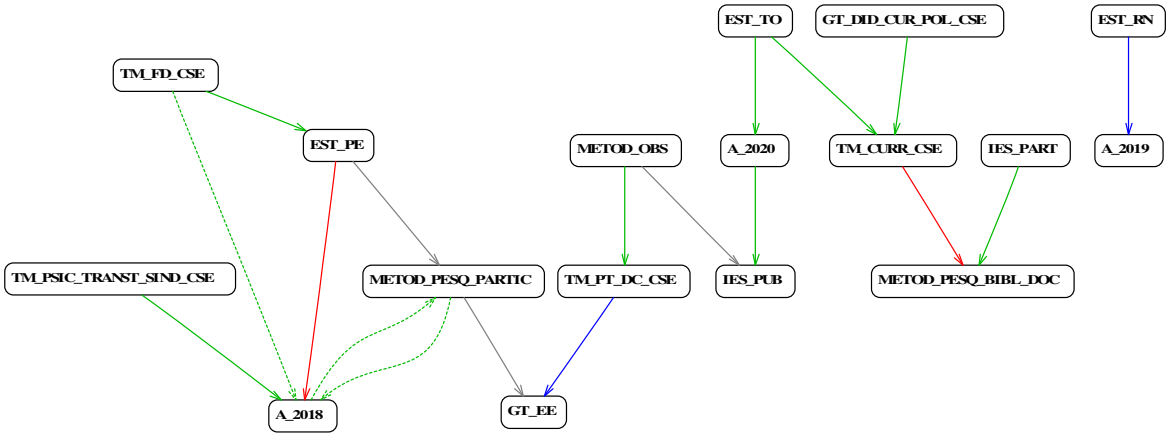
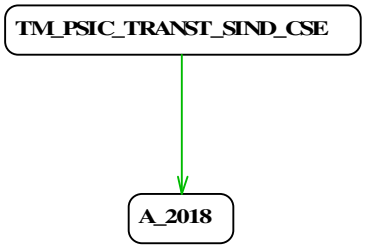


Figura 1– Grafo implicativo completo

Fonte: Dados da pesquisa

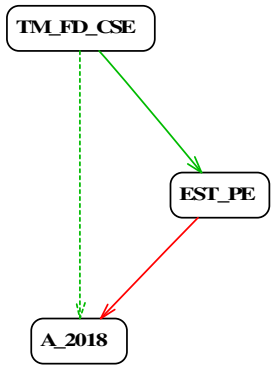
Com base no grafo acima, apresentamos a seguir as relações de implicação e as interpretações realizadas.



(TM_PSIC_TRANST_SIND_CSE ⇒ A_2018): O tema de abrangência psicologia/transtornos/síndromes teve uma tendência de ser apresentado no ano de 2018 com uma intensidade de implicação de 0,81.

Figura 2– Grafo implicativo

Podemos inferir através da ASI, num contexto geral, dentro do universo das competências socioemocionais, destacando no ano de 2018 tendências de implicação para os temas: psicologia e síndromes no contexto escolar e isso nos remete, segundo as produções pesquisadas, que as CSE quando inseridas na vivência escolar reduz sintomas de depressão, ansiedade, bullying e violência, promovendo um ambiente mais harmônico e tolerante às diferenças, levando a uma cultura de paz.



TM_FD_CSE ⇒ EST_PE⇒ A_2018: O tema de abrangência formação docente tende a ser no estado de Pernambuco com uma intensidade de implicação de 0,81. Por sua vez, as pesquisas no estado de Pernambuco tiveram uma tendência mais forte de se apresentarem no ano de 2018 no CONEDU com uma intensidade de implicação de 0,90. Ressaltamos que em 2018 o CONEDU foi em Pernambuco, o que facilitou a presença de pesquisadores pernambucanos nesse evento.

Figura 3– Grafo implicativo

TM_FD_CSE ⇒ A_2018: Observamos também uma quase transitividade (com uma intensidade de 0,78) indicando que o tema de abrangência formação docente foi mais abordado no evento do CONEDU que ocorreu em 2018.

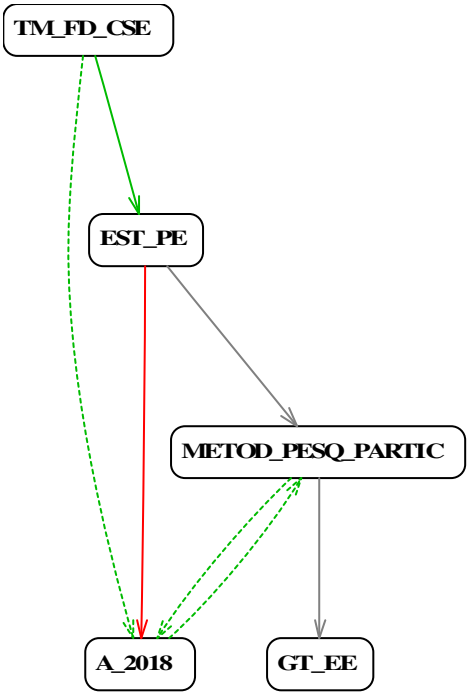


Figura 4– Grafo implicativo

TM_FD_CSE ⇒ EST_PE⇒ METOD_PESQ_PARTIC: ⇒GT_EE: O tema de abrangência formação docente implica no Estado de Pernambuco com um índice de implicação entre 0,81. O Estado de Pernambuco implica em método de pesquisa participante com uma intensidade de implicação de 0,74. A variável método de pesquisa participante possui uma quase implicação com o grupo de trabalho de educação emocional com uma intensidade de implicação de 0,74.

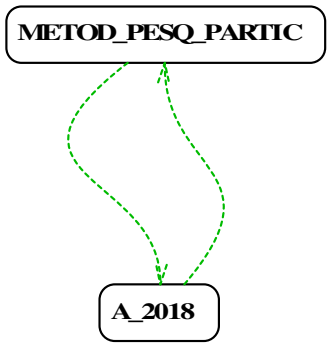


Figura 5– Grafo implicativo

METOD_PESQ_PARTIC ⇔ A_2018

Observamos uma relação de quase equivalência com uma intensidade de implicação de 0,79. Essa relação nos mostra que temos uma tendência do método de pesquisa participante a ser utilizado nas pesquisas que foram apresentadas no ano de 2018, ao mesmo tempo que no ano de 2018 temos uma tendência dos trabalhos usarem como método de pesquisa a pesquisa participante.

Em parte, essa influência pode estar associada às produções de pesquisadores pernambucanos que, pelo fato do CONEDU ter sido realizado em 2018 em Pernambuco, tiveram uma maior participação nesse evento.

O tema formação docente com pesquisas inseridas no GT de Educação Emocional demonstrou tendência de implicação para o Estado de Pernambuco em que o tipo de metodologia mais utilizado foi a pesquisa participante com proposta de intervenção e atuação dos pesquisadores nas instituições de ensino.

O tema formação docente nos traz como relevância o formato de formação docente com os vieses das Competências Socioemocionais que possibilitariam entender e atender

às demandas atuais para a educação do século XXI, que consistem em processos formativos que possam viabilizar:

- Um contexto teórico-metodológico pautado na transformação social humanizada em detrimento da racionalidade a partir da tomada de consciência de sentimentos e emoções por parte do profissional, no seu processo de aprendizagem emocional;
- Um olhar diferenciado, focado na educação das emoções como forma de autonomia pessoal e de atuação profissional na práxis educativa; Caminho para confrontar os preconceitos e estereótipos que alguns profissionais guardam dentro de si e que muitas vezes são responsáveis por impedi-lo de enxergar a realidade e contribuir para seu empoderamento e transformação pessoal;
- Estratégias de auto-organização de uma rede de profissionais hábeis emocionalmente para os desafios da vida escolar; Integralização dos pilares: indivíduo, profissional e cidadão;
- A autoconsciência considerada o alicerce sobre qual são construídas todas as outras competências, ou seja, a consciência de si mesmo, do outro e do ambiente. Numa perspectiva de mão dupla, como o profissional afeta e é afetado, levando-o ao compromisso com o próprio amadurecimento e com propósitos humanísticos para orientar suas ações;
- Técnicas que consistem em uma comunicação não violenta, escuta ativa e estimulação para soluções criativas, incluindo o respeito, a individualidade do profissional e também, da sua rede de auto, eco- organização na prática docente;
- A autoformação do profissional como início para conhecer as camadas mais profundas do seu ser e levá-lo a perceber que todos estão interligados por relações que dependem da maneira como sentimos, pensamos e agimos.

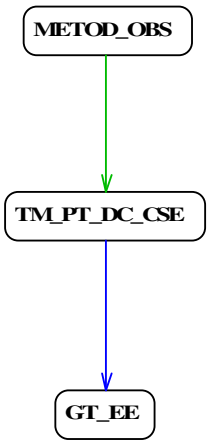


Figura 6– Grafo implicativo

METOD_OBS ⇒ TM_PT_DC_CSE⇒ GT_EE: O método de pesquisa observação implica no tema de abrangência prática docente com uma intensidade de implicação de 0,79. A variável tema de abrangência prática docente implica no grupo de trabalho educação emocional com uma intensidade de implicação de 0,82.

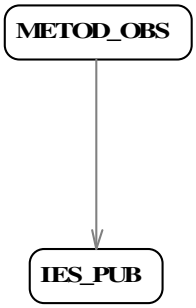


Figura 7– Grafo implicativo

METOD_OBS ⇒ IES_PUB: o método de pesquisa observação tem uma tendência, com uma intensidade de implicação de 0,75, de ser realizado em uma instituição de ensino superior pública.

O grafo implicativo na figura 7 aponta que o emprego do método de pesquisa observação tem uma tendência a ser utilizado (dentro da amostra selecionada) por pesquisadores vinculados a instituições de ensino superior públicas. Isso nos remete que hoje a prática docente vai além da mera cognição, numa perspectiva integradora de ensino em que a BNCC insere as CSE no âmbito curricular para atender a demanda de educação para o século XXI. Consideramos relevante a preocupação dos pesquisadores das instituições públicas de ensino superior, avançando dessa forma com a formação de professores. Esta também é uma preocupação do Grupo de Trabalho Educação Emocional.

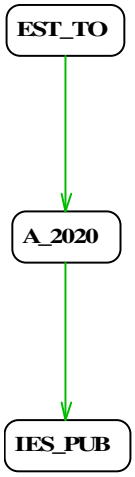
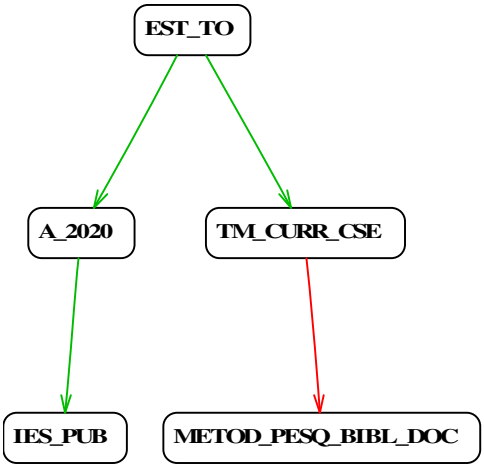


Figura 8 – Grafo implicativo

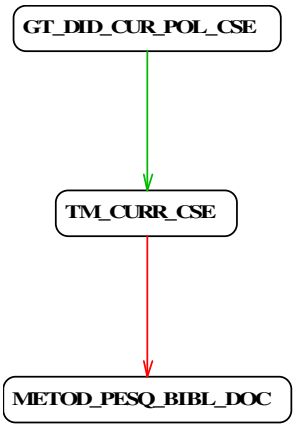
EST_TO ⇒ A_2020 ⇒ IES_PUB: A variável estado de Tocantins implica no ano de 2020 com uma intensidade de implicação de 0,78. O ano de 2020 implica em uma instituição de ensino superior com uma intensidade de implicação de 0,80.



EST_TO ⇒ TM_CURR_CSE ⇒ METOD_PES_BIBL_DOC: A variável estado de Tocantins implica na temática abrangente currículo com uma intensidade de implicação de 0,78. A temática abrangente currículo implica na metodologia pesquisa bibliográfica/documental com uma intensidade de implicação de 0,97.

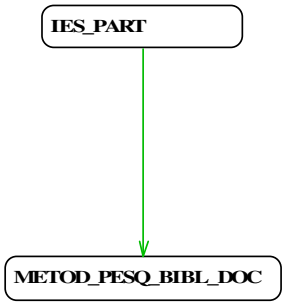
Figura 9– Grafo implicativo

O Estado de Tocantins sinalizando implicações no ano de 2020, com pesquisas vinculadas a IES públicas com temática curricular de pesquisas documentais e bibliográficas expressaram relevância ao tema CSE, carência de pesquisas na área no respectivo estado, comprovado através de pesquisas em estado da arte e ressaltaram a preocupação da inserção das CSE no âmbito curricular do estado.



GT_DID_CUR_POL_CSE⇒ TM_CURR_CSE ⇒ METOD_PES_BIBL_DOC: O grupo de trabalho didática, política e educação implica no tema abrangente currículo com uma intensidade de implicação de 0,78. A Temática abrangente Currículo implica na Metodologia Pesquisa Bibliográfica/Documental com uma intensidade de implicação de 0,97.

Figura 10– Grafo implicativo



IES_PAR⇒ METOD_PES_BIBL_DOC: A variável instituição de ensino superior particular implica na metodologia pesquisa bibliográfica/documental com uma intensidade de implicação de 0,80.

Figura 11– Grafo implicativo

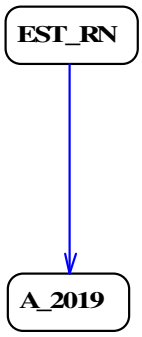


Figura 12– Grafo implicativo

EST_RN ⇒ A_2019: A variável "estado do Rio Grande do Norte" implica no ano de 2019 com uma intensidade de implicação de 0,82. O CONEDU de 2019 foi no estado do Ceará, vizinho ao Rio Grande do Norte.

A proximidade dos estados pode ter afetado a produção do Rio Grande do Norte neste ano. Foram observadas 3 produções do Rio Grande do Norte e todas as três foram nesse ano. Esse efeito, foi observado com uma intensidade de implicação mais forte no CONEDU 2018 em Pernambuco, no qual tivemos uma participação maior de pesquisadores Pernambucano. Poderia se questionar por que a produção mais intensa não foi dos pesquisadores do Ceara.

Ressaltamos, que essa intensidade de implicação é em relação aos artigos selecionados com os descritores "Competências Socioemocionais" e "Habilidades Socioemocionais". Assim, não estamos avaliando o conjunto das produções no evento que é muito superior as produções selecionadas, pois o nosso objetivo é olhar para os descritores indicados.

Desta forma, ao observarmos os dados, ao longo do CONEDU o Ceará teve uma participação maior nesses descritores com 7 produções contra 3 do Rio Grande do Norte. O ano com maiores produções dos pesquisadores do Ceara nesses descritores foi o ano em que o evento foi sediado no Ceara (com 3 publicações nessa área), contudo nos demais anos tivemos também produções como em 2018 duas produções e 2020 com duas produções. O que levou a em 2019 ter uma intensidade de implicação maior para o Rio Grande do Norte, uma vez que as únicas produções do estado nesses descritores foram nesse ano.

Apesar do número pequeno de artigos nesses descritores, observamos um crescimento no conjunto das publicações, eles podem indicar um aumento das produções nessa área. Na tabela 1, temos as produções nesses descritores por evento.

2017	2018	2019	2020
1	7	12	7

Tabela 5. Produções nos descritores pesquisados em cada CONEDU

Em 2018 tivemos apenas uma produção com esses descritores, no ano de 2018 tivemos 7 artigos, no ano de 2019 tivemos 12. No ano de 2020 tivemos uma redução, modificando a tendência de crescimento. Contudo, essa redução também foi no número de artigos total do CONEDU, em que houve uma redução de 56% de 2019 para 2020. Considerando essa redução para o conjunto, a redução em relação as publicações analisadas que passaram de 12 para 7 foi menor (42% contra 56% do total das produções). Acreditamos que a pandemia impactou na redução das produções neste evento no ano de 2020. Que como muitos outros eventos foram repensados para o modelo on-line. Ressaltamos que a educação socioemocional vem a ganhar destaque em função das inserções curriculares proveniente da BNCC (MEC, 2019) , pois na Base Nacional Curricular Comum - BNCC competência é definida como: A mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas, cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho (BRASIL, 2017, p.8).

Solidariedade, amizade, responsabilidade, colaboração, empatia, organização, ética, cidadania, honestidade. Esses valores (ou características) – tão desejáveis nos relacionamentos humanos e cada vez mais requisitados e necessários nos dias de hoje – deverão ser ensinados, praticados ou pelo menos estimulados também nas escolas. É o que dizem as novas diretrizes da Base Nacional Comum (BNCC).

Outro aspecto relevante é que a partir de 2020, todas as escolas brasileiras terão de incluir as habilidades socioemocionais nos seus currículos. Ou seja, haverá a necessidade de adaptar os programas escolares e treinar os professores para que possam ministrar essas novas competências _ que têm foco em habilidades não cognitivas, muito mais relacionadas ao comportamento e à administração das próprias emoções, mas que impactam positivamente o indivíduo e a relação dele com o mundo ao seu redor.

Destacando que a orientação é que os currículos considerem a articulação dessas competências como norteadoras do trabalho com os direitos de aprendizagem e campos de experiência (Educação Infantil) e com as áreas do conhecimento (Ensino Fundamental e Ensino Médio). Organizar um currículo pautado pelo desenvolvimento integral que as competências gerais inspiram (que estudante queremos formar?) é o primeiro passo para que estados e municípios construam suas políticas de educação integral.

Inclusive, pode-se arriscar a ligação de que Pernambuco e em especial o Ceará, que são Estados de alto índice de IDEB, avaliações externas que se mostraram de excelente desempenho nos monitoramentos educacionais.

Ainda foi mostrado nas implicações acerca do Grupo de Trabalho (GT) Didática, Currículo e Política educacional com tendências a temas de inserção curriculares das CSE com metodologias de pesquisas voltadas para pesquisas bibliográficas e documentais.

A BNCC que define o currículo obrigatório para o Brasil, aprovada em 2017, apresenta dentro desse currículo a inserção das CSE, o que torna essas habilidades como obrigatórias para a educação básica do Brasil.

Por ser uma temática apesar de antiga, mas que veio a ser discutida e cobrada formalmente em documentos curriculares a partir de 2020 os envolvimentos dos pesquisadores na área em questão estão despertando para tal compromisso e o primeiro GT de trabalho acerca do tema é considerado relativamente recente criado em 2017.

A metodologia de pesquisa bibliográfica e documentais foram sinalizadas por pesquisas provenientes de IES particulares e por fim, também no ano 2019, o estado do Rio Grande do Norte demonstra participação relacionada à temática.

Além do grafo implicativo, apresentamos também a árvore coesitiva, na qual podemos observar a formação de classes que definem metarregras.

4.2 Análise e interpretações dos dados a partir da árvore coesitiva

Na figura 13, apresentamos a árvore coesitiva, consideramos como valor mínimo 0,70 do índice de coesão (indicamos uma linha delimitando esse corte). Com o corte realizado, temos 4 classes principais sendo uma delas na forma de metarregras. Apresentamos a seguir a análise de cada uma dessas classes.

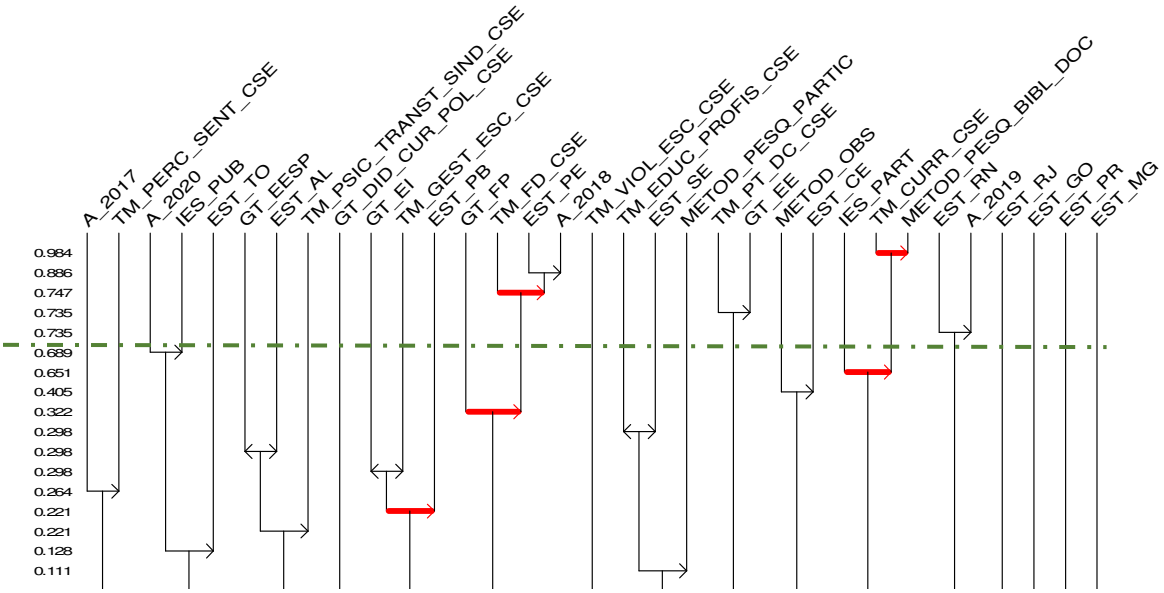


Figura 13– árvore coesitiva

Classificação ao nível: 1 : (TM_CURR_CSE ⟹METOD_PESQ_BIBL_DOC)

Tal como observado no grafo implicativo, temos que as pesquisas sobre a Temática abrangente Currículo tende a ser usada na metodologia Pesquisa Bibliográfica/Documental com um índice de coesão de 0,984. Destaca-se a temática ainda muito no plano teórico demonstrado através dos dados acima a carência de pesquisas experimentais, pesquisa-ação e mais atuação prática da temática e seus desdobramentos na vivência ensino-aprendizagem nas instituições.

Classificação ao nível: 2 : (EST_PE A_2018) Coesão : 0.886

Classificação ao nível: 3 : (TM_FD_CSE (EST_PE A_2018)) Coesão : 0.747

Com um índice de coesão de 0.886, temos que os eventos no estado de Pernambuco tendem a serem realizados no ano de 2018. Por uma vez temos uma metarregra, com um índice de coesão de 0.747, que indica as pesquisas que abordam a formação docente tem uma propensão a serem realizadas no ano de 2018 por pesquisadores pernambucanos. Consideramos o professor como um elo fundamental na educação socioemocional e esse resultado vem a mostrar que as pesquisas em Pernambuco caminham na direção de observar a formação docente como um caminho necessário na educação socioemocional.

Classificação ao nível: 4 : (TM_PT_DC_CSE ⟹ GT_EE) Coesão : 0.735

Com um índice de coesão de 0,735, temos que as pesquisas que abordaram prática docente tendem a serem realizadas no grupo de trabalho educação emocional.

Classificação ao nível: 5 : (EST_RN A_2019) Coesão : 0.735

Com um índice de coesão de 0,735, observa-se que as pesquisas no estado do Rio Grande do Norte têm uma inclinação a serem apresentadas no CONEDU realizado no ano

de 2019. Destacamos que no ano de 2019 o CONEDU foi realizado no estado do Ceará, vizinho ao Rio Grande do Norte. Isso indica importância das mudanças nas localidades dos eventos como forma de incentivar e viabilizar a vinda de pesquisadores nas proximidades do evento.

5 Considerações finais

Através do quadro teórico da Análise Estatística Implicativa, podemos observar as relações entre as produções selecionadas nas edições do CONEDU que abordaram os descritores "Competências Socioemocionais" e "Habilidades Socioemocionais" e o tema das produções, o grupo de trabalho, a natureza da pesquisa, o tipo de instituição no qual são realizadas determinados tipos de pesquisas e os estados que tiveram uma maior inserção em determinados anos dos eventos. As interpretações dessas relações de quase implicação que surgem através do tratamento estatístico, possibilitaram identificar tendências que auxiliam a compreender como se desenvolvem as pesquisas com esses descritores.

Através do tratamento com o grafo implicativo também foi possível focar a atenção sobre as pesquisas que estão sendo desenvolvidas no grupo específico sobre educação emocional. Observa-se que temos uma tendência de pesquisa com formação inicial de professores e o uso da metodologia de pesquisa participante. Consideramos relevantes esses dados e bastante coerente as investigações nessa área uma vez que se faz necessário atuar desde a formação inicial sobre a educação emocional dos futuros docentes e através da pesquisa podemos aperfeiçoar essa formação. Uma das formas de investigação que possibilita pesquisadores e pesquisados interagirem e refletirem sobre a própria realidade seria a pesquisa participante, nesse sentido é bastante coerente no quadro da educação emocional.

Muito se tem falado sobre as competências socioemocionais na prática docente, porém é fato que para isso o profissional da linha de frente, o professor (mediador), precisa estar preparado para desenvolver sua práxis educacional intencional pautada também nesses saberes, que encontram-se atrelados à BNCC e que caminham otimizando fortemente junto ao aparato cognitivo dos discentes. Mas, como anda esse preparo, essa formação docente? Como essa atuação vem acontecendo nas instituições de ensino?

Logo, somos levados a compreender a importância da Educação Emocional e das Competências Socioemocionais na esfera educacional e fortalecer a temática na formação inicial e continuada dos docentes, além do fundamental papel das IES e fomento de inserção curricular desses contextos integradores tão relevantes.

Referências

[1] Abed, A. L. Z. (2016), O desenvolvimento das habilidades socioemocionais como caminho para a aprendizagem e o sucesso escolar de alunos da educação básica. *Construção psicopedagógica*, 24 (25), 8-27.

[2] Alzina, R. (2003), Educación emocional y competencias básicas para la vida. *Revista de Investigación Educativa*, 2003, 21(1), 7-43. Disponível em: http://revistas.um.es/rie/article/view/99071/94661. Acessado em: 10 jun. 2020.

[3] Alzina, R. (2005), A educación emocional en la formación del profesorado *Revista Interuniversitaria de Formación del Profesorado*, 19(3), 95-114. Disponível em: http://www.redalyc.org/pdf/274/27411927006.pdf. Acessado em: 01 jun.2020.

[4] Alzina, R, Paniello, S. (2017), Psicología positiva, educación emocional y el programa aulas felices. *Papeles del Psicólogo* / Psychologist Papers, 38(1), 58-65. Disponível em: http://www.papelesdelpsicologo.es/pdf/2822.pdf. Acessado em: 10 jun. 2021.

[5] Bisquerra, R . A. (2000), *Educación Emocional y Bienstar*. Barcelona: Wolter Kluwer Educación.

[6] Brasil (2019) Base Nacional Comum Curricular: Educação Infantil e Ensino Fundamental. Brasília: MEC/Secretaria de Educação Básica.

[7] Casassus, J. (2009) *Fundamentos da educação emocional*. Brasília: UNESCO, Liber Livro Editora.

[8] Dantas, H. (1992), Afetividade e a construção do sujeito na psicogenética de Wallon. In: La Taille, Y.; Dantas, H. et al. *Piaget, Vygotsky e Wallon: teorias psicogenéticas em discussão.* São Paulo: Summus, 85-98.

[9] Ferreira, A. L., Acioly-Régnier, N. M. (2010), Contribuições de Henri Wallon à relação cognição e afetividade na educação. *Educar em Revista* (Impresso), 21-38.

[10] Freire, P. (1992), *Comunicação e extensão*. 10. ed. Rio de Janeiro: Paz e Terra.

[11] Goleman, D. (2011), *Inteligência emocional*. Tradução Marcos Santarrita. Rio de Janeiro: Objetiva.

[12] Gonsalves, E. P. (2015), *Educação e emoções*. Campinas: Alínea.

[13] Libâneo, J. C. (1994), *Didática*, São Paulo: Cortez Editora.

[14] Libâneo, J. C. (2001), Pedagogia e pedagogos: inquietações e buscas. *Educar*, n. 17, 153-176.

[15] Mahoney, A.; Almeida, L. R. de (org.). (2009), *Henri Wallon Psicologia e Educação*. São Paulo: Edições Loyola.

[16] Minayo, M. C. S. (2011), O desafio da Pesquisa Social. In: M.C.S. Minayo (dir.) *Pesquisa social: teoria, método e criatividade*. 32. ed. Petrópolis: Vozes, 7-79.

[17] Morin, E.; Roger, E. C.; Motta, R. (2003), *Educar na era Planetária*, São Paulo: Cortez; Brasília, DF: UNESCO.

[18] Piaget, J. (1993) *Gênese das estruturas lógicas elementares*. Rio de Janeiro: Zahar.

[19] Pimenta, S. G. (2006) *Pedagogia e Pedagogos: caminhos e perspectivas*. São Paulo: Cortez.

[20] Régnier, J.-C. et Andrade, V. L. V. X. de (2020a). A Análise Estatística Implicativa e Análise de Similaridade. Dans J.-C Régnier et V. L. V. X. de (dir.),

Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC. (p. 41-84). Recife: EDUFRPE, 2020. Récupéré le 18 février 2021 du site de l'Université Fédérale Rural de Pernambuco (Brésil) : http://www.editora.ufrpe.br/ASI.

[21] Régnier, J.-C. et Andrade, V. L. V. X. de (2020b). Usando o software CHIC. Dans J.-C Régnier et V. L. V. X. de (dir.), *Análise estatística implicativa e análise de similaridade no quadro teórico e metodológico das pesquisas em ensino de ciências e matemática com a utilização do software CHIC*. (p. 85-164). Recife: EDUFRPE, 2020. Récupéré le 18 février 2021 du site de l'Université Fédérale Rural de Pernambuco (Brésil) : http://www.editora.ufrpe.br/ASI.

[22] Valente, J. A., Almeida, M. E. B. (2015). Uso do Chic na Formação de Educadores: à guisa de apresentação dos fundamentos e das pesquisas e foco, Rio de Janeiro: Letra Capital.

[23] Vallés, A., Vallés, C. (2003). La autorregulación para el afrontamiento emocional. En Vallés, A. y Vallés, C. *Psicopedagogía de la Inteligencia Emocional*. Valencia: Promolibro.

[24] Vygotsky, L. S. (2003) *Psicologia pedagógica*. Porto Alegre: Artmed.

CONCEPÇÕES SOBRE O ENSINO DE ÁLGEBRA: UMA INVESTIGAÇÃO A PARTIR DA ANÁLISE ESTATÍSTICA IMPLICATIVA

Flávia Pereira Righi[1], Leonardo Dalla Porta[2], Greice Scremin[3]

CONCEPTIONS SUR L'ENSEIGNEMENT DE L'ALGÈBRE: UNE ENQUÊTE À PARTIR DE L'ANALYSE STATISTIQUE IMPLICATIVE

CONCEPTIONS ABOUT THE TEACHING OF ALGEBRA: AN INVESTIGATION FROM THE IMPLICATIVE STATISTICAL ANALYSIS

RESUMO

Este trabalho teve por objetivo investigar acerca das concepções sobre o ensino de álgebra que professores que ensinam matemática no Ensino Fundamental manifestam. Este interesse é motivado pela recente publicação da Base Nacional Comum Curricular – BNCC no Brasil, em que o ensino de álgebra passa a ter como finalidade o desenvolvimento do pensamento algébrico. Nessa perspectiva, este trabalho orientou-se por uma pesquisa quanti-qualitativa, em que os dados foram obtidos por meio de questionário *online* e interpretados com apoio da Análise Estatística Implicativa – ASI e fundamentados na revisão bibliográfica. A partir do grafo implicativo produzido pelo *software* CHIC 7.0 , foi possível observar uma possível tendência de que os professores que ensinam matemática no Ensino Fundamental concebem a álgebra como uma forma de pensamento, entretanto, não viabilizam meios didáticos de promover o seu desenvolvimento, e com isso, retomam o ensino técnico da álgebra com o objetivo de formar a base para o Ensino Médio.

Palavras-chave: *Pensamento algébrico, Base Nacional Comum Curricular, Didática da matemática, Ensino Fundamental.*

RÉSUMÉ

Ce travail visait à étudier les conceptions des professeurs de mathématiques sur l'enseignement de l'algèbre dans l'enseignement primaire. La motivation tire son origine dans la récente publication de la Base Nationale Commune Curriculaire – BNCC au Brésil, où l'enseignement de l'algèbre vise à développer la pensée algébrique. Le travail a été réalisé par une recherche quantitative et qualitative où les données ont été collectées par un questionnaire on-line et les informations traitées à travers de l'Analyse Statistique Implicative – ASI. À partir du graphe implicatif produite par le logiciel CHIC 7.0, il a été possible d'observer une tendance que les professeurs de mathématiques comprennent l'algèbre comme une forme de pensée, mais n'utilisent pas de ressources didactiques pour promouvoir son développement, l'enseignement que de manière technique.

Mots-clés : *Pensée algébrique, Base Nationale Commune Curriculaire, Didactique mathématiques, École élémentaire.*

ABSTRACT

This paper aimed to investigate about the conceptions about teaching algebra that teachers who teach mathematics in elementary school manifest. This interest is motivated by the recent publication of the National Common Curricular Base - BNCC in Brazil, in which the teaching of algebra starts to aim at the development of algebraic thinking. In this

[1]Universidade Franciscana, Brasília/DF, Brasil, flaviarighi87@gmail.com
[2]Universidade Franciscana, Santa Maria/RS, Brasil, leodallaporta@gmail.com
[3]Universidade Franciscana, Santa Maria/RS, Brasil, greicescremin@gmail.com

perspective, this work was guided by a quantitative-qualitative research, in which the data were obtained through an online questionnaire and interpreted with the support of the Implicative Statistical Analysis – ASI and based on the bibliographic review. From the implicit graph produced by the CHIC 7.0 software, it was possible to observe a trend that teachers who teach mathematics in elementary school conceive algebra as a way of thinking, however, they do not provide didactic means to promote its development, and with that, they resume teaching algebra technician in order to form the basis for high school.

Keywords: *Algebraic thinking, National Common Curricular Base, Didactics of mathematics, Elementary school.*

1 Introdução

Os documentos que regem a Educação Básica no Brasil evidenciam a importância da álgebra nos currículos escolares, uma vez que atribuem a esse campo da Matemática uma unidade temática exclusiva para estudo de seus objetos de conhecimento. Atualmente, o ensino da Matemática no contexto escolar brasileiro assume como sua finalidade fundamental dotar os estudantes de competências matemáticas adequadas que lhes permitam enfrentar as demandas de seus ambientes social e cultural (Brasil, 2017).

No tocante a álgebra, a principal competência a ser desenvolvida nos estudantes é o pensamento algébrico, entendido como "essencial para utilizar modelos matemáticos na compreensão, representação e análise de relações quantitativas de grandezas e, também, de situações e estruturas matemáticas, fazendo uso de letras e outros símbolos" (Brasil, 2017, p. 270). Nesse contexto, as ideias fundamentais vinculadas a esta unidade são: equivalência, variação, interdependência e proporcionalidade.

Entretanto, essa concepção de educação algébrica voltada para o desenvolvimento do pensamento algébrico é relativamente recente, dada pela publicação da Base Nacional Comum Curricular – BNCC, em dezembro de 2017, e pela pouca oferta de cursos de formação continuada de professores, em que a mesma fosse difundida e viabilizada por diferentes metodologias. O que se tem é uma nova proposta para o ensino de álgebra sem a iniciativa de discutir com os professores que ensinam matemática novos processos de ensino e aprendizagem que o viabilize.

Definir a álgebra bem como compreender e caracterizar o pensamento algébrico não é uma tarefa simples. Admite-se, neste trabalho, a definição de Kieran (2007, p. 5, tradução nossa[4]) em que

> [...] a Álgebra não é apenas um conjunto de procedimentos envolvendo os símbolos em forma de letras, mas consiste também na atividade de generalização e proporciona uma variedade de ferramentas para representar a generalidade das relações matemáticas, padrões e regras. Assim, a Álgebra passou a ser vista não apenas como uma técnica, mas também como uma

[4] Algebra was not merely a set of procedures involving the letter-symbolic form, but also that it consisted of generalizing activity and provided a range of tools for representing the generality of mathematical relationships, patterns, and rules (e.g., Mason, 2005). Thus, Algebra came to be seen not merely as technique, but also as a way of thinking and reasoning about mathematical situations.

> forma de pensamento e raciocínio acerca de situações matemáticas.

Apoiados nessa definição, entende-se o pensamento algébrico como sendo o processo de generalização de ideias matemáticas a partir de um conjunto de casos particulares (Blanton & Kaput, 2005). Nesse processo, é possível identificar algumas características do pensamento algébrico manifestadas pelo indivíduo, conforme apontam Almeida e Câmara (2017): estabelecer relações; generalizar; modelar; construir significado; e operar com o desconhecido, sendo a capacidade de estabelecer relações o centro propulsor desse pensamento.

Esta concepção de educação algébrica não é estática e, também não está inerte ao longo do tempo. Em uma breve releitura acerca da história e desenvolvimento da álgebra, é possível perceber as diferentes concepções que se teve sobre a mesma e sua influência nos processos de ensino e aprendizagem. Algumas vezes, um novo redirecionamento para o ensino de álgebra não implicou, necessariamente, na formulação automática de uma concepção por parte dos professores e, tampouco, de métodos didáticos que atendam a esse novo objetivo.

Faz-se necessário, portanto, programas de formação continuada de modo a refletir, discutir e atualizar sobre novas concepções educacionais com o objetivo de minimizar o lapso de tempo de assimilação e adequação pedagógica em prol de uma efetiva mudança de contexto. Foi considerando esses pressupostos, que este trabalho se deu com o objetivo de identificar as concepções de professores de matemática a respeito do ensino de álgebra, uma vez que sua finalidade passou a ser o desenvolvimento do pensamento algébrico.

Para tal, elaborou-se uma pesquisa quanti-qualitativa com professores de matemática do Ensino Fundamental, da rede pública e privada de Santa Maria – RS e região. Os participantes da pesquisa responderam, de modo *online*, um questionário de perguntas fechadas que versou sobre suas concepções sobre a álgebra e seu ensino. A apreciação e inferência dos dados se deu por meio da Análise Estatística Implicativa - ASI, pautado na revisão bibliográfica acerca do tema.

Este artigo está estruturado em cinco seções, da seguinte maneira: na seção 2, tem-se a fundamentação teórica do trabalho, em que são discutidos alguns elementos históricos da álgebra e da linguagem algébrica, as concepções de álgebra de Usiskin (2003), e o ensino de álgebra no contexto brasileiro. Na seção 3, trata-se dos aspectos metodológicos desenvolvidos nesta pesquisa. Na seção 4, são apresentados e discutidos os resultados obtidos pela ASI. E por fim, na seção 5, apresentam-se as considerações finais.

2 Revisão bibliográfica

2.1 Alguns elementos históricos sobre o desenvolvimento da álgebra e da linguagem algébrica

O desenvolvimento histórico da álgebra pode ser dividido em duas grandes etapas temporais: a da Álgebra Clássica, desenvolvida até o final do século XVIII, que compreendia o estudo das equações e o método de resolvê-las; e a da Álgebra Abstrata, produzida a partir do século XIX, que trata do estudo das estruturas matemáticas, como grupos, anéis e corpos (Gomes, 2013; Santos & Borges, 2013).

O desenvolvimento da linguagem algébrica, assim como o da própria álgebra, também sofreu modificações ao longo dos séculos e pode ser definida em três estágios distintos: Retórica; Sincopada; e Simbólica. Segundo Gomes (2013), a distinção desses três estágios na evolução da linguagem algébrica foi estabelecida pelo alemão Georg Heinrich Friedrich Nesselmann (1811-1881), em 1842. Segundo Nesselmann apud Gomes (2013), a distinção das três fases se dá pela consideração de como a representação formal das equações e operações é realizada.

É possível identificar a linguagem retórica na álgebra dos babilônicos (2000 a 1600 a. C.), dos egípcios (2.000 a. C.) e dos gregos (300 a. C.). De acordo com Gomes (2013, p. 25) "a álgebra na qual os cálculos são expressos completa e detalhadamente por meio da linguagem comum, sem qualquer uso de símbolos e abreviações, é descrita como retórica".

Apesar da similaridade, a álgebra dos gregos se diferenciava pelo seu caráter geométrico, devido a ideia de número estar vinculada a um segmento de reta e as operações algébricas estarem descritas em terminologia geométrica (Vailati & Pacheco, 2011). Assim, a álgebra grega encontrada nos primeiros trechos da obra *Os Elementos* de Euclides de Alexandria, pode-se verificar que Euclides enuncia a identidade $(a + b)^2 = a^2 + 2ab + b^2$ na forma: "Dividindo-se uma reta em duas partes, o quadrado sobre a reta toda é igual à soma dos quadrados sobre as partes juntamente com o dobro do retângulo contido pelas partes" (Eves, 2010, p. 108). Percebe-se, portanto, que a álgebra de Euclides é construída pelo prisma geométrico e caracterizada pela linguagem retórica.

Foi durante o século III d. C., que Diofanto de Alexandria dedicou seus estudos ao uso de símbolos para facilitar a escrita e os cálculos matemáticos. Segundo Gomes (2013, p. 26) "a segunda fase proposta por Nesselmann para a linguagem algébrica é a fase sincopada, em que a exposição ainda é de natureza retórica, mas faz uso de abreviações consistentes para certos conceitos e operações que ocorrem frequentemente". Para Santos (2016), Diofanto foi o primeiro a utilizar os símbolos para representação da incógnita por meio da letra "sigma" do alfabeto grego e com isso introduziu a álgebra grega à linguagem sincopada.

De acordo com Gomes (2013, p. 27), a linguagem algébrica dos europeus até a metade do século XVII estava inserida na fase sincopada, como no caso do "algebrista italiano Gerolamo Cardano (1501-1576) que, com a expressão "cubus p. 6 rebus aequalis 20", exprimia a equação que, na linguagem simbólica posterior, corresponderia a $x^3 + 6x = 20$".

A álgebra simbólica, último estágio do desenvolvimento da linguagem algébrica, de acordo com Gomes (2013), "caracteriza-se pela expressão das ideias mediante o uso de símbolos, dispensando-se completamente o recurso às palavras da linguagem comum". Esse estágio teve início no século XVI e está vigente até hoje. Apesar da contribuição de François Viète ao simbolismo algébrico ao propor o uso de vogais para representação de incógnitas e consoantes para representação de coeficientes, a passagem para uma álgebra completamente simbólica ficou reconhecida na obra de René Descartes (1596-1650) *A Geometria*, em 1637. Segundo Gomes (2013, p. 27), "Descartes emprega as primeiras letras do alfabeto para simbolizar as quantidades fixas e as últimas para nomear as incógnitas; ele também usou as notações x, xx, x^3, x^4, ... para as potências".

Para Vailati e Pacheco (2011) a evolução da linguagem algébrica (da retórica à simbólica) possibilitou um aprofundamento no pensamento algébrico ao passar da

"solução manipulativa de equações" para o estudo de suas propriedades teóricas. Assim, influenciou nas finalidades do estudo da álgebra e nas concepções que se têm sobre a mesma.

2.2 Concepções sobre a álgebra segundo Usiskin (2003)

Diante do desenvolvimento da linguagem algébrica, Usiskin (2003) afirma que as finalidades da álgebra são determinadas por concepções diferentes da álgebra que se correspondem aos diversos usos das variáveis. Diante disso, o autor identifica quatro diferentes concepções de álgebra: a álgebra como aritmética generalizada; a álgebra como estudo de procedimentos para resolver certos tipos de problemas; a álgebra como estudo de relações entre grandezas; e a álgebra como estudo das estruturas.

Na primeira concepção, a álgebra como aritmética generalizada, as variáveis são vistas como generalizadoras de modelos, ou seja, o objetivo é traduzir e generalizar. Por exemplo, em 3 + 5 = 5 + 3 generaliza-se o conceito de que a ordem das parcelas não altera a soma através da expressão $a + b = b + a$ (Gomes, 2013).

Na concepção de álgebra como estudo de procedimentos para resolver certos tipos de problemas, as variáveis são incógnitas ou constantes, e os objetivos são simplificar e resolver. No problema - Adicionando 3 ao quíntuplo de um certo número, o resultado é 43. Ache o número - primeiramente é preciso traduzir o problema para a linguagem algébrica, conforme a primeira concepção, ou seja, o problema é escrito como $5x + 3 = 43$. A partir disso, o trabalho é resolver a equação, conforme a segunda concepção, através da manipulação algébrica e simplificação.

$$5x + 3 + (-3) = 43 + (-3). (1)$$

$$5x = 40$$

$$x = 8.$$

Segundo Gomes (2013, p. 15) "o aluno, nessa concepção, precisa dominar não apenas a capacidade de equacionar os problemas (isto é, traduzi-los para a linguagem algébrica em equações), como também precisa ter habilidades em manejar matematicamente essas equações até obter a solução". Destaca-se que nessa concepção, a letra representa um valor a ser encontrado e, por isso, é uma incógnita.

Na terceira concepção, a álgebra como estudo de relações entre grandezas, as variáveis realmente se alteram, ou seja, dentro desta concepção, a álgebra se ocupa de modelos e leis funcionais que descrevem ou representam as relações entre duas ou mais grandezas variáveis. O objetivo não é encontrar um valor determinado para a variável, mas sim, estabelecer a relação entre duas ou mais variáveis (Gomes, 2013).

Gomes (2013, p. 15) ilustra essa concepção através do exemplo: "o que ocorre com o valor de 1/x quando x se torna cada vez maior?". Nesse caso, não está sendo questionado o valor de x e, portanto, x não é uma incógnita e também não está sendo pedido que o aluno traduza o problema. Segundo a autora, "há um modelo a ser generalizado, mas não se trata de um modelo que se pareça com a aritmética (não teria sentido perguntar o que aconteceria com o valor de ½ quando 2 se torna cada vez maior). Trata-se de um modelo fundamentalmente algébrico".

Na quarta e última concepção, a álgebra como estudo das estruturas, pode-se reconhecer nas propriedades das operações com números reais e polinômios as características dessa concepção. Ou seja, o que caracteriza a "variável" na concepção da álgebra como estudo de estruturas é o fato de ser pouco mais do que um símbolo arbitrário. No problema - Determine $(a + x).(b - 1)$- fatorar a expressão não coincide com nenhuma das concepções já discutidas. A autora sintetiza as diferentes concepções de álgebra de acordo com os diferentes usos das variáveis, conforme proposto por Usiskin (2003), no Quadro 1.

Concepção de álgebra	Uso ou papel das variáveis
Aritmética generalizada	Generalizadoras de modelos (traduzir, generalizar)
Estudo de procedimentos para resolver problemas	Incógnitas, constantes (resolver, simplificar)
Estudo de relações entre grandezas	Argumentos, parâmetros (relacionar, fazer gráficos)
Estudo de estruturas	Sinais arbitrários no papel (manipular, justificar)

Fonte: Gomes (2013, p. 17).

Quadro 1. Concepções da álgebra conforme Usiskin (2003).

Para Gomes (2013), as atividades conhecidas como de cálculo algébrico na educação básica (produtos notáveis, fatoração, operações com monômios e polinômios) situam-se nesse âmbito da quarta concepção, estudo de estruturas. O estudo de funções situa-se no âmbito do estudo de relações entre grandezas, enquanto o estudo de equações e inequações encontram-se no âmbito de procedimentos para resolver problemas.

É possível perceber que o desenvolvimento da linguagem algébrica exerceu significativa influência no desenvolvimento da própria álgebra bem como da concepção algébrica. Foi pela passagem da linguagem retórica para a simbólica que objetos de estudo algébricos puderam ser investigados, como no caso das estruturas matemáticas, e também a mudança de uma concepção de álgebra para resolução de problemas, conforme proposto por Trajano (1947), por outra em que o objetivo é aprofundar-se no pensamento algébrico.

2.3 O ensino de álgebra no Brasil

Ao analisar a evolução histórica do ensino de álgebra no Brasil, identificam-se marcos importantes que determinaram o *modus operandi* de seu ensino revelando as diferentes concepções que se desenvolveram acerca da mesma. Segundo Gomes (2013), o ensino da álgebra no Brasil pode ser caracterizado a partir de três momentos históricos: antes, durante e depois do Movimento da Matemática Moderna.

Conforme Miguel et al. (1992) o ensino de álgebra entrou nos currículos brasileiros quando as cadeiras de Aritmética, Geometria e Trigonometria já estavam estabelecidas. Através da Carta Régia de 19 de agosto de 1799 que seu ensino foi introduzido na forma

de aulas avulsas. Nesse período, o ensino se dava em compartimentos estanques, ou seja, primeiro se estudava todos os conteúdos de Aritmética, depois os conteúdos de Álgebra, e, por fim, os conteúdos de Geometria, sem articulação entre esses campos (Miguel et al., 1992).

Somente com a Reforma Francisco Campos (1931) – primeira Reforma Educacional Brasileira - que o ensino de Aritmética, Álgebra, Geometria e Trigonometria recebeu o nome unificado de Matemática e seu intuito era organizar o ensino secundário, estabelecendo definitivamente um currículo seriado (Miguel et al., 1992). Apesar dessa reestruturação proposta pela Reforma, Miguel et al. (1992), em análise de livros-textos anteriores à década de 60, salientam que os tópicos de álgebra permaneceram inalterados e trabalhados de modo mecanizado e automatizado, seguindo os padrões de ensino da aritmética, apresentados na forma de regras e procedimentos.

Segundo Araujo (2008), isso demonstra que a aprendizagem da álgebra era baseada em procedimentos sem significados. Desse modo, era marcada pelo caráter reprodutivo, enraizado no objetivo de resolver problemas, conforme encontra-se em Trajano (1947, p. 7), o qual define: "Álgebra é a parte das matemáticas que resolve os problemas e demonstra os teoremas quando as quantidades são representadas por letras".

Fiorentini et al. (1993) revelam que durante todo o século XIX e metade do século XX, tanto no Brasil quanto em outros países, o ensino de álgebra estava pautado na noção de transformismo algébrico, ou seja, era necessário e suficiente que o aluno compreendesse os processos de transformações algébricas (expressões algébricas equivalentes entre si) empregando regras e propriedades válidas na busca de resolver problemas (Fiorentini et al., 1993; Gomes, 2013; Aguiar, 2014). Fiorentini et al. (1993) denominam essa primeira concepção de *linguístico-pragmática.*

Com o surgimento do Movimento da Matemática Moderna, nacional e internacionalmente, desde o final do século XIX na Europa e a partir de 1960 no Brasil, abriu-se espaços importantes de discussão com vistas ao ensino de matemática. Esse movimento revelou uma possibilidade de ensino mais contextualizado e menos complexo, acessível a todos os alunos, contrapondo-se a concepção tecnicista da álgebra evidenciada no momento anterior (Claras & Pinto, 2008; Coelho & Aguiar, 2018).

Um dos objetivos desse movimento de modernização da matemática era a unificação de seus três campos fundamentais (Aritmética, Álgebra e Geometria) através da introdução de elementos unificadores, como a teoria dos conjuntos, funções e as estruturas algébricas (Miguel et al., 1992; Araujo, 2008). Segundo Coelho e Aguiar (2018, p. 173), durante o Movimento da Matemática Moderna

> [...] os tópicos algébricos foram reorganizados de tal sorte que se ensinassem em primeiro lugar os conjuntos numéricos, suas propriedades estruturais, as sentenças abertas e fechadas, o conjunto-universo e o conjunto-verdade e as equações e inequações do 1° grau. Só após esse estudo é que seriam abordadas as expressões algébricas, valores numéricos, operações e a fatoração. Por fim, na sequência, seriam introduzidos os novos conteúdos algébricos, isto é, as funções.

Nesse sentido, a álgebra deixou de ter o caráter pragmático e passou a fundamentar os outros campos da matemática escolar (Aguiar, 2014). Para Araujo (2008), a álgebra passou a ocupar um lugar de destaque, recebendo um maior rigor e assumindo a

responsabilidade com os aspectos lógico-estruturais dos conteúdos matemáticos e a precisão da linguagem. Fiorentini et al. (1993) denominam essa segunda concepção de *fundamentalista-estrutural*.

Pode-se perceber que o caráter da concepção linguístico-pragmática está associado à necessidade de resolver problemas, enquanto, na concepção fundamentalista-estrutural a ênfase está na precisão matemática do conceito de equação e na linguagem exata para expressá-lo (Gomes, 2013).

Contudo, na segunda metade da década de 1970, o Movimento da Matemática Moderna entrou em declínio em todo o mundo e críticas aos seus pressupostos surgiram. Diante desse novo contexto, surge uma terceira concepção para o ensino de álgebra, a qual aponta para uma síntese das duas concepções anteriores. Portanto, recuperar o valor instrumental da álgebra, na perspectiva da concepção prevalecente antes do movimento da matemática moderna, contudo, mantendo o caráter de justificação das passagens presentes no transformismo algébrico, de natureza fundamentalista (Gomes, 2013; Aguiar, 2014).

Essa nova concepção busca justificar as passagens algébricas através de recursos analógicos geométricos, isto é, visuais. Segundo Aguiar (2014), os seguidores dessa ideia acreditam que uma álgebra a partir da visualização de elementos geométricos torna certas identidades algébricas didaticamente visíveis, sendo superior a qualquer forma de abordagem estritamente lógico-simbólica. Em decorrência disso, Fiorentini et al. (1993) denominam essa terceira concepção de *fundamentalista-analógica*. Os autores salientam que exemplos da concepção fundamentalista-analógica são as situações em que os produtos notáveis, o cálculo ou a fatoração de expressões algébricas são apresentadas com o apoio de ilustrações de quadrados e retângulos.

A concepção fundamentalista-analógica é a perspectiva mais atual na educação algébrica, entretanto, Fiorentini et al. (1993) ressaltam que no contexto escolar continua-se a conferir o papel principal ao transformismo algébrico. Aguiar (2014) afirma que apesar dos esforços em formação inicial e continuada,

> [...] algumas deficiências do ensino de álgebra são resultado de um processo histórico desse ensino e que, atualmente, existem outras concepções de ensino na literatura que apontam para o desenvolvimento de um pensar algébrico como marca dessa concepção (Aguiar, 2014, p. 36).

Para Gomes (2013, p. 39) "a visão da álgebra, nas concepções linguístico- pragmática, fundamentalista-estrutural e fundamentalista-analógica, é centrada na vertente simbólica, e, assim, o aspecto mais importante é a aprendizagem da manipulação dos símbolos". Em suma, o que está implícito é que o domínio de regras algébricas possibilita ao estudante aplicar corretamente a álgebra na resolução de situações reais.

Mais de trinta anos depois do Movimento da Matemática Moderna, em 1997, o Governo Federal elabora os Parâmetros Curriculares Nacionais (PCN) com o objetivo de orientar os educadores a respeito dos elementos considerados relevantes para cada um dos componentes curriculares integrantes da Educação Básica (Brasil, 1997). Nesse documento, os conteúdos matemáticos estavam organizados em quatro blocos temáticos: Números e Operações; Espaço e Forma; Grandezas e Medidas; e Tratamento da Informação.

Sendo assim, em tal documento, os conteúdos algébricos pertenciam ao bloco Números e Operações, ainda muito enraizado com o desenvolvimento da aritmética, e tinham por finalidade desenvolver e exercitar a capacidade de abstração e generalização para a resolução de problemas (Brasil, 1997). Além disso, os PCN orientavam que os objetos de estudo do campo da álgebra deveriam ser trabalhados a partir do 7º ano do Ensino Fundamental, pois considerava a capacidade de conexões lógicas e de abstração que os alunos passariam a ter nessa fase, evidenciando uma estreita relação às ideias piagetianas quanto às fases de desenvolvimento cognitivo da criança (Piaget, 1971). Pois, há uma tendência a priorizar a introdução dos estudos de álgebra apenas no período das operações formais que se caracteriza pela transformação dos esquemas cognitivos, operados concretamente em esquemas baseados na realidade imaginada.

Mais recentemente, em dezembro de 2017, teve-se mais uma atualização nas orientações curriculares da Educação Básica no Brasil, dada a publicação da BNCC. O objetivo desse documento é direcionar a educação brasileira para a formação humana integral e para a construção de uma sociedade justa, democrática e inclusiva (Brasil, 2017). Nesse documento normativo, os diferentes campos da matemática estão divididos em cinco unidades temáticas: Números, Álgebra, Geometria, Grandezas e Medidas, e Probabilidade e Estatística (Brasil, 2017).

Nesse caso, é possível observar que a álgebra compõe uma unidade exclusiva, intitulada Álgebra*,* evidenciando a sua importância em conjunto com as demais unidades matemáticas. Além disso, o objetivo do ensino de álgebra passa a ser o desenvolvimento do pensamento algébrico, por meio de ideias fundamentais como equivalência, variação, interdependência e proporcionalidade (Brasil, 2017).

Quanto a introdução da álgebra no currículo, diferentemente dos PCN, a BNCC sugere que as primeiras noções algébricas sejam trabalhadas desde os anos iniciais do Ensino Fundamental, por meio da observação de regularidades, generalizações de padrões e algumas propriedades da igualdade (Brasil, 2017). Essas orientações coincidem com as ideias de Zabala (1998), quando considera outras formas de vincular os diferentes conteúdos de aprendizagem, pois refere que, ao longo da história, os conhecimentos foram alocados em disciplinas, em uma lógica da organização curricular linear e em níveis de complexidade progressivos. Sendo assim, observa-se o longo caminho que a concepção algébrica atual percorreu nos currículos brasileiros desde sua introdução até o seu reconhecimento e valorização na BNCC.

3 Aspectos metodológicos

Tendo em vista o alcance do objetivo proposto, esta pesquisa caracteriza-se como pesquisa quanti-qualitativa quanto a abordagem do problema, a qual, entende-se assim como Minayo e Sanches (1993, p. 247) que

> A relação entre quantitativo e qualitativo, entre objetividade e subjetividade não se reduz a um *continuum*, ela não pode ser pensada como oposição contraditória. Pelo contrário, é de se desejar que as relações sociais possam ser analisadas em seus aspectos mais "ecológicos" e "concretos" e aprofundadas em seus significados mais essenciais. Assim, o estudo quantitativo

> pode gerar questões para serem aprofundadas qualitativamente, e vice-versa (Minayo; Sanches, 1993, p. 247).

Primeiramente, desenvolveu-se a revisão bibliográfica em diferentes literaturas relativas ao desenvolvimento da álgebra e suas concepções. De acordo com Gil (2006), a pesquisa bibliográfica é desenvolvida a partir de material já publicado com o intuito de analisar posições diversas em relação a determinado assunto. Para tal, constituíram-se como objetos de análise diferentes referenciais teóricos publicados, como livros, teses, dissertações, monografias, artigos científicos em anais e revistas que possibilitaram que o tema proposto fosse fundamentado.

Posteriormente, foi elaborado como instrumento de pesquisa um questionário enviado de modo *online*, via Google Forms, composto por perguntas fechadas (dados quantitativos), em que o objetivo era identificar as concepções acerca da álgebra e seu ensino manifestadas por professores de matemática do Ensino Fundamental. Os participantes da pesquisa referem-se a 12 professores de matemática que atuam na rede pública e privada de ensino do município de Santa Maria – RS e região.

No Quadro 2 são apresentadas as perguntas realizadas neste questionário e suas alternativas de respostas com as respectivas siglas, as quais são necessárias visando a entrada, leitura e interpretação dos dados no *software* CHIC 7.0. Para cada objetivo específico do questionário foram elaborados quadros de perguntas.

Objetivo específico	**Perguntas e siglas**
Identificar a formação dos participantes e a sua relação com a álgebra, na perspectiva do próprio conhecimento algébrico e na perspectiva da didática desse conhecimento.	**1) Sua formação acadêmica atual:** a. Magistério (1MAG) b. Graduação/Licenciatura (1GLI) c. Especialização (1ESP) d. Mestrado (1MES) e. Doutorado (1DOU)
	2) Na sua formação, você teve alguma disciplina específica sobre Álgebra? a. Sim, na graduação (2SG) b. Sim, na pós-graduação (2SPG) c. Não (2NAO)
	3) Na sua formação, você teve alguma disciplina voltada para o ENSINO de álgebra? a. Sim, na graduação (3SG) b. Sim, na pós-graduação (3SPG) c. Não (3NAO)
Investigar o modo como esses professores concebem a álgebra e a sua melhor definição.	**4) Marque três alternativas que estejam mais relacionadas com as suas concepções sobre a álgebra:** a. Extensão da aritmética (4ARIT) b. Resolução de problemas (4RP) c. Transformismo algébrico (4TA) d. Relação entre grandezas distintas (4RGD) e. Generalizações (4GEN) f. Modelagem matemática (4MOM) g. Estruturas lógico-matemáticas (4ELM) h. Uso de números e letras (4NL) i. Operações aritméticas com letras (4OAL)
	5) Para você, a melhor definição de Álgebra é: a. Conjunto de procedimentos envolvendo símbolos em forma de letras (5CJP) b. A atividade de generalização e proporciona uma variedade de ferramentas de representação (5ATG)

	c. É uma técnica para resolução de problemas apoiada em operações aritméticas (5TER) d. É a relação entre grandezas que permite estruturar algebricamente procedimentos para resolver certos tipos de problemas (5REG) e. É uma forma de pensamento e raciocínio acerca de situações matemáticas (5FPE)
Identificar e relacionar o conhecimento desses professores quanto às orientações da BNCC e suas atividades para o ensino de álgebra.	**6) Você já se apropriou sobre as orientações da Base Nacional Comum Curricular - BNCC (2017) para o ensino de álgebra no Ensino Fundamental?** a. Sim (6SIM) b. Não (6NAO) c. Parcialmente (6PAR)
	7) As atividades de álgebra que você propõe aos seus alunos, em sua maioria, visam: a. A manipulação algébrica (7MA) b. A resolução de problemas (7RP) c. O desenvolvimento de um modo de pensar (7DPA) d. Formar a base necessária para o Ensino Médio (7FEM)
Investigar sobre a utilização desse recurso didático e suas expectativas de ensino a partir dele, bem como questionamos sobre a origem dos problemas de álgebra que esses professores utilizam em aula.	**8) Sua escola possui livros didáticos atualizados e organizados para uso em sala de aula?** a. Sim (8SIM) b. Não (8NAO)
	9) Você utiliza o livro didático (pode marcar mais de uma alternativa): a. No planejamento das aulas (9PLA) b. Na elaboração de trabalhos, projetos e avaliações (9TRA) c. Em sala de aula juntamente com os alunos (9EMA) d. Como fonte de problemas e exercícios (9PEX) e. Como apoio teórico aos conteúdos (9APT) f. Como fonte de pesquisa e consulta por parte dos alunos (9PAL) g. Não utilizo livro didático (9NAO)
	10) Quanto aos conteúdos de álgebra, os problemas contidos em livros didáticos atendem às tuas expectativas de ensino. a. Muito satisfatório (10MS) b. Satisfatório (10S) c. Pouco satisfatório (10PS) d. Não utilizo livros didáticos (10NU)
	11) Caso você não utilize os livros didáticos, marque abaixo a origem dos problemas de álgebra que você utiliza em aula (pode marcar mais de uma alternativa). a. Utilizo os problemas com os quais eu aprendi (11AP) b. Sites diversos da internet (11IN) c. Artigos científicos, dissertações e teses (11TR) d. Eu crio meus próprios exemplos (11CR)

Fonte: elaborado pela autora.

Quadro 2. Perguntas do questionário, suas siglas e objetivos.

Após as devolutivas do questionário, as respostas dos professores foram tabuladas em planilha eletrônica de modo a possibilitar a leitura pelo *software*, o qual requer que os dados sejam apresentados em escala binária. A seguir, apresenta-se brevemente a Análise Estatística Implicativa como método para análise de dados quantitativos.

3.1 Análise Estatística Implicativa – ASI

A Análise Estatística Implicativa (ASI) é um método de análise quantitativa e de classificação de dados multidimensionais, desenvolvido no âmbito da Didática da Matemática francesa a partir de situações didáticas (Dalla Porta, 2019). Segundo Gras e Régnier (2015, p. 22) "a análise estatística implicativa (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".

De acordo com Gras et al. (2009, 2013, 2017), a ASI tem como objetivo principal a estruturação de dados no cruzamento de indivíduos e variáveis, a partir da contingência de regras e determinando os conceitos de intensidade e envolvimento, a coesão de classe, o significado dos níveis hierárquicos, a contribuição de adicionais etc. Este método de análise permite extrair as relações entre sujeito e variáveis, bem como a associação entre variáveis medido através de um índice de qualidade dessa associação (Dalla Porta, 2019).

Segundo Gras e Almouloud (2005), a ASI possibilita classificar os elementos de um banco de dados, assim como estabelecer um grau de implicação entre as variáveis desse banco, isto é, em que medida uma variável implica em outra. Para isso, a ASI tem encontrado apoio dentro da informática no *software* CHIC (Classificação Hierárquica, Implicativa e Coesiva) cuja construção foi iniciada por Régis Gras em 1985, aperfeiçoado em seguida dentro das teses de Saddo Ag Almouloud (1992) e de Harrison Ratsimba-Rajohn (1992).

O *software* CHIC é composto por uma série de funções (cálculo de parâmetros de estatística básica) e resulta na construção de uma árvore de similaridades, onde são empregados índices de probabilidade (Dalla Porta, 2019). Desse modo, o *software* permite construir a partir de um banco de dados, diferentes formas de representar os valores das intensidades de implicação associadas a cada par de variáveis principais obtidas, são elas: árvore de similaridades; árvore de coesão; grafo implicativo.

A representação que optamos em nossa investigação refere-se ao grafo implicativo, conforme Figura 1, em que o resultado traduz a rede de relações quase implicativas entre as variáveis, e o intervalo de confiança da aparição de arcos e flechas do grafo é controlável. Adotamos a implicação segundo a teoria clássica e, em função do número limitado da amostra, optamos para o tratamento a modelagem probabilística da variável aleatória utilizando uma distribuição binomial, com o grau de implicação de $\geq 0{,}75$, ou seja, consideramos implicações a partir de 0,75, que, conforme Gras, Régnier e Guillet (2009) possui um valor significativo do ponto de vista estatístico.

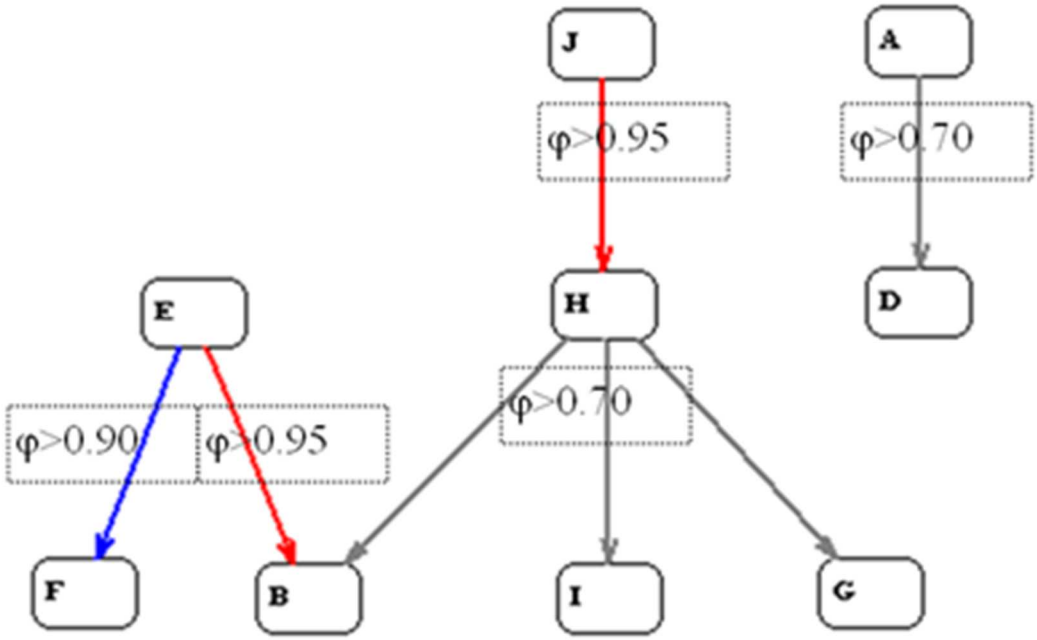


Figura 1 – Exemplo de grafo implicativo obtido com o *software* CHIC. Fonte: Gras e Régnier (2015, p.26).

Segundo Almouloud, Gras e Régnier (2014, p.625) "a Análise Estatística Implicativa, tem por objetivo a extração de conhecimentos, invariantes, e atribuição de uma medida para proposições do tipo: 'quando A está escolhido, tende-se a escolher B',". Diante dessa proposição, acreditamos que a ASI nos permitirá inferir sobre as concepções que os professores de matemática manifestam quanto a álgebra e seu ensino.

4 Análise e discussão dos resultados obtidos pelos tratamentos ASI

Com as respostas obtidas pelo questionário, o *software* CHIC 7.0 nos permitiu analisar os dados a partir do grafo implicativo, cuja intensidade de implicação considerada foi ≥0,75 que, conforme Gras, Régnier e Guillet (2009) possui um valor significativo do ponto de vista estatístico, como mostra a Figura 2.

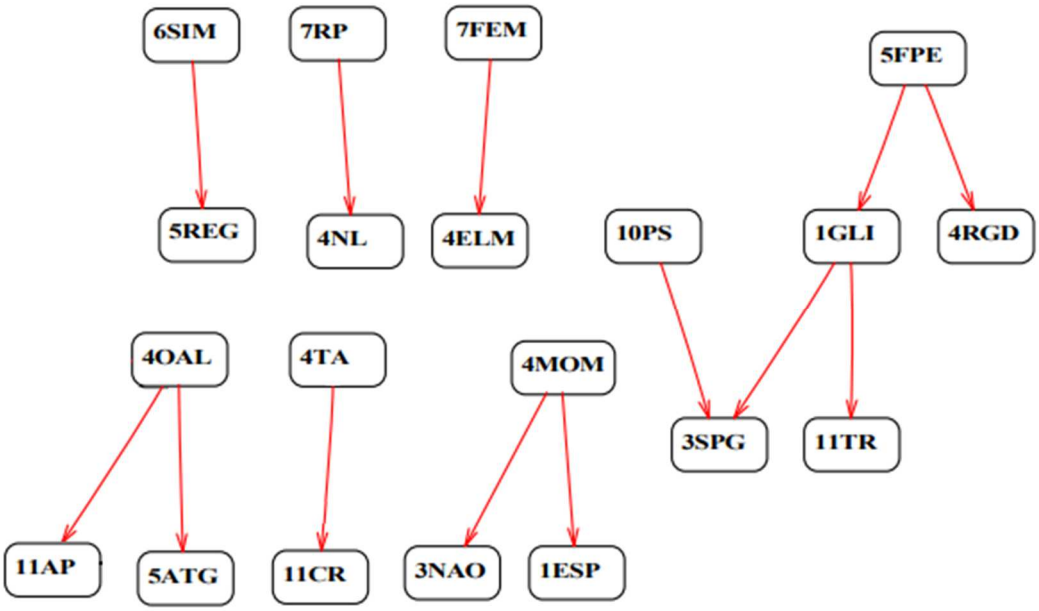


Figura 2 – Grafo implicativo. Fonte: dados da pesquisa.

Considerando o grafo implicativo construído, pode-se analisar os seguintes caminhos:

O caminho **6SIM→5REG** mostra a implicação de professores que já se apropriaram das orientações da BNCC, quanto ao ensino de álgebra, a definem como sendo a relação entre grandezas que permite estruturar algebricamente procedimentos para resolver certos tipos de problemas. Pode-se observar, portanto, que há possível tendência desses professores conceberem a álgebra dentro das quatro concepções de Usiskin (2003) em que a álgebra é entendida como: aritmética generalizada; estudo de procedimentos para resolver problemas; estudo das relações entre grandezas; e estudo das estruturas.

Os caminhos **4OAL→5ATG** e **4OAL→11AP** mostram uma possível implicação de professores que relacionam sua concepção de álgebra às operações aritméticas com letras, tanto definem a álgebra como sendo uma atividade de generalização que proporciona uma variedade de ferramentas de representação, como utilizam os problemas com os quais aprenderam álgebra em suas atividades didáticas. Pode-se observar, portanto, que há indícios de que esses professores manifestam a concepção da aritmética generalizada de Usiskin (2003). Além disso, observa-se que definem a álgebra conforme a vertente da Aritmética Generalizada de Kaput (2008) uma vez que a generalização está no centro e é fruto das operações aritméticas e suas propriedades.

Já no caminho **4TA→11CR** tem-se a tendência de implicação de professores que relacionam sua concepção de álgebra com as noções de transformismo algébrico criam seus próprios problemas para utilização em aula. Nessa concepção, que Fiorentini, Miguel e Miorim (1993) denominam de linguístico-pragmática, é necessário e suficiente que o aluno compreenda os processos de transformação algébrica, empregando regras e propriedades válidas na busca da resolução de problemas. Portanto, esta implicação pode indicar que apesar dessa concepção caracterizar a álgebra da metade do século XX, os professores consideram importante trabalhar o caráter instrumental da álgebra.

Além disso, o caminho **4MOM→3NAO** mostra a possível implicação de professores que relacionam sua concepção de álgebra com a modelagem matemática e não tiveram disciplinas voltadas para o ensino de álgebra em sua formação. Segundo Kaput (2008), a modelação é uma das vertentes do pensamento algébrico e representa a generalização a partir de situações matemáticas ou de fenômenos. Portanto, ainda que sem uma formação para a didática, há indícios de que estes professores concebem a álgebra como um processo de generalização, conforme propõe Blanton e Kaput (2005).

O caminho **5FPE→4RGD** pode indicar a implicação de professores que definem a álgebra como uma forma de pensamento e raciocínio acerca de situações matemáticas a concebem como uma relação entre grandezas distintas. Ou seja, é possível que haja uma implicação entre a definição de álgebra conforme propõe Kieran (2007) e a terceira concepção de Usiskin (2003), em que a álgebra se preocupa com a construção de modelos e leis funcionais que descrevem ou representam as relações entre variáveis.

Os caminhos **5FPE→1GLI→3SPG** e **5FPE→1GLI→11TR** também podem indicar a implicação de professores que definem a álgebra como sendo uma forma de pensamento e raciocínio acerca de situações matemáticas à professores que tiveram na sua formação alguma disciplina voltada para o ensino de álgebra e que buscam seus problemas em artigos científicos, dissertações e teses para trabalhar em sala de aula. Pode-se observar, portanto, que é possível que a definição de álgebra que esses professores manifestam, assim como proposta por Kieran (2007), também está relacionada a didática da matemática, ao "aprender a ensinar" e não somente ao "saber resolver". Isto é, muitas vezes a formação em matemática prioriza o conhecimento do conteúdo em detrimento das metodologias de ensino. Além disso, o caminho **5FPE→1GLI→11TR** pode indicar que conceber a álgebra como uma forma de pensamento implica na busca por problemas que satisfaçam esse objetivo de ensino em materiais produzidos no campo científico.

Os caminhos **7RP→4NL** e **7FEM→4ELM** indicam para possíveis implicações de professores que propõem aos seus alunos atividades de álgebra visando a resolução de problemas, relacionam sua concepção de álgebra com o uso de números e letras. Enquanto, professores que propõem aos seus alunos atividades de álgebra visando formar a base para o Ensino Médio, relacionam sua concepção de álgebra com o estudo das estruturas lógico-matemáticas. Aqui há possibilidade de que o objetivo das atividades propostas pelos professores seja reflexo de suas concepções sobre a álgebra. Isto é, para o objetivo de resolver problemas associa-se à concepção da álgebra como estudo de procedimentos para resolução de problemas, de Usiskin (2003) em que o objetivo é resolver e simplificar. Por outro lado, para o objetivo de formar a base para o Ensino Médio associa-se à concepção da álgebra como estudo das estruturas, em que o objetivo é manipular e justificar, alinhada também com a concepção fundamentalista-estrutural de Fiorentini, Miguel e Miorin (1993).

Por fim, pelo caminho **10PS→3SPG** observamos uma possível implicação de professores que consideram os problemas contidos em livros didáticos pouco satisfatórios tiveram em sua formação alguma disciplina voltada para o ensino de álgebra. Isso corrobora com os argumentos expostos anteriormente, uma vez que a formação para o ensino de álgebra possibilita um olhar mais atento às metodologias e materiais de ensino.

Diante do exposto, pode-se inferir que as concepções dos professores sobre a álgebra a constituem como uma atividade de generalização, dada a partir da relação entre grandezas distintas, que visa uma forma de pensamento e raciocínio acerca de situações

matemáticas. Isso vai ao encontro das concepções de álgebra de Usiskin (2003), da definição de álgebra de Kieran (2007) e das orientações da BNCC. Entretanto, destaca-se que, apesar desse alinhamento entre as concepções de álgebra da base teórica e as manifestadas pelos professores, o *software* não apresentou uma implicação (dentro do parâmetro de implicação$\geqslant$0,75) entre as variáveis **5FPE** e **7DPA**, as quais indicariam que professores que definem a álgebra como uma forma de pensamento e raciocínio acerca de situações matemáticas propõem aos seus alunos atividades de álgebra que visam o desenvolvimento desse modo de pensar.

A ausência dessa implicação sugere uma "quebra" no processo de didatização do saber de álgebra, uma vez que há indícios de que os professores a concebem como uma forma de pensamento, mas não visam explorar meios de desenvolvê-la. Nesse sentido, infere-se que, apesar de manifestarem a álgebra como uma forma de pensamento em seus discursos, os professores não sabem ainda como lidar metodologicamente com seu desenvolvimento na prática pedagógica. Possíveis razões para isso podem ser a recente concepção de álgebra em que se objetiva o desenvolvimento do pensamento algébrico, a recente chegada dos livros didáticos do PNLD/20 nas escolas, e ainda, as poucas investigações sobre o tema em nível de formação pedagógica e programas de pós-graduação.

5 Considerações finais

Neste trabalho, aplicou-se um questionário *online* cujo objetivo era investigar a respeito das concepções sobre a álgebra e seu ensino a partir de informações fornecidas por professores de matemática do Ensino Fundamental. Nesse contexto, houve indícios de que os professores participantes da pesquisa concebiam a álgebra como uma forma de pensamento, contudo não visavam meios didáticos de desenvolver esse modo de pensar, e com isso, retomavam o ensino de álgebra de modo técnico, em que o objetivo era formar a base necessária para o Ensino Médio. Assim, tal antagonismo pode provocar uma "quebra" no processo de didatização da álgebra.

Perante isso, percebeu-se que os professores estavam atualizados quanto às orientações da BNCC para a álgebra, porém, não haviam consolidado um método didático que possibilitasse o desenvolvimento do pensamento algébrico. Pode-se dizer que isso ocorreu devido à BNCC estar em fase de implementação, na qual os professores estão vivenciando as formações sobre o documento e participando de sua consolidação ao nível de escola. Salienta-se, também, que no contexto de isolamento social e suspensão das aulas presenciais neste ano, devido à pandemia da Covid-19, os professores desenvolveram os conteúdos listados na BNCC de modo remoto, o que modificou suas metodologias de ensino e pode ter influenciado seus discursos nesta pesquisa.

Referências

[1] Aguiar, M. (2014). *O percurso da didatização do pensamento algébrico no Ensino Fundamental:* uma análise a partir da Transposição Didática e da Teoria Antropológica do Didático (Tese de doutorado). Faculdade de Educação, Universidade de São Paulo, São Paulo, SP, Brasil.

[2] Almeida J. R., & Câmara, M. S. (2017). Pensamento algébrico: em busca de uma definição. *Revista Paranaense de Educação Matemática*, 6(10), 34-60. Recuperado de http://rpem.unespar.edu.br/index.php/rpem/article/view/1124.

[3] Araujo, E. A. (2008). Ensino de álgebra e formação de professores. *Educação Matemática Pesquisa,* 10(2), 331-346. Recuperado de https://revistas.pucsp.br/index.php/emp/article/view/1740.

[4] Blanton, M. L., & Kaput, J. (2005). Characterizing a classroom practice that promotes algebraic reasoning. *Journal for Research in Mathematics Education*, 36(5), 412-446. Recuperado de https://www.researchgate.net/publication/285754350_Characterizing_a_classroom_practice_that_promotes_algebraic_reasoning.

[5] Brasil. (1997). Secretaria de Educação Fundamental. *Parâmetros curriculares nacionais:* matemática/Secretaria de Educação Fundamental. – Brasília: MEC/SEF. Recuperado de http://portal.mec.gov.br/seb/arquivos/pdf/livro03.pdf.

[6] _____. (2017). Ministério da Educação. *Base Nacional Comum Curricular.* Versão Final. Brasília: MEC. Recuperado de http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=79601-anexo-texto-bncc-reexportado-pdf-2&category_slug=dezembro-2017-pdf&Itemid=30192.

[7] Claras, A. F., & Pinto, N. B. (2008). O Movimento da Matemática Moderna e as iniciativas de formação docente. *Anais do VIII Congresso Nacional de Educação da PUC-PR* – Educere, Curitiba. Recuperado de https://educere.bruc.com.br/arquivo/pdf2008/863_662.pdf.

[8] Coelho, F. U., & Aguiar, M. (2018). A história da álgebra e o pensamento algébrico: correlações com o ensino. *Revista Estudos Avançados*, 32(94). Recuperado de http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142018000300171.

[9] Dalla Porta, L. (2019). *Formação do raciocínio estatístico na conceptualização da estimação estatística:* estudo exploratório de um dispositivo pedagógico no ensino superior (Tese de doutorado). Universidade Franciscana, Santa Maria, RS, Brasil.

[10] Eves, H. (2010). *Introdução a História da Matemática.* Campinas: Editora Unicamp.

[11] Fiorentini, D., Miorin, A., & Miguel, A. (1993). Contribuição para um Repensar a Educação. Algébrica Elementar. *Pró-posições*, 4(1), 78-91. Recuperado de https://www.fe.unicamp.br/pf-fe/publicacao/1761/10-artigos-fiorentinid_etal.pdf.

[12] Gomes, M. L. M. (2013). *Álgebra e funções na educação básica.* Belo Horizonte: CAED – UFMG.

[13] Kaput, J. (2008). What is algebra? What is algebraic reasoning? In: Kaput, J., Carraher, D., & Blanton, M. (Orgs.), *Algebra in the Early Grades*. Nova York: Lawrence Erlbaum Associates.

[14] Kieran C. (2007). Developing algebraic reasoning: the role of the sequenced tasks and teacher questions from the primary to the early secondary school levels. *Quadrante,* 16(1), 5-26.

[15] Minayo, M. C. S., & Sanches, O. (1993). Quantitativo-qualitativo: oposição ou complementaridade? *Cadernos de Saúde Pública*, 9(3), 239-262. Recuperado de https://www.scielo.br/j/csp/a/Bgpmz7T7cNv8K9Hg4J9fJDb/?lang=pt&format=pdf.

[16] Piaget, J. (1971). *A Epistemologia Genética.* Trad. Nathanael C. Caixeira. Petrópolis: Vozes.

[17] Santos A. O. (2016). *A álgebra no ensino fundamental como ferramenta de generalização* (Dissertação de mestrado). Universidade Federal de Alagoas, Maceió, AL, Brasil.

[18] Santos, C. A. O., & Borges, M. F. (2013). Evolução da simbologia algébrica: um passeio pela história evolutiva do pensar matemático humano. *Anais do XVII Encontro Brasileiro de Estudantes de Pós-graduação em Educação Matemática*. Recuperado de http://www.editorarealize.com.br/revistas/ebrapem/trabalhos/0c3d4ab019fbd17a8287b8cffe8ae4ec.pdf.

[19] Trajano, A. (1947). *Álgebra Elementar.* 22 ed. Rio de Janeiro: Livraria Francisco Alves.

[20] Usiskin, Z. (2003). Concepções sobre a álgebra da escola média e utilizações das variáveis. In: Coxford, A. F., & Shulte, A. P. (Orgs.), *As ideias da álgebra.* Trad. Hygino H. Domingues. São Paulo, SP: Atual.

[21] Vailati. J. S., & Pacheco, E. R. (2011). *Usando a história da matemática no ensino da álgebra.* Curitiba: Secretaria de Estado da Educação. Recuperado de http://www.diaadiaeducacao.pr.gov.br/portals/pde/arquivos/702-4.pdf.

[22] Zabala, A. (1998). *A prática educativa:* como ensinar. Porto Alegre: Editora Artes Médicas Sul Ltda.

CHANGEMENT RADICAL DES MÉTHODES D'ENSEIGNEMENT UNIVERSITAIRES VERS LE NUMÉRIQUE FACE À LA PANDÉMIE DE COVID-19 À MADAGASCAR : Cas de l'université de Antsiranana

Bakys Bruno Ralahady[1] , Christian Razanatsoavina[2], Andrianasy Angelo Djistera[3], Jean-Claude Laberche[4], André Totohasina[5]

RADICAL CHANGE IN UNIVERSITY TEACHING METHODS TOWARDS THE DIGITAL COVID-19 PANDEMIC IN MADAGASCAR: CASE OF THE UNIVERSITY OF ANTSIRANANA

RÉSUMÉ

La pandémie de covid-19 génère une crise profonde du secteur de l'éducation, dans les Universités de Madagascar. Les modes de transmissions pédagogiques ont changé radicalement en privilégiant les outils numériques, sans prendre en compte la non-application complète du système de LMD dans le système éducatif du pays. L'objectif de cet article est d'étudier et de mettre en évidence les réactions des étudiants face aux obligations de changement de méthode d'apprentissage lors de la pandémie de covid-19. Sont ainsi rapportées dans cet article : 1) Les études d'acquisition de données au moyen d'un questionnaire proposé à 200 étudiants confinés, 2) L'analyse statistique des résultats dont l'ASIMGK et leurs conséquences sur la gestion des affaires universitaires. Les résultats précisent une analyse des réalités du terrain. D'abord, une partie des étudiants a dû assumer les coûts générés pour une connexion Internet et a été obligée de fréquenter des cyber-cafés, pour accéder aux documents mis en ligne. Ces étudiants ont également changé leurs stratégies d'apprentissage. En conséquence, la majorité des étudiants a ressenti une baisse de leur niveau et estime insuffisante leur formation. Malgré tout, la majorité des étudiants ne déclare pas avoir l'intention d'abandonner leurs études même s'ils rencontrent des difficultés.

Mots-clés : *Codiv-19, TIC, méthode pédagogique, enseignement supérieur, ASI.*

ABSTRACT

The pandemic of covid-19 generates a deep crisis in the education sector, in the Universities of Madagascar. The modes of pedagogical transmissions have changed radically to the benefit of digital, without taking into account the non-complete application of the LMD system in the educational system. The objective of this article is to show and to study the reactions of the students in front of the obligations to change their learning method during the pandemic of Covid-19. The following are reported in this article: 1) Data acquisition studies with a questionnaire proposed to 200 confined students, 2) Statistical analysis of the results including the ASI MGK and their implications on the management of academic affairs.

The results specify an analysis of the realities of the field. First, a large proportion of students spent money on an internet connection and were forced to go to cyber cafes to search for

[1]Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, ralahadybru@yahoo.fr

[2]Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, cmrazanatsoavina@gmail.com

3Faculté de Droit, d'Economie, de Gestion et de Mathématiques, Informatique et Applications, Université de Toamasina, angelo.djistera@yahoo.fr

4Professeur des Universités, émérite, jeanclaudelaberche@gmail.com

5Ecole Normale Supérieure pour l'Enseignement Technique, Université d'Antsiranana-Madagascar, andre.totohasina@gmail.com

materials; they also changed their learning strategies. As a result, the majority of students has noticed a decline in their level and feels that their training is insufficient. Despite this, the majority of students do not intend to abandon their studies even if they encounter difficulties.

Keywords: *Codiv-19, TICE, teaching method, high school, ASI.*

1 Introduction

Dès l'annonce officielle du confinement à Madagascar suite à la progression de la Covid 19 dans la Grande Ile, , le gouvernement a pris des mesures entraînant la fermeture générale des écoles et universités (Instat, 2020). Ceci c'est traduit par des activités pédagogiques et administratives ralenties dans l'ensemble du pays (Plan Multisectoriel D'urgence Madagascar, 2020). Ces mesures ont grandement perturbé les méthodes et le fonctionnement des enseignements et des apprentissages. Devant cette fermeture, les parents et les étudiants s'inquiétèrent grandement de leur avenir surtout que cette suspension des cours durant une longue et difficile période de crise fut l'occasion d'une révision des stratégies éducatives que ce soient celles du Ministère de l'Education Nationale ou celles de l'Enseignement Supérieur et de la Recherche Scientifique.

Cette révision des stratégies éducatives s'inscrit parfaitement dans le processus de «Bologne», mise en œuvre depuis plus d'une dizaine d'années par le gouvernement malgache en 2008, comme l'a évoqué Razafimbeloen (2019). Il est caractérisé au niveau des étudiants par une diminution des enseignements en présentiel au profit de recherches personnelles qui devraient représenter à échéance plus ou moins lointaine près de 80% du temps consacré à l'éducation (Andriamampianina, 2012).

Comme le détaille très bien Andriambelo, (2014), ce nouveau LMD impose aux établissements publics universitaires de disposer d'infrastructures et d'outils liés aux Technologies de l'Information et de la Communication(TIC) pour que les étudiants assurent leurs recherches personnelles tout au long de leur parcours universitaires. Mais force est de constater que la situation des universités au moment de la pandémie en bibliothèques, médiathèques et même réseaux WiFi n'était guère brillante et le confinement imposé, n'a fait qu'augmenter le manque d'organisation permettant d'assurer l'enseignement selon ces TIC.

Lors de l'annonce du déconfinement, en juin 2021, la reprise de l'année universitaire a été annoncée par le gouvernement avec pour seule contrainte le respect des mesures sanitaires, donc l'interdiction des regroupements de plus de 100 personnes. Libres à toutes les universités malgaches de choisir leurs propres modalités pour terminer les cours et les examens de fin d'études. Ceci c''est immanquablement traduit par un renforcement du travail individuel des étudiants au détriment des cours en présentiel.

Notre article se focalise sur les étudiants de l'université d'Antsiranana (au nord de Madagascar) qui, dès le premier confinement, ont subi de multiples changements dans leur manière de travailler consécutivement aux nouvelles orientations pédagogiques basées sur les TIC et décrétées par l'université

L'objectif de notre étude est de montrer l'impact de ces changements brutaux sur les étudiants quant aux méthodes pédagogiques développées sous la double contrainte des assauts du virus, toujours présent, et des mesures sanitaires difficiles à appliquer. Pour connaître ces impacts, qui modifient les modes d'acquisition avec le développement du

numérique, les étudiants ont été interrogés au moyen d'un sondage, d'abord sur leur perception (positive négative ou constante) des changements de stratégies pédagogiques, puis sur les transformations matérielles de leur vie quotidienne que ces changements induisent. Enfin, synthèse de tous ces changements, leur ressenti d'abandon ou d'amélioration de leurs études lors de cette période.

Pour approfondir les décomptes sur les observations obtenues lors du dépouillement des données du sondage, une analyse statistique utilisant la méthode ASI-MGK permet de préciser les résultats. Cette méthode permet de mesurer la qualité des règles d'association (Ralahady et al, 2019a)

2 Matériels et méthodes

Dans cette étude, il y a obligatoirement deux phases qui se succèdent:

- mesurer les changements subis par la population des étudiants lors du confinement quant aux modifications induites dans les méthodes pédagogiques, c'est le sondage
- dépouiller, analyser les résultats du sondage.

A chaque phase correspond des matériels et des méthodes spécifiques.

2.1 Mesurer les vécus et les changements des stratégies d'apprentissage chez les étudiants lors du confinement

Fin mars 2021, nous avons réalisé une enquête auprès de 200 étudiants, sur un total d'environ 400 étudiants présents étayant été confinés, sur le campus universitaire de l'Université d'Antsiranana, située au Nord de Madagascar. L'étude a été réalisée sans distinguer le sexe ni l'âge, ni leur établissement d'origine ni le niveau d'études.

Nous avons choisi un type de questionnaire semi-directif, avec une réponse ouverte ou fermée. Dans l'objectif de l'étude, il est important de savoir comment les étudiants ont fait pour adapter leurs études face aux bouleversements brutaux du mode de transmission vers le numérique lors de la pandémie du covid-19 et notamment durant le confinement prolongé. Les questions sur la connexion internet, les frais de connexion, les fréquences, leur adaptation et, leur motivation sont abordées avec un caractère subjectif clair. Le détail du questionnaire est en annexe 1.

Nous avons présenté notre questionnaire, porte à porte chez les étudiants présents dans leurs logements respectifs et pour avoir leurs réponses, nous leur avons demandé de remplir directement et librement le questionnaire ramassé immédiatement.

Avec un échantillon proche de la moitié de la population cible nous considérons notre échantillon comme représentatif. Ceci ne l'empêche pas d'avoir des limites. En effet l'échantillon peut ne pas représenter exactement la composition de la population des 2000 étudiants de l'Université d'Antsiranana. Au moment du sondage, certains d'entre eux avaient quitté le campus universitaire pour retourner chez leurs parents et certains autres ont eu peur d'ouvrir leur porte. Cette dernière catégorie a constitué un frein inattendu à l'obtention d'un plus grand nombre d'étudiants interrogés.

2.2 Analyses des données

Les données issues des questionnaires sont dépouillées selon les méthodes décrites par (Ralahady *et al*. 2019c). Le premier tableau d'enregistrement des données permet de connaître plusieurs caractéristiques intéressantes de l'échantillon. Comme nous sommes intéressés à évaluer la dépendance entre deux variables qualitatives binaires, nous avons utilisés deux méthodes différentes:

2.2.1 Le tableau de contingence

C'est un tableau à double entrée permettant d'estimer la dépendance entre deux caractères.

Pour illustrer notre propos, on choisit, comme exemple, les données des deux variables présentes à deux niveaux + ou -):la proposition de cours en ligne (PropLign+ ou PropLign-) et l'idée d'abandon des études ou non (PensAbd+ ou PensAbd-)

	Effet+ (Y)	Effet– (Y)
Cause +(X)	A	B
Cause - (X)	C	D

	PensAbdNon+	PensAbdNon-
PropLign+	100	29
PropLign-	40	31

Tableau 1: Principe (à gauche) et résultats (à droite)des tableaux de contingences pour deux variables (proposition de cours en ligne et idée d'abandon des études) présentes dans le questionnaire à deux niveaux (+ ou -)

Nous utilisons les notations suivantes :

- PropLign+ PensAbdNon + (a) : nombre d'étudiants (100) à qui on a proposé les cours en ligne et qui ne pensent pas abandonner leurs études.
- PropLign- PensAbdNon + (b) : nombre d'étudiants (40) à qui on n'a pas proposé les cours en ligne et qui pensent abandonner leurs études.
- PropLign+ PensAbdNon - (c) : nombre d'étudiants (29) à qui on a proposé les cours en ligne et qui ne pensent pas abandonner leurs études.
- PropLign- PensAbdNon - (d) : nombre d'étudiants (31) à qui on n'a pas proposé les cours en ligne et qui ne pensent pas abandonner leur étude.
- Effectif (N) : nombre total d'individus 200

2.2.2 Analyse selon le risque relatif et l'odds ratio

Une approche non paramétrique permettant de mesurer l'association entre deux variables binaires a été effectuée grâce aux calculs du Risque Relatif, relié à l'Odds Ratio.

Le risque relatif RR (relative risk, R.R)

C'est le risque de survenue d'un événement (comme l'abandon) dans un groupe exposé à un facteur d'exposition par rapport à un autre groupe non exposé. Le risque relatif est un ratio qui mesure à quel point le test se comporte mieux lorsqu'il est positif par rapport au cas où il est négatif.

Avec les données de variables: proposition de cours en ligne (PropLign+ou PropLign-) et l'idée d'abandon ou non (PensAbd+ou PensAbd-), on obtient :

$$RR = \frac{a/(a+b)}{c/(c+d)}, RR = \frac{100/(100+29)}{40/(29+40)} = 1.376$$

L'odds ratio OR (*odds ratio*)

C'est le rapport de la cote d'exposition chez les sujets ayant présenté l'événement (Abandon Non) et ceux ne l'ayant pas présenté (Abandon Oui). L'odds ratio indique à quel point un individu a plus de chance d'être positif si le test est positif, par rapport au cas où le test est négatif.

Toujours avec le même exemple que pour les tableaux de contingence, on obtient:

$$OR = \frac{a/c}{b/d}, OR = \frac{100/40}{29/31} = 2.672$$

Ces deux indices s'interprètent de la même façon. L'intervalle de variation est compris entre [0 à +∞[:

- Si =1, il n'y a pas de dépendance (l'association est non significative)
- Si <1, l'association est négative
- Si >1, l'association est positive

Dans les données présentées comme exemple, nous avons le *relative risk* et l'*odds ratio* tous les deux supérieurs à 1, ce qui signifie une dépendance significative entre la proposition de cours en ligne et non abandon d'étude.

2.2.3 Analyse selon M_{GK} (Mesure de Guillaume Khenchaf)

La méthode statistique ASI-MGK est l'une des méthodes d'Analyse Statistique Implicative utilisant la mesure de qualité des Règles d'Association (RA)M_{GK}. D'abord, elle optimise la validation et l'évaluation de la force d'attraction ou de répulsion entre deux couples de modalités. Elle est très sélective quant au choix d'orientation d'une relation qui peut avoir deux couples de modalité ($A \longrightarrow B$ *ou* $B \longrightarrow A$, ...). Ainsi, pour une RA de type $A \longrightarrow B$: la modalité A est l'antécédent, la modalité B est le conséquent, la probabilité P(A) est sa couverture et P(B) est sa prévalence et finalement M_{GK} est sa force. Nous rappelons ci-après comment on calcule ces paramètres. (Ralahaddy et al; 2021b).

Couverture $P(X)$ — C'est le ratio du nombre de transactions sur le nombre de transactions supportant l'antécédent

$$P(X) = \frac{a+c}{n}, \quad P(PropLign) = \frac{100+40}{200} = 0.7$$

Prévalence P(Y) —C'est le ratio du nombre de transactions sur le nombre de transactions supportant le conséquent.

$$P(Y) = \frac{a+b}{n}, \quad P(\mathrm{PensAbd}) = \frac{100+29}{200} = 0.64$$

Confiance $Conf$ — C'est le ratio du nombre de transactions supportant la règle (antécédent et conséquent) sur le nombre de transactions supportant l'antécédent.

$$Conf(X \longrightarrow Y) = \frac{P(X \cap Y)}{P(X)} = \frac{a}{a+c}$$

Si $Conf(X \longrightarrow Y) > P(Y)$ alors l'évènement X favorise l'évènement Y

$$Conf(\mathrm{PensAbd} \longrightarrow \mathrm{PropLign}) = \frac{100}{100+40} = 0.71$$

$$Conf(\text{PropLign} \longrightarrow \text{PensAbd}) = \frac{100}{100+29} = 0.77$$

Ici $Conf(\text{PensAbd} \longrightarrow \text{PropLign}) > P(\text{PropLign})$ et $Conf(\text{PropLign} \longrightarrow \text{PensAbd}) > P(\text{PensAbd})$ alors l'évènement PensAbd et évènement PropLign se favorisent.

M_{GK} —Mesure de Guillaume Khenchaf (2000), C'est un ratio de probabilité conditionnelle incrémentée (CPIR), mesure implicative, orientée et normalisée (ION): définie par Totohasina (2008), dont sa partie favorisante est définie par **:**

$$M_{GK}^{f}(X \longrightarrow Y) = \frac{na - (a+b)(a+c)}{(a+c)(n - a - b)}$$

Dans notre approche, une règle d'association $X \longrightarrow Y$ est donc dite valide si $P(X \cap Y)$ est supérieur au seuil minimal de support, et $M_{GK}^{f}(X \longrightarrow Y) > M_{GK}^{f}cr$.

La valeur critique de M_{GK}^{f} s'appuie sur la statistique de Khi-carré (χ^2) de Pearson à un degré de liberté, (Ralahady 2019)

$$M_{GK}^{f}cr = \sqrt{\frac{(n-a-c)(a+b)}{n(a+c)(n-a-b)}}\chi_{\alpha}$$

Ladite règle est appelée «règle exacte» si son M_{GK}^{f} est 1, et approximative, dans le cas contraire.

$$M_{GK}^{f}(\text{PropLign} \longrightarrow \text{PensAbd}) = \frac{200 \times 100 - (100+29)(100+40)}{(100+40)(200 - 100 - 29)} = 0.25$$

$$M_{GK}^{f}(\text{PensAbd} \longrightarrow \text{PropLign}) = \frac{200 \times 100 - (100+40)(100+29)}{(100+29)(200 - 100 - 40)} = 0.19$$

$$M_{GK}^{f}(\text{PropLign} \longrightarrow \text{PensAbd}) > M_{GK}^{f}(\text{PensAbd} \longrightarrow \text{PropLign})$$

Et en calculant la valeur significative de $M_{GK}^{f}(\text{PropLign} \longrightarrow \text{PensAbd})$ au risque d'erreur égal à 5%, nous obtenons :

$$M_{GK}^{f}cr = \sqrt{\frac{(200-100-29)(100+40)}{200(100+29)(200-100-40)}}0.45 = 0{,}053$$

Ainsi $M_{GK}^{f}(\text{PropLign} \longrightarrow \text{PensAbd})$ supérieur à $M_{GK}^{f}cr$ traduit une dépendance orientée positive de PropLignquiimplique PensAbd; c'est donc le caractère significatif et implicatif de cette association.

Pour réaliser nos analyses, nous avons utilisé deux outils d'analyses statistiques à savoir : l'outil ASI-MGK pour extraire et évaluer les Règles d'Associations valides selon la mesure de Guillaume Khenchaf et le logiciel Epi Info version 7.0 pour vérifier les dépendances des couples de modalité extraites selon l'analyse statistique classique utilisant le Risque Relatif et l'Odds Ratio.

3 Présentation des résultats et interprétations

Avant de présenter les résultats des analyses statistiques effectuées sur les données du sondage, nous présenterons d'abord les caractéristiques de notre échantillon à partir des statistiques générales observées lors du dépouillement des données.

3.1 Les caractéristiques de notre échantillon.

C'est le logiciel Epi info 7 qui nous a fourni une approche de notre échantillon. L'ensemble des statistiques pour chaque critère étudié est présenté sous forme de diagrammes circulaires sectorisés en annexe 1.

Ici pour mieux comprendre l'échantillon nous en retenons deux. Il s'agit du genre et des classes d'âges de l'échantillon (figure 1).

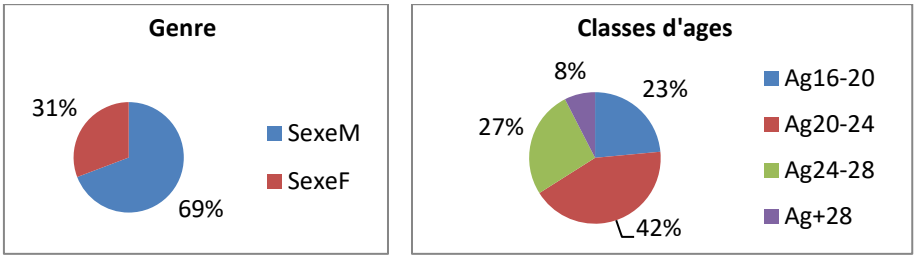


Figure 1: Genre et classes d'âge des étudiants de l'échantillon interrogé.

Nombre d'observations 200, les résultats sont exprimés en % par rapport au nombre d'observations. Nous sommes en présence d'une population très jeune dominée par le sexe masculin d'âge moyen entre 20 à 24 ans. Après ces données générales permettant d'appréhender le groupe nous préciserons les valeurs complémentaires observées lors des différentes analyses statistiques effectuées.

3.2 Les analyses statistiques

Les analyses statistiques se focalisent sur l'ASI-MGK. C'est donc tout naturellement que nous avons effectué une analyse sur toutes les règles (les relations entre les items) valides (figure 2). On pourrait même dire qu'il représente la synthèse des réactions et des changements des étudiants par rapport aux moyens de TIC de leur université. Pour l'obtenir nous avons d'abord exploré l'ensemble des relations valides selon M_{GK} au minimum support égal à 35% et au risque d'erreur égal à 5% à l'aide d'ASI-MGK. Puis nous avons analysé les dépendances de chaque paire de modalités des règles valides à l'aide de l'outil Epi Info (tableau de contingence et les analyses de tableaux simples).

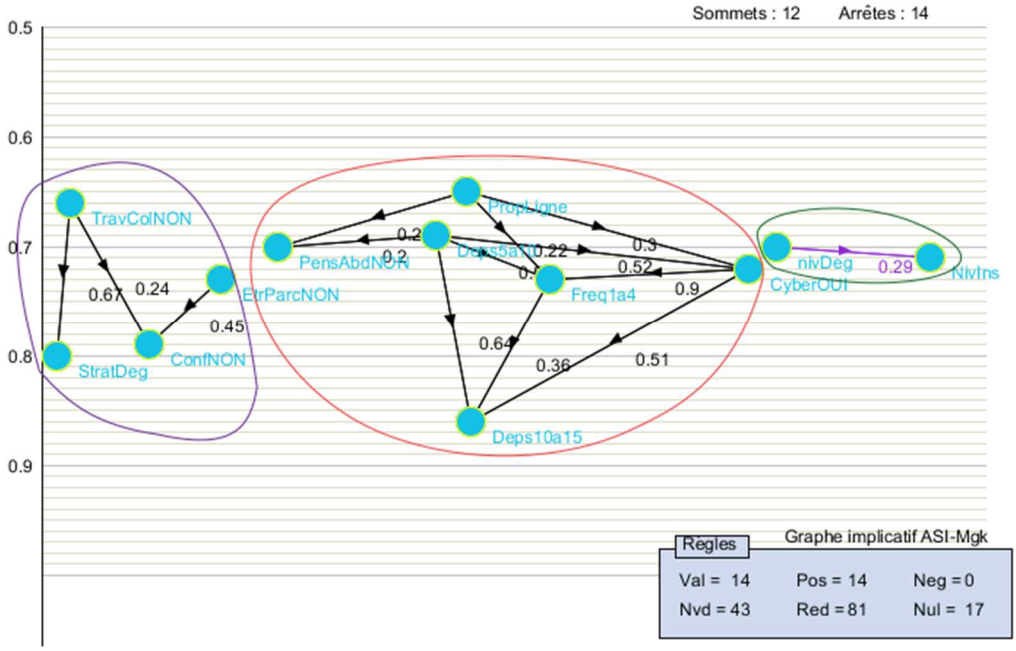


Figure 2: Ensemble des résultats de l'ASI-MGK
Il y a14 relations significatives entre 12 items dont la description sera précisée lors de chaque analyse. Le graphe comporte trois composantes connexes (classes de résultats entourées)

Légendes communes à toutes les figures présentant des ASI-MGK

- Les nœuds représentent les modalités
- Les arcs évalués représentent l'orientation et la force de la relation de deux couples de modalité
- La graduation sur l'axe vertical représente la valeur de couverture et de prévalence d'une relation.

3.2.1 Analyse 1 : Les conséquences des changements de stratégies des méthodes d'apprentissage sur leurs habitudes quotidiennes

Nous allons commencer notre analyse par la première composante (figure 3) de notre résultat d'ensemble présenté dans la figure 2.

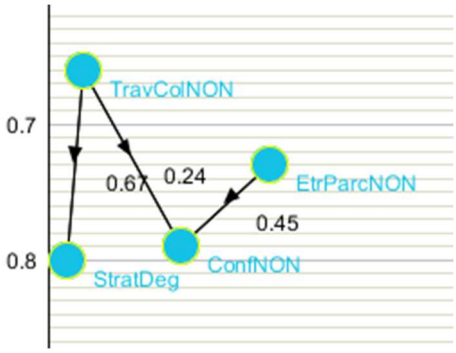


Figure 3 : Changements de stratégies et méthode d'apprentissage

Description des items :
TravColNON : Pas de travail collaboratif, P(TravColNON)=0,66
StratDeg : Stratégies et méthodes d'apprentissage dégradées, P(StratDeg)=0.80

ConfNoN : Aucune participation aux conférences en ligne P(ConfNoN)=0.79
EtrParcNoN : Aucune participation aux cours en ligne P(EtrParcNoN)=0.74

Les résultats montrent que depuis de la mise en place des barrières sanitaires pour lutter contre la propagation de pandémie du covid-19, dont l'interdiction de se regrouper 66% des étudiants interrogés ne peuvent plus faire d'études en groupe (TravColNON). Deux tiers des étudiants respectent les consignes.

Figure 4 : Résultat du croisement TravColNON × StratDeg par le logicielEpi Info ; il existe une dépendance positive et significative entre la modalité TravColNON et la modalité StratDeg

Face à ces mesures drastiques, les 80% des étudiants interrogés disent que leurs stratégies d'apprentissage et l'enseignement ont été brutalement changés. ($M^{f}_{GK}(\text{TravColNON} \longrightarrow \text{StratDeg}) = 0.67,\ \text{OR} = 12.14$ et $\text{RR} = 1.76$). En fait, l'ASI confirme bien que l'absence de travail collaboratif implique une dégradation de stratégie d'apprentissage. Lorsque l'ancienne méthode d'enseignement (en présentiel) a été interdite, les nouvelles méthodes d'enseignement et d'apprentissage basées sur les travaux personnels imposent l'utilisation des ressources du web alors que ces ressources ne sont pas ordinairement utilisées et surtout ne sont pas encore maîtrisées par la majorité des étudiants interrogés. Ces situations créent une dégradation bien ressentie par ces étudiants.

3.2.2 Analyse 2 : Le suivi des cours ou des conférences en ligne avec les enseignants

Le croisement de la figure 2 (ASI) et de l'Epi Info, montre une relation fortement significative telle que ($M^{f}_{GK}(\text{EtrParcNon} \longrightarrow \text{ConfNON}) = 0.45,\ \text{OR} = 6.82$ et $\text{RR} = 1.67$), cela signifie que le non suivi des enseignements en présentiel se retrouve aussi bien pour les cours magistraux que pour les conférences.qui complètent normalement les cours.

Ce résultat complète bien ce qui déjà apparaissait dans les tableaux de contingence et dans l'odds ration (figure 5)

Figure 5 : Résultat du croisement EtrParcNon × ConfNON de l'Epi Info .

Une dépendance positive et significative est observée entre la modalité EtrParcNon et la modalité ConfNON. Les étudiants sont obligés de rester dans leur appartement parfois sans cours. Interviennent certes les carences liées aux TCI mais aussi probablement le fait que certains enseignants n'ont pas proposé ces outils numériques et sont restés à attendre le retour à la situation antérieure pour reprendre les cours et les activités pédagogiques. Le système de télé-enseignement massif avec un système d'e-learning, notamment lors de la pandémie ne semble pas avoir fonctionné à Antsiranana. Peut-être n'en était-il pas de même dans les autres universités de la grande Ile comme ceci a été signalé par la presse à Antananarivo, la capitale. C'est le reflet des inégalités géographiques qui se traduisent dans le système éducatif. Cette constatation n'est pas nouvelle, elle avait déjà été dénoncée par Duret en 2005 et malheureusement elle perdure. Pourtant, l'éducation est aussi un droit universel et l'Etat en est le garant. Pour le moment, l'e-learning n'est pas adapté ou n'est pas à la portée de la majorité des étudiants d'Antsiranana et probablement de tout Madagascar.

3.2.3 Analyse 4 : Leur utilisation actuelle du Cyber-café

Nous poursuivons notre analyse sur la deuxième composante de la figure 2 en nous focalisant sur les items basés sur l'accès à internet des étudiants interrogés (figure6)

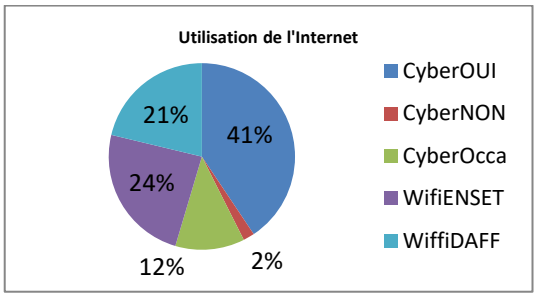


Figure 6 : L'utilisation de l'internet par les étudiants

Si peu d'étudiants n'utilisent pas internet (2%) un peu moins de la moitié (41%) vont dans les cyber-cafés en ville alors qu'un quart utilisent les ressources de l'ENSET et de la DAFF. Ces simples pourcentages montrent d'abord que l'accès à Internet est un besoin pour pratiquement tous les étudiants. Si 45% ont trouvé une solution sur le campus les autres ont dû aller en ville dans les cyber-cafés.

On peut s'étonner du faible pourcentage des étudiants ayant trouvé une solution Internet sur place. En effet comme le signale un document de l'Ambassade de France à Madagascar:

« Au niveau local, le ministère (malgache) a mis en place, avec l'appui technique et financier du projet MADES[6], une politique volontaire de développement de l'accès au numérique,(ainsi)depuis juillet 2012, les six universités publiques sont toutes dotées de réseaux de campus en fibre optique et de réseaux métropolitains de type WIMAX. »

(https://www.diplomatie.gouv.fr/IMG/pdf/MADAGASCAR_10_avril_2013x_cle01ae2c.pdf)

Effectivement 45% des étudiants confinés ont eu recours a ces réseaux mais les autres, pourquoi ont-ils choisi les cyber-cafés? Peut-être s'agit-il d'un manque d'information auprès des étudiants? Peut-être s'agit-il de problèmes techniques ou administratifs comme les horaires d'accès, le manque de personnel, ayant limité l'accès? L'enquête n'apporte actuellement aucune réponse. Il sera intéressant d'approfondir cet aspect pour rendre plus facile et plus libre l'accès des étudiants à Internet.

La figure 7 donne plus de détails sur l'accès des étudiants à Internet. Ils sont obtenus par l'analyse ASI-MGK

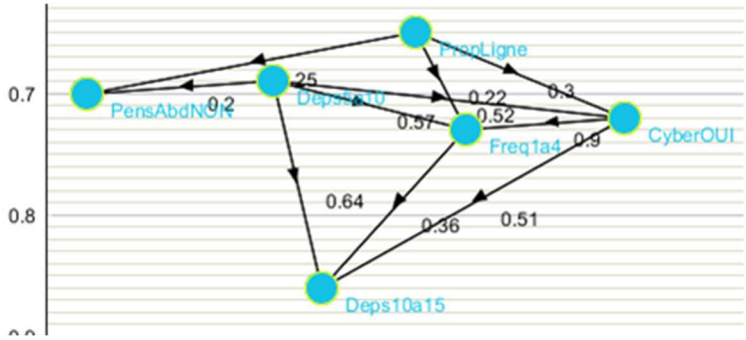


Figure 7: Accès des étudiants à l'internet

Description des items
$ProPLign$: Proposition de cours en ligne comme une nouvelle stratégie, P(ProPLign)=0,63
$PensAbdNON$: Etudiant qui ne pense pas à abandonner leurs études, P(PensAbdNON)=0,70
$CyberOUI$: Etudiants qui peuvent au moins connectés au Cyber-café, P(CyberOUI)=0,72
$Freq1a4$: Fréquentation d'une à quatre fois par semaine à l'internet, P(Freq1a4)=0,74
$Deps5a10$: Dépense de 500 à 1000Ariary par semaine à la connexion internet, P(Deps5a10)=0, 69
Deps10a15 : Dépense de 1000 à 1500Ariary par semaine à la connexion internet, P(Deps10a15)=0,87

Les valeurs de P(CyberOUI), P(Freq1a4) et P(Deps10a15) nous donnent le taux d'étudiants ayant accédé au cyber-café, le taux de fréquentation au cyber-café et puis leurs dépenses hebdomadaires de la connexion internet durant le confinement.

La règle ProPLigne ⟶ CyberOUI du graphe implicatif selon l'ASI et le tableau de contingence (figure 2) de l'Epi info montrent (figure 8) que les étudiants qui ont pu suivre

[6]Le projet MADES en collaboration avec la France, avait pour finalité d'aider l'enseignement supérieur malgache à s'internationaliser.

les enseignements en ligne ont une relation significative avec ceux qui utilisent le cyber-café ($M^{f}_{GK}(ProPLigne \longrightarrow CyberOUI) = 0.30,\ OR = 3.04$ et $RR = 1.39$).

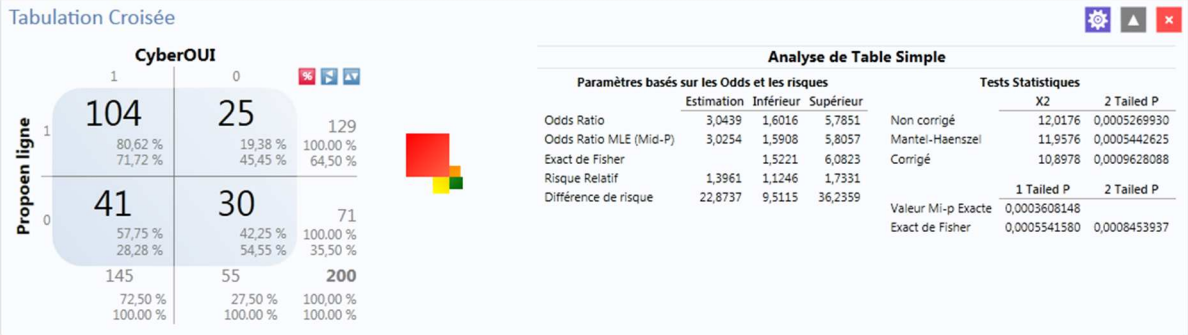


Figure 8 : Résultat du croisement ProPLigne × CyberOUI de l'Epi Info, une dépendance positive et significative entre la modalité ProPLigne et la modalité CyberOUI.

Ces résultats montrent que les étudiants ont beaucoup utilisé l'internet pour leurs études lors de la pandémie du covid-19. De plus, la règle $ProPLigne \longrightarrow Freq1a4$ où $M^{f}_{GK}(ProPLigne \longrightarrow Freq1a4) = 0.2$, signifie que la proposition de cours en ligne implique une fréquentation d'une à quatre fois par semaine à l'accès internet. En plus, la règle $Freq1a4 \longrightarrow Deps10a15$ où $M^{f}_{GK}(Freq1a4 \longrightarrow Deps10a15) = 0.22$, montreune fréquentation d'une à quatre fois par semaine d'un accès internet implique une dépense de 1000 à 1500 ariarys[7].De même l'accès à des cours en ligne est réclamée par les étudiants. Ce sont eux qui fréquentent et dépensent pour accéder à Internet.

Madagascar est déjà engagé dans le système LMD. (Gioan, 2012); toutefois, l'investissement en équipements, en infrastructure et en recrutement du personnel compétent dans le domaine des TIC malgré les efforts ne se concrétisent pas dans les enseignements (Andriambelo, 2014). La fracture numérique reste aujourd'hui un problème flagrant à Antsiranana et crée une contrainte majeure pour l'égalité des chances au sein de l'enseignement supérieur malgache. Ces manques obligent trop étudiants à payer de leurs propres deniers les charges de connexion.

3.2.4 Analyse 4 : Covid-19 et leur niveau d'études

Nous passons maintenant à l'analyse de la troisième composante de la figure 2, qui traite la relation entre l'item Niveau dégradé, (nivDeg) et l'item Niveau insuffisant (NivIns)(figure 9).

[7]1,00 Euro =4 600 Ariarys malgaches environ. Chaque étudiant boursier dispose en moyenne pour vivrede100000Ariarys par mois (cf. données de l'Université d'Antananarivo (univ-antananarivo.mg) soit 22€ soit moins de 0,75€par jour pour se nourrir, se loger et avoir unevie sociale.

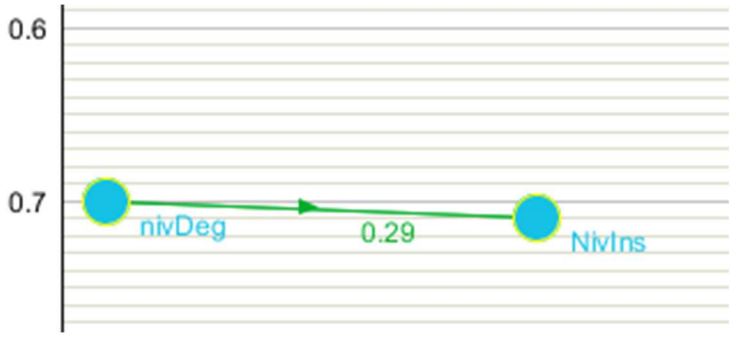


Figure 9 : Le niveau d'études se dégrade

Description des items
nivDeg : Niveau dégradé, P(nivDeg) = 0.70,
NivIns : Niveau insuffisant, P(NivIns) = 0.71

Les résultats nous montrent qu'un niveau d'études dégradé implique le niveau insuffisant pour les étudiants interrogés ($M^{f}_{GK}(\text{nivDeg} \longrightarrow \text{NivIns}) = 0.29$, $\text{OR} = 3.58$ et $\text{RR} = 1.53$) (figure 9).

Figure 10 : Résultat du croisement nivDeg × NivIns de l'Epi Info ; une dépendance positive et significative entre la modalité nivDeg et la modalité NivInsest observée

70% des étudiants interrogés ont répondu que leur niveau d'études se dégrade de plus en plus et 71% des étudiants ont aussi dit que leur niveau d'études est insuffisant. Par un manque de dispositions du net pour les étudiants et d'un côté une carence des enseignants, l'accès aux savoirs des étudiants était limité pendant le confinement.

Des cours en ligne n'ont pas été organisés ainsi que des conférences en ligne ne sont pas envisagées. D'ailleurs l'université d'Antsiranana ne dispose pas d'une infrastructure compatible à ces exigences engendrées par les nouvelles technologies. Certes le Wifi existe à l'ENSET et à la DAFF mais les autres composantes ne mettent pas à la disposition des étudiants ces réseaux faute d'installation *ad hoc*.

En conséquence, les étudiants risquent de ne plus avoir des moyens d'apprendre et ce phénomène engendra une chute de niveau des étudiants.

3.2.5 Analyse 5 : Et maintenant, quel est le ressenti des étudiants sur leur avenir universitaire?

Enfin, nous terminons notre analyse en revenant à la deuxième composante de notre résultat (figure 2) sur la partie de supérieure gauche (figure 11).

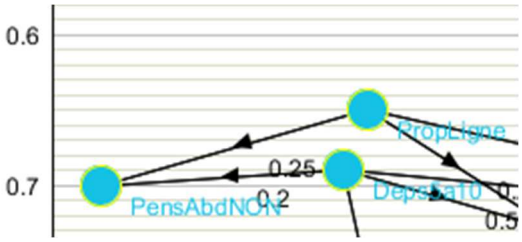


Figure 11 : Avenir universitaire des étudiants

Selon la figure ci-dessus, la relation entre PropLing et PensAbdNON est significative, (M^{f}_{GK}(PropLing ⟶ PensAbdNON) = 0.25, OR = 2.67 et RR = 1.37). La proposition de cours en ligne implique le choix de ne pas abandonner leur étude.

Avant même la pandémie de covid-19, les auteurs comme Doré, (2012) ou (Djistera), 2019 ont évoqué que les bibliothèques n'ont pas pu acquérir de nouveaux livres ni de nouveaux matériels, qu'il manque des salles de cours, consécutivement aux sureffectifs en étudiants. Inversement, les enseignants permanents ne sont pas en nombre proportionnel aux effectifs des étudiants et n'arrivent plus à suivre l'augmentation de la population étudiante, ce qui a conduit à une hausse spectaculaire du ratio étudiants par enseignant. De son côté Andriambelo (2014) signale l'absence de documentation, la rareté, le coût élevé des ouvrages, etc., Tout cet aspect est et demeure problématique pour les universités malgaches.

Puis lors de confinement, à cause des mesures d'urgence sanitaires dont l'interdiction des regroupements ou, les rassemblements de personnes y compris les étudiants, il s'en suit que l'accès à la documentation physique, aux recherches dans la bibliothèque et dans les laboratoires est impossible Les respects des gestes barrière comme la distanciation bloquent et mettent en désordre les stratégies classiques d'apprentissage chez les étudiants (figure 11).

Figure 11: Le ressenti des étudiants sur leur avenir universitaire.

Résultat du croisement PropLing × PensAbdNON de l'Epi Info ; une dépendance positive et significative entre la modalité PropLing et la modalité PensAbdNON

Il apparait les étudiants interrogés ne se sentent pas complètement dépourvus. Ils proposent une alternative aux dispositifs d'apprentissage classiques en utilisant les nouvelles Technologies d'Information et de la Communication comme entre autres les plus cités : les webconférences sur les réseaux sociaux, les centres de documentations numériques accessible gratuitement en ligne comme Wikipédia, les plateformes de travail collaboratif libre, etc. Même si la constatation de la A l'inverse de la crainte exprimée par

Banque mondiale en mai 2020 qui envisageait que la pandémie entraine une diminution de l'apprentissage va diminuer et une augmentation des abandons scolaires, notamment chez les plus défavorisés, la majorité des étudiants de l'Université d'Antsiranana n'ont pas du tout l'intention d'abandonner leurs études universitaires. Même si certains étudiants ont peur du retard engendré ou de l'allongement de l'année universitaire, ils se trouvent en définitive résilients.

4 Conclusions et perspectives

Cet article étudie dans les vécus des étudiants confinés de l'Université d'Antsiranana les conséquences des changements liées aux transformations brutales consécutives à la propagation du covid-19. La virulence de ce virus a obligé le gouvernement malgache à prendre des mesures drastiques pour terminer l'année universitaire. En particulier, il a interdit les enseignements en présentiel et à l'opposé il a souhaité développer l'enseignement à distance. Nous avons examiné à travers une enquête les manières des étudiants pour que perdurent leurs enseignements et leurs manières d'apprentissage. Nous nous sommes intéressés également à leur ressenti sur leur formation qui se manifeste, pour eux, par une dégradation de leur niveau d'études. Nous avons remarqué aussi que si la motivation de poursuivre des études universitaires se manifeste dans l'immense majorité des étudiants ils se sentent frustrés et découragés par le manque flagrant de moyens disponibles. Le ministère concerné rencontre toujours des difficultés à assurer pleinement ses missions envers toutes les universités de Madagascar. En dépit des budgets alloués aux TIC, juste une partie d'étudiants ont un accès informatique correct en adéquation avec leurs activités d'enseignement et de recherche. Evoquée tout le temps, la formation à distance en s'appuyant sur les TIC n'est toujours pas une réalité dans la plupart des établissements à Madagascar en particulier à l'université d'Antsiranana. La pandémie met, hélas, en exergue les faiblesses dans la l'accès au système numérique. Les dépenses chez les étudiants ont augmenté et ils se sentent perturbés au niveau de leur pédagogique et ressentent également une baisse de leur niveau d'études. Pour éviter des dégâts, il faut chercher à tout prix la promotion de TIC, avec une méthode simple et adaptée aux situations de l'ensemble des étudiants et des enseignants ainsi que des personnels administratifs des universités. Les TIC ont des avantages dans l'enseignement si les conditions favorables d'accès et d'utilisation sont réunies. A l'issue de cette analyse de données, nous avons reconfirmé l'intérêt de l'Analyse Statistique Implicative comme (ASI-MGK, CHIC, ...) par rapport aux mesures statistiques basées sur le risque relatif et de l'odds ratio. Elle est également pertinente pour explorer les relations valides à partir de ses valeurs critiques. Elle dispose aussi de la capacité d'identifier la cause et la conséquence dans une relation. Elle s'avère donc être la méthode la plus efficace pour une analyse causale. Enfin ces informations nous seront très utiles lors de l'étude prospective à effectuer sur l'élaboration et la mise en place de plateformes d'apprentissage collaboratif numérique.

Références

[1] Andriambelo M. (2014). *L'introduction des TIC dans l'enseignement supérieur : cas de la faculté DEGS*. Mémoire. Université d'Antananarivo. Madagascar.

[2] Andriamampianina, L. (2012). *Etude en vue de la mise en place du fonds compétitif et d'innovation pour l'amélioration de l'enseignement supérieur et de la recherche scientifique.* Collecte de documents et de données et élaboration des manuels de procédures. 3ème rapport d'activités. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[3] ADEA- Abidjan, (2020). *Assurer l'éducation à domicile dans les Etats membres africains dans le contexte de la pandémie du covid-19 :* Rapport sur la situation dans les pays- ADEA- Abidjan).

[4] Bashir, S. (2008). *Mieux former la population active pour préparer l'avenir. La transformation de l'enseignement post-fondamental à Madagascar*, Document de travail de la Banque mondiale, N° 157.

[5] Banque mondiale, (2020). *Covid-19 et l'éducation : les mesures à prendre pour faire face aux chocs.* 7 mai 2020

[6] Banque mondiale Mai (2020). *Pandémie de Covid-19 : chocs pour l'éducation et réponses stratégiques. Groupe de la Banque mondiale, éducation.*

[7] Djistera, A. A. (2019). *Évolution et enjeux de l'enseignement supérieur à Madagascar*, Vol. 4, **Annales UMCS section** Educationova, pp. 21-37.

[8] Doré, G. (2012). *Etude pour l'amélioration des services offerts aux étudiants de l'enseignement supérieur malgache*. Rapport final. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[9] Duret, Elsa. (2005). Atteindre les objectifs d'Éducation Pour Tous en Afrique subsaharienne : état des lieux, marges de manœuvre et défis pour les politiques publiques, *Mondes en développement*, no 132, p. 57-84

[10] Etienne, S, Rakotoson, Y, (2012). *TIC pour apprendre, TIC pour enseigner, colloque international sur les TIC*, Ecole Normale

[11] Gioan, P.A. (2012). « *Etude en vue de la mise en place du fonds compétitif et d'innovation pour l'amélioration de l'enseignement supérieur et de la recherche scientifique* ».Rapport final. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[12] Instat, (2020). *Impact du covid-19 sur les conditions de vie des ménages, enquête à haute fréquence par téléphone auprès des ménages (EHTM),* 13 juin 2020, Madagascar.

[13] Institut de statistique de l'UNESCO (ISU) (2020), Disponible sur : http://uis.unesco.org/fr

[14] Nations unies - Conseil économique et social, 27 février 2014. *Les technologies de l'information et de la communication pour un développement économique et social équitable.*

[15] Ndiaye, C. F., (2020). *Appel d'Urgence pour Covid 19- Madagascar.* juin-août 2020. Nations-Unies.

[16] Raharimihaja Zakariasy L., Leaby N., Rafanomezantsoa L. A., Rakotomalala, M., Razafindradina, B.et Razanakolona J-M., (2014). *Projet Université Numérique Malgache (UNuM).* Rapport final – draft. Mai 2014.

[17] Rakoto, D.B. (2012). *« Collecte de documents et de données et élaboration de manuels de procédures dans le cadre de l'étude pour l'amélioration des services aux étudiants ».* Rapport final. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[18] Ralahady B. B. and Totohasina A. (2019a) *Experimental study of the valid rules according to the measure* M_{GK} in IJCST vol.3 N°.1.

[19] Ralahady B. B. and Totohasina A. (2019b) *ASI-MGK: Implicative Statistical Analysis tool based on* M_{GK} in IJCST vol.3 N°.1

[20] Ralahady, B. B., Totohasina, A. et Simon, J. (2019c). *Apports intéressants de l'indice asi-mgk et des courbes d'apprentissages en didactique de l'informatique*. Analyse Statistique Implicative, 10:304.

[21] Randriamanambintsoa, M. (2012). *« Rapport sur l'exécution de l'étude sur l'amélioration des données statistiques de l'éducation post-fondamentale ».*Washington, DC, World Bank.

[22] Razafimbelo, J. (2012). *« Etat des lieux de l'enseignement supérieur à distance (public et privé) ».* Rapport final. Antananarivo. Ministère de l'Enseignement supérieur et de la recherche scientifique.

[23] Razafimbelo A. N., (2019). *Recherche de l'efficacité de l'enseignement supérieur professionnel à Madagascar par l'employabilité des sortants : cas de l'IST-T.* Universitéd'Antananarivo.

[24] Repobolikan'i Madagascar. (2013). *La stratégie nationale de la recherche et scientifique à Madagascar.* Ministère de l'Enseignement supérieur et de la Recherche Scientifique. Madagascar.

[25] Repobolikan'i Madagascar. (2020). *Plan Multisectoriel D'urgence Madagascar*, 1er juillet 2020.

[26] Guillaume, S., Khenchaf, A., & Briand, H. (2000). *Generalizing Association Rules to Ordinal Rules*. In IQ (pp. 268-282).

[27] Totohasina, A. (2008). *Contribution to the study of measures of quality of association rules: normalization and constraints in five cases and MGK, properties, composite base and extension rules for applying statistical and physical sciences*, University of Antsiranana, Madagascar, HDR

Annexe : Questionnaire

1. Numéro du questionnaire ……. Date …….
2. Sexe ? ☐ Masculin ☐ Féminin
3. Quel âge avez-vous ? ……..
4. Dans quel quartier résidez-vous au campus universitaire ?
   ☐Quartier Bloc☐Quartier PV ☐Quartier PJ☐Quartier Belle rose
5. A cause du Covid-19, comment trouvez-vous votre étude en général ?
   Niveau ☐Moyenne ☐ Insuffisant☐Médiocre
   Progression ☐améliorant☐Constant☐Dégradé
6. Lors de notre Président République a annoncé la fermeture de tous les établissements, qu'avez-vous faits?
7. Pouvez-vous continuer et améliorer votre connaissance dans ce confinement ?
   ☐Oui ☐Non
8. Par quel moyen vous-y -arrivez ?
   ☐Travail Individuel ☐ Facebook ☐Internet ☐Amis
9. Est-ce que vous trouvez des changements de stratégies de la représentation de ce confinement ou bien c'est la même de votre habitude ?
   ☐ Dégrade ☐ Constante ☐ Augmente
10. Avez-vous des occasions de faire des études ou des devoirs ensembles avec vos camarades de classe lors de la pandémie du covid-19 ?
    ☐ Oui ☐ Non
11. Par quel moyen avez-vous fait ces études ou devoirs ensemble pendant l'interdiction de se regrouper lors du confinement ?
12. Où trouvez-vous tous les documents pour vos études ?.
13. Connectiez-vous sur internet et par quel moyen ?
    ☐ Oui ☐ Non ☐ Occasionnelle
    ☐ CYBER ☐ Internet Mobil (Orange /Telma/Airtel) ☐WIFI ENSET ☐WIFI DAFF
14. Dans une semaine combien de fois pouviez-vous vous connecter sur l'Internet ?
15. Est-ce qu'avez-vous occasions de faire des cours ou des études ou des conférences en ligne avec vos camarades de classe lors du confinement ?
    ☐ Oui ☐ Non
16. A votre avis ? ☐Facile ☐Moyenne ☐Difficile
17. Est-ce que vous avez participé de telle sorte des cours en ligne ?☐ Oui ☐ Non
18. A votre avis, quelle différence si vous comparez avec votre cours en classe habituelle ? ☐Très diffèrent ☐Applicable
19. Pendant ce confinement quelle sorte de connaissances avez-vous reçu à part celles de votre établissement ?
20. Après le bouleversement d'année scolaire, est-ce que vous déjà pensez d'abandonner votre étude ? Oui/ Non/ Abandonné

ESTUDIO SOBRE VIOLENCIA EN RELACIONES SEXO-AFECTIVAS DE JÓVENES DE LA PROVINCIA DE NEUQUÉN-ARGENTINA MEDIANTE A.S.I.

Pablo Carranza[1], Federico Vasches[2], Ailin Gamoneda[3], María Cardieri[4], Iván Benzaquén[5], Erika Pedersen[6]

ÉTUDE SUR LA VIOLENCE DANS LES RELATIONS SEX-AFFECTIVES DES JEUNES DE LA PROVINCE DE NEUQUÉN-ARGENTINE PAR A.S.I.

STUDY OF VIOLENCE IN SEX-AFFECTIVE RELATIONSHIPS OF YOUNG PEOPLE FROM THE PROVINCE OF NEUQUÉN- ARGENTINA THROUGH A.S.I.

RESUMEN

Las violencias contra las mujeres constituyen una vulneración de los Derechos Humanos, siendo una problemática social compleja, amplia, multifacética y presente en todas las clases sociales, franjas etarias y niveles socioeducativos sin distinción. Frente a esta problemática, el Observatorio de la Violencia Contra las Mujeres de la provincia del Neuquén y el Observatorio de Políticas Públicas y Sociales de la provincia de Río Negro, ambos de la Patagonia Argentina, se propusieron realizar un trabajo que pudiera generar algún tipo de aporte. En base a los datos recolectados por una encuesta con variables mayoritariamente modales, y por medio del análisis implicativo se extrajo información sobre fenómenos relacionados a violencia de género en jóvenes neuquinos. El estudio permitió construir conocimiento no solo sobre la violencia de género en la juventud sino también identificar individuos en situación de riesgo, posibilitando así brindar herramientas de intervención a equipos especializados.

Palabras claves: *violencias, mujeres, información, prevención, intervención*

RÉSUMÉ

La violence à l'égard des femmes constitue une violation des droits de l'homme, étant un problème social complexe, vaste et multiforme qui est présent dans toutes les classes sociales, tranches d'âge et niveaux socio-éducatifs sans distinction. Face à ce problème, l'Observatoire de la violence à l'égard des femmes de la province de Neuquén et l'Observatoire des politiques publiques et sociales de la province de Río Negro, tous deux en Patagonie Argentine, ont entrepris de réaliser un travail susceptible de générer une sorte de contribution. Sur la base des données collectées par une enquête avec des variables principalement modales, et à travers l'analyse implicative, des informations ont été extraites sur les phénomènes liés à la violence de genre chez les jeunes de Neuquén. L'étude a permis de construire des connaissances non seulement sur la violence de genre chez les jeunes mais aussi d'identifier les individus à risque, fournissant ainsi des outils d'intervention aux équipes spécialisées.

Mots-clés: *violence, femmes, information, prévention, intervention*

1Universidad Nacional de Río Negro. Observatorio de Políticas Públicas y Sociales de Río Negro **pfcarranza@gmail.com**

2Universidad Nacional de Río Negro. Observatorio de Políticas Públicas y Sociales de Río Negro **observatoriodepoliticasrn@gmail.com**

3Observatorio de la Violencia Contra las Mujeres **observatorioviolencia@outlook.com**

4Observatorio de la Violencia Contra las Mujeres

5Observatorio de la Violencia Contra las Mujeres

6Observatorio de la Violencia Contra las Mujeres

ABSTRACT

Violence against women constitutes a violation of Human Rights, being a complex, broad, multifaceted social problem that is present in all social classes, age groups and socio-educational levels without distinction. Faced with this problem, the Observatory of Violence Against Women of the province of Neuquén and the Observatory of Public and Social Policies of the province of Río Negro, both in Patagonia Argentina, set out to carry out work that could generate some kind of contribution. Based on the data collected by a survey with mostly modal variables, and through the implicative analysis, information was extracted on phenomena related to gender violence in young people from Neuquén. The study made it possible to build knowledge not only about gender violence in youth but also to identify individuals at risk, thus providing intervention tools to specialized teams.

Keywords: *violence, women, information, prevention, intervention*

1 Introducción

En esta comunicación compartiremos un trabajo realizado entre el Observatorio de la Violencia Contra las Mujeres (OVCM) de la provincia del Neuquén y el Observatorio de Políticas Públicas y Sociales (OPPSRN) de la provincia de Río Negro, ambos de la Patagonia Argentina. Nos abocamos, de manera general, al estudio de la violencia de género, y en esta oportunidad a la violencia en las relaciones sexo-afectivas de los jóvenes que habitan la provincia de Neuquén.

El OPPSRN fue creado por un grupo interdisciplinario de profesionales, que trabaja por ejes temáticos, desde fines del año 2019. Entre sus principales objetivos se encuentra el de acompañar, fortalecer y mejorar las administraciones y las políticas públicas en los territorios.

El OVCM depende de la Subsecretaría de las Mujeres, Ministerio de Ciudadanía, gobierno de la provincia del Neuquén, creado por la ley provincial 2887/13. El Observatorio tiene como propósito *"desarrollar un sistema de información permanente que brinde insumos para el diseño, implementación y monitoreo de políticas públicas, tendientes a la prevención y erradicación de la violencia contra las mujeres"* (Art. 2).

Las violencias contra las mujeres constituyen una vulneración de los Derechos Humanos. Son una problemática social compleja, amplia, multifacética y presente en todas las clases sociales, franjas etarias y niveles socioeducativos sin distinción. Se encuentran ligadas a normas socioculturales profundamente arraigadas, cuyos efectos y consecuencias afectan a la sociedad en su conjunto, lo cual dificulta su comprensión, detección y abordaje. En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró la pandemia de COVID-19. Esto conllevó a medidas de aislamiento y distanciamiento social, generando un nuevo contexto que implica una transformación sustancial en la vida cotidiana de las personas, que no sólo tiene y tendrá profundas consecuencias económicas sino también sociales y psicológicas. La Organización de las Naciones Unidas (ONU) afirma que el mayor impacto social de la pandemia recae sobre las mujeres y se destaca una tendencia de crecimiento de la violencia hacia las mujeres en el ámbito doméstico. Las estrategias de autoaislamiento y cuarentena en hogares donde ya existían situaciones de violencia, incrementaron la tensión y el riesgo.

La provincia de Neuquén no es ajena a los sucesos ocurridos a nivel mundial. En el año 2020, la provincia de Neuquén tuvo una población estimada, de 664.057 habitantes

de los cuales 251.674 son mujeres de 15 años y más(Censos, 2020). Según el Informe anual de la violencia contra las mujeres(Mujeres, 2020) realizado por el OVCM:

- 8.260 mujeres de 15 años y más, solicitaron asistencia por situaciones de violencia en una área gubernamental fuente de información
- 14.815 intervenciones de asistencia por situaciones de violencia
- 8.197 denuncias policiales

Frente a esta problemática, los observatorios consideraron la viabilidad de llevar adelante diversos proyectos, optando como primera experiencia de trabajo colaborativo el análisis de las encuestas realizadas por el Programa "Pintó Enamorarse".

"Pintó Enamorarse" es un programa de capacitación y sensibilización para prevenir la violencia en las relaciones sexo-afectivas de jóvenes, como así también, la disminución de futuras situaciones de maltrato o violencia en las relaciones de pareja en la adultez. El mismo comenzó en el año 2013 y actualmente lo lleva adelante la Dirección de Equidad de Género de la Subsecretaría de las Mujeres, Ministerio de Ciudadanía, Gobierno de la provincia de Neuquén (DEGN).

Entre las actividades desarrolladas en el marco de este programa y durante los años 2018 y 2019, un equipo técnico integrante del mismo, diseñó, ejecutó y sistematizó una encuesta a jóvenes escolarizados de nivel secundario que participaron de las jornadas de capacitación y sensibilización en distintas localidades de la provincia de Neuquén.

En base a los datos recolectados por las encuestas, la DEGN confeccionó un informe donde, con herramientas de estadística descriptiva univariada se analizaron las tendencias observadas. Esta base de datos no resultó de fácil abordaje desde la estadística tradicional, entre otras cuestiones por el tipo de variables consideradas (modales y no numéricas en general). Sin embargo, ambos observatorios consideraron que subyacía a la misma, información de potencial interés y que no estaba pudiendo ser extraída mediante los métodos estadísticos descriptivos univariados utilizados por la DEGN hasta el momento.

2 Fundamentación

El trabajo en conjunto entre los dos observatorios se orientó entonces a intentar extraer información adicional de la encuesta "Pintó Enamorarse" elaborada por la DEGN. Tomando en cuenta los resultados obtenidos en ese estudio se observó que el fenómeno de violencia entre los jóvenes y en particular el de género no era mayoritario. Por ejemplo, ante la pregunta realizada: "Cuando se enojan o discuten, ¿sentís que tu vida está en peligro?". La distribución de frecuencias absolutas fue: nunca: 646, alguna vez: 49 y frecuentemente: 6.

Las variables con las que se confeccionó la encuesta eran, principalmente, del tipo modal. En efecto, la mayoría asumía valores del tipo "nunca", "a veces", "frecuentemente"; raras resultaban ser numéricas (Edad por ejemplo). Recordamos que la misma había sido ya elaborada por la DEGN, por lo que tanto su diseño como los datos recolectados pre-existían a la intención del trabajo colaborativo entre los dos observatorios.

En el análisis de los datos que aquí utilizamos, los observatorios pretendían no solo mejorar la comprensión del fenómeno de la violencia en general sino también identificar

a las personas (individuos) en situación de riesgo. Interesaba entonces tanto una mirada macro como también una micro. Ambas miradas fueron consideradas entonces como complementarias. En efecto, naturalmente lo micro (el individuo) alimenta lo macro (la tendencia). Pero, a su vez, la identificación de ciertas tendencias resultaba de interés para "regresar" al individuo, y en el caso de la problemática en cuestión, identificar jóvenes en situación de violencia.

Estos factores llevaron a considerar como método de análisis de los datos al Análisis Estadístico Implicativo (Gras & Bailleul, 2000; Gras, Couturier, Guillet, & Spagnolo, 2005). En principio, el interés por regresar a lo micro (identificación de los individuos en riesgo) se abordaría con la tipicalidad de los individuos que responden a una regla o un conjunto de reglas. Además, el método ASI puede proponer reglas incluso cuando la frecuencia es baja; y por último, el método no presenta problemas para tratar variables modales como las de la encuesta a disposición.

3 Objetivo

Realizar un estudio exploratorio mediante la metodología ASI (Gras et al., 1996) sobre la base de datos elaborada a partir de la encuesta del programa "Pintó Enamorarse" de los años 2018 y 2019.

Más precisamente, el trabajo se orienta a:

- Reconocer relaciones entre los fenómenos vinculados a la violencia entre jóvenes
- Identificar individuos en riesgo.

4 Marco Teórico

La violencia es una problemática social profundamente compleja que no puede ser interpretada ni abordada en su total magnitud, si no se aplica una perspectiva de género que ponga el foco en la forma en que varones y mujeres y otras identidades se relacionan y cuáles son los estereotipos, los roles y los mandatos naturalizados, que perpetúan las desigualdades afectando principalmente a las mujeres.

Según las leyes vigentes en Argentina y en particular en la provincia del Neuquén, la violencia contra las mujeres se define como toda conducta, acción u omisión, que de manera directa o indirecta, tanto en el ámbito público como en el privado, basada en una relación desigual de poder, afecte su vida, libertad, dignidad, integridad física, psicológica, sexual, económica o patrimonial, como así también su seguridad personal, incluyendo a aquellas perpetradas desde el Estado o por sus agentes (Argentina, 2009)

Cabe mencionar que cuando se habla de "mujeres" se hace referencia a aquellas personas que sienten subjetivamente su identidad o expresión de género mujer, de acuerdo o no al sexo asignado al momento del nacimiento, tomando en cuenta su vivencia interna e individual, como así también la corporal, involucrando o no la modificación de la apariencia o de su cuerpo a través de medios farmacológicos, quirúrgicos o de otra índole, cualquiera sea su orientación sexual, siempre que ello sea escogido libremente. (Neuquén, 2015).

La violencia constituye una relación social basada en la fuerza, el control y la dominación de un miembro de la pareja sobre el/la otro/a y puede ser ejercida de diferentes maneras y asumir diversas modalidades.

Los tipos de violencia que fueron considerados para este estudio son:

- la física: entendida como aquella violencia que se emplea contra el cuerpo de la mujer produciendo dolor, daño o riesgo de producirlo y cualquier otra forma de maltrato o agresión que afecte su integridad física.
- la psicológica: la cual causa daño emocional y disminución de la autoestima o perjudica y perturba el pleno desarrollo personal, que busca degradar o controlar sus acciones, comportamientos, creencias y decisiones, mediante amenaza, acoso, hostigamiento, restricción, humillación, deshonra, descrédito, manipulación o aislamiento. Incluye también la culpabilización, la vigilancia constante, la exigencia de obediencia y sumisión, coerción verbal, persecución, insulto, indiferencia, abandono, celos excesivos, chantaje, ridiculización, explotación y limitación del derecho de circulación o cualquier otro medio que cause un daño a su salud psicológica y a su autodeterminación.
- la sexual: caracterizada como cualquier acción que implique la vulneración en todas sus formas, con o sin acceso genital, del derecho de la mujer de decidir voluntariamente acerca de su vida sexual o reproductiva a través de amenazas, coerción, uso de la fuerza o intimidación, incluyendo la violación dentro del matrimonio o de otras relaciones vinculares o de parentesco, exista o no convivencia, así como la prostitución forzada, la explotación, la esclavitud, el acoso, el abuso sexual y la trata de mujeres. (Argentina, 2009).

El presente estudio aborda el tema de la violencia en las relaciones de noviazgo de jóvenes de 12 a 23 años. Desde una perspectiva de los Derechos Humanos podemos entender a la juventud como una etapa de la vida de las personas marcada por el crecimiento, el desarrollo de la propia subjetividad y con necesidades específicas. De ningún modo se lo puede asociar a una etapa transitoria de formación para la vida adulta, sino por el contrario debe ser considerada como un momento de la vida en sí mismo, donde se desarrollan vínculos propios, espacios de socialización e interacción con otros/as (escuela, club deportivo, la calle, el barrio, el boliche, etc.). Así se profundizan las relaciones con sus pares, la amistad y en muchas ocasiones las primeras relaciones sexo-afectivas o de noviazgo(Arévalo, Romero, & Ferreyra, 2018)

En este sentido, la violencia en el noviazgo de jóvenes debe ser abordada y entendida como una problemática social que se encuentra atravesada por las particularidades de esta etapa de la vida, enmarcada en un contexto histórico y en una determinada cultura y no como un fenómeno social que responde únicamente a aspectos biológicos e individuales.

La violencia en las relaciones de pareja entre jóvenes presenta características que la diferencian de la violencia en el matrimonio o entre adultos/as, debido a:

- la edad de los miembros involucrados es relativamente menor a las de las parejas convivientes o casadas, ya que se trata de adolescentes o jóvenes en temprana edad.
- las razones por las cuales se presenta la violencia en estas relaciones, pueden ser distintas a las de la violencia entre convivientes, ya que no hay

> obligaciones filiales o carga familiar, no hay obligaciones económicas de ninguna naturaleza, ni contractuales, ni de dependencia, ni responsabilidad de auto-sostén o del otro miembro. Las parejas de novios/as se encuentran o se citan para salir, para conocerse, para entretenerse, divertirse juntos, compartir etc. pero no conviven. Ambos miembros tienen su propio entorno familiar que es el ámbito de contención emocional y de sustento. Incluso en muchos casos, aún no tienen proyectos en común, al menos a largo plazo (Arévalo et al., 2018).

De este modo, es posible distinguir aspectos propios de la violencia de género contra las mujeres que se reproducen a lo largo de todas las etapas donde las personas establecen vínculos sexo-afectivos y otros que aparecen en la etapa de la juventud de manera singular.

A grandes rasgos, la violencia es un proceso que generalmente se presenta de forma cíclica, lo cual da cuenta de una secuencia repetitiva característica de este tipo de relaciones, no como eventos aislados sino como situaciones que adquieren cierta cronicidad.

El Ciclo de la violencia puede desarrollarse en tres fases, cuya duración y características dependerá de cada situación. Debe tenerse en cuenta que el tiempo entre cada fase tiende a acortarse a medida que se prolonga la relación violenta, y que en cada nueva fase o episodio se irá acrecentando la intensidad. A medida que el ciclo se repite, el proceso violento se agrava, dinámica que se da de forma paulatina. Esto produce un estado de permanente tensión, temor y parálisis; motivos por los cuales se acrecienta el aislamiento de la víctima y la dificultad para pedir ayuda. A continuación, se presentan las fases de este ciclo:

- Acumulación de tensión: En esta fase la violencia es sutil, por lo general en forma psicológica; como chistes o comentarios que descalifican. Se produce una serie de incidentes que van incrementándose en ansiedad y hostilidad. La persona sometida tratará de calmar al agresor, justificará el incidente culpando a causas externas o a sí misma, experimentando sentimientos de culpa y vergüenza que le impiden socializar su problemática con su entorno.
- Estallido de violencia: La violencia acumulada estalla en un episodio que rompe el patrón con el que venía sucediéndose. Así, por ejemplo, si la violencia era ante todo psicológica, ahora se manifiesta en forma física. La gravedad o estallido que caracteriza a esta segunda fase suele aumentar con el paso del tiempo. De las primeras cachetadas, empujones y pellizcos se va pasando a los golpes de puño, patadas y hasta al uso de armas. Cabe aclarar que no hay nada que pueda hacer la persona que padece la violencia para evitar que el episodio agudo estalle, es una situación impredecible y puede ser generada por cualquier motivo, hasta el más trivial. La exposición prolongada a este tipo de maltrato y de violencia produce en la persona un deterioro de sí misma que la hace dudar de sus propias percepciones.
- Luna de miel: En esta fase, quien agrede muestra arrepentimiento, pide perdón y jura no volver a hacerlo, sobreviene un período de seducción y la promesa que nunca más volverá a ocurrir. En esos momentos ambos creen que ya superaron los problemas de violencia. Luego el ciclo inicia nuevamente.

Cabe mencionar, que es posible detectar tempranamente rasgos del ciclo de la violencia en los vínculos sexo-afectivos de los y las jóvenes, tal como se verá en los resultados de este estudio.

5 Metodología

El presente trabajo consiste en una aproximación exploratoria que busca obtener una caracterización más exhaustiva de las situaciones de violencia en la etapa de la juventud, en consonancia con los objetivos planteados originalmente en la encuesta realizada por el equipo técnico del programa "Pintó enamorarse".

En este caso, y como ocurre en muchas situaciones de minería de datos, estos últimos preceden en existencia al planteamiento de los objetivos del estudio. Se trata entonces de una base de datos que reúne las respuestas de 853 estudiantes de nivel secundario de la provincia de Neuquén, que fue puesta a disposición para la realización de este estudio.

La encuesta fue realizada en formato papel durante los años 2018 y 2019 y contaba con un total de 26 preguntas. La mayoría, del tipo modal, donde las opciones de respuesta (nunca, a veces, frecuentemente) se refieren a la intensidad de apreciación respecto a un hecho de violencia o de indicio de violencia padecido. El anexo I reúne las preguntas propuestas en el cuestionario y sus posibles respuestas.

Cabe acotar que se observó un número importante de datos faltantes que tenían como origen la no respuesta de los jóvenes a una determinada pregunta. En este caso, la falta de respuesta puede interpretarse como un silencio por parte de los y las jóvenes y dado el contexto de la problemática, ese silencio podría representar una respuesta de interés oculta.

Lamentablemente la estructura del cuestionario no permitió proponer relaciones significativas con ese tipo de respuestas vacías. Sin embargo, ellas fueron retenidas en el estudio y conservadas como variables secundarias en el análisis estadístico implicativo.

Las variables modales cuyas respuestas admiten niveles de intensidad en la apreciación de los encuestados (nunca, a veces, frecuentemente, etc.) fueron consideradas como binarias y esto a los efectos de facilitar la interpretación. En efecto, reviste interés que la semántica de las reglas obtenidas resultase de fácil acceso a interlocutores no familiarizados con este tipo de estudios, en particular, el personal de organismos gubernamentales a los que les podría representar de utilidad el conocimiento extraído a los fines de intervención en el terreno. La conversión de variables modales a variables binarias extendió el número de variables hasta un total de 130 variables.

Dada la cantidad de individuos que respondieron la encuesta, se retuvo la implicación del tipo entrópica y la ley de Poisson, aunque claro está, se exploraron también las reglas obtenidas bajo la teoría clásica con la ley binomial. En este último caso, la cantidad de reglas obtenidas perturbaban la búsqueda de reglas significativas en función de los objetivos del trabajo, por lo que se retuvo la primera opción mencionada.

Las respuestas vacías fueron agrupadas en variables codificadas según su variable de origen. Por ejemplo, la variable binaria "1vac" asume valor 1 para individuos que no respondieron a la pregunta 1 y 0 para aquellos que la respondieron. Lo mismo ocurre con

la variable binaria “2vac”, etc. Las variables binarias vacías fueron asignadas en modo variable suplementaria.

6 Resultados

Bajo la teoría entrópica (Gras & Régnier, 2009) y utilizando la ley de Poisson, las primeras reglas aparecen con un nivel de significación de 0.98 (no presentadas aquí por razones de espacio). Si el nivel de significación se lo desciende a 0.95, emergen otras reglas de interés. Por ejemplo, el de la Figura 1 (se detalla solamente el cardinal de los antecedentes retenidos):

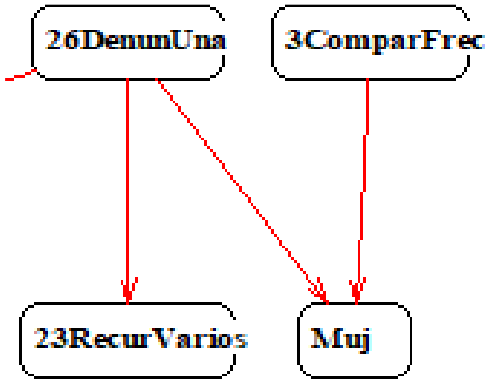


Figura 1. Grupo Mujeres. Nivel de significación: 0.95 Fuente: Elaboración propia

Este entramado de reglas reúne las siguientes variables. En los antecedentes:

26DenunaUna:
Pregunta26: Hiciste en el último año denuncias en la policía por violencia contra tu pareja, novia, novio o ex:
Respuesta: Una vez
Cantidad de individuos:10

3ComparFrec:

Pregunta3: ¿Te compara con sus ex o sale con otras personas?

Respuesta: Frecuentemente

Cantidad de individuos: 21

En los consecuentes:

23RecurVarios:

Pregunta23: Si sufriste violencia en una relación de noviazgo en el último año recurriste a:

Respuesta: Varios Organismos

Muj:

Pregunta: Género?

Respuesta: Mujer

Este grupo de reglas, siempre con un umbral de implicación de 0.95, informa sobre las jóvenes encuestadas: son mujeres las personas que resultan ser frecuentemente comparadas con las ex de su pareja. De la misma manera, son mujeres quienes, ante un acto de violencia, han denunciado una vez a sus parejas, cabe acotar que los varones recurren a sus amigos (regla no visualizada aquí).

Otro entramado de reglas resulta evidente, al mismo nivel de significación (0.95), en este caso vinculado al género masculino.

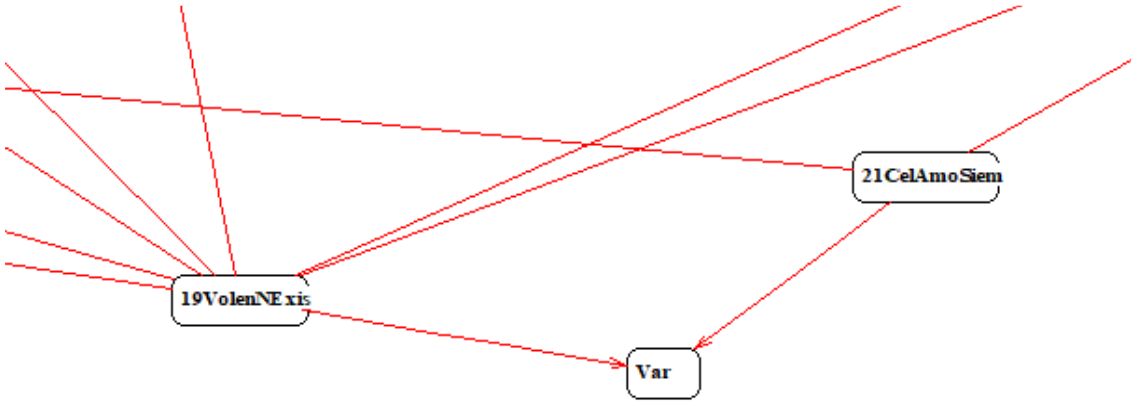


Figura 2. Grupo Varones. Nivel de significación: 0.95Fuente: Elaboración propia

Este entramado de reglas reúne las siguientes variables. En los antecedentes:

21CelAmoSiem:

Pregunta21: ¿Vos pensás que los celos son una demostración de amor?

Respuesta: Siempre

Cantidad de individuos: 17

19VolenNExis:

Pregunta19: ¿Qué opinas? La violencia en las relaciones de pareja entre adolescentes o jóvenes es:

Respuesta: No existe

Cantidad de individuos: 7

En el consecuente:

Var:

Pregunta: Género?

Respuesta: Varón

Este entramado de reglas, a un nivel de significación de 0.95, caracteriza a un grupo de interés en relación a la violencia de género: Son varones quienes piensan que la violencia en las parejas adolescentes no existe, al mismo tiempo que quienes piensan que los celos son un gesto de amor siempre, resultan ser varones.

Los entramados de reglas presentados anteriormente son sólo un ejemplo de la información que pudo ser extraída a un nivel de significación de 0.95. Hubo muchas otras reglas que resultaron de gran interés a los integrantes de los observatorios y que no representamos aquí por razones de espacio.

A continuación presentamos otros conjuntos de reglas, esta vez a partir de un conjunto de variables de interés. Para ello, se recurrió al modo Cono del software CHIC. La Figura 5 muestra las reglas alrededor de las variables de interés a un nivel de significación de 0.90. Estas variables son:

- 12. Cuando se enojan o discuten ¿sentís que tu vida está en peligro?

- 13. ¿Te ha golpeado, empujado, pellizcado o tirado de los pelos? ¿Dado una bofetada?
- Género.

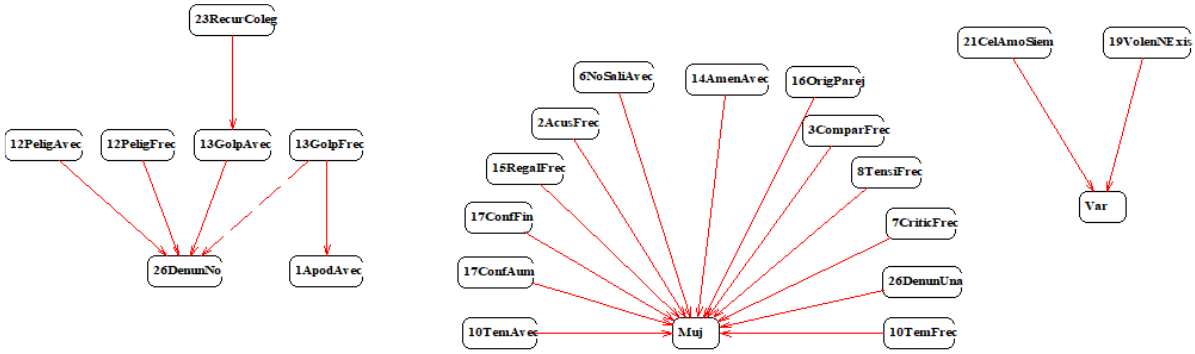


Figura 3. Mujeres y Varones. Diferencias en fenómenos observados. Nivel de significación: 0.90
Fuente: Elaboración propia

A los efectos de facilitar la interpretación, en la Figura 3 se sintetiza la semántica de algunas de las reglas donde las mujeres y varones son consecuentes.

Regla 14AmenFrec → 15RegalFrec →Muj

Son mujeres quienes reciben amenazas de muerte con frecuencia cuando se enoja la pareja. Son mujeres también, a quienes con frecuencia, se le obsequia regalos o se le promete cambios luego de una discusión fuerte con su pareja.

10TemAvec →Muj ←10TemFrec

También son mujeres quienes han cedido a deseos sexuales de la pareja por temor o presión.

12Pelig (Avec y Frec) →DenunNo→Muj

Son mujeres también, quienes habiéndose sentido en peligro, no han hecho denuncias, al menos el último año.

14AmenFrec y 13GolpFrec →DenunNo→Muj

Mujeres también son quienes han recibido amenazas con frecuencia o las han golpeado, pellizcado o tirado de los pelos con frecuencia, y no han hecho una denuncia el último año.

Son mujeres también quienes son acusadas con frecuencia de ser infieles o de coquetear con otras personas (2AcusFrec). Son mujeres también quienes son criticadas con frecuencia por su cuerpo, en público o en privado (7CriticFrec).

Respecto a los varones, se mantienen las reglas analizadas en el punto anterior.

El siguiente entramado de reglas es obtenido aplicando el modo Cono (ver Figura 4) sobre dos variables de interés (nivel de significación 0.90):

- 15 - Después de una discusión fuerte, se muestra cariñoso/a y atento/a, ¿te regala cosas y te promete que nunca más volverá a suceder y que “todo cambiará”?
- 10 - Si has cedido a sus deseos sexuales. ¿Sientes que ha sido por temor o presión?

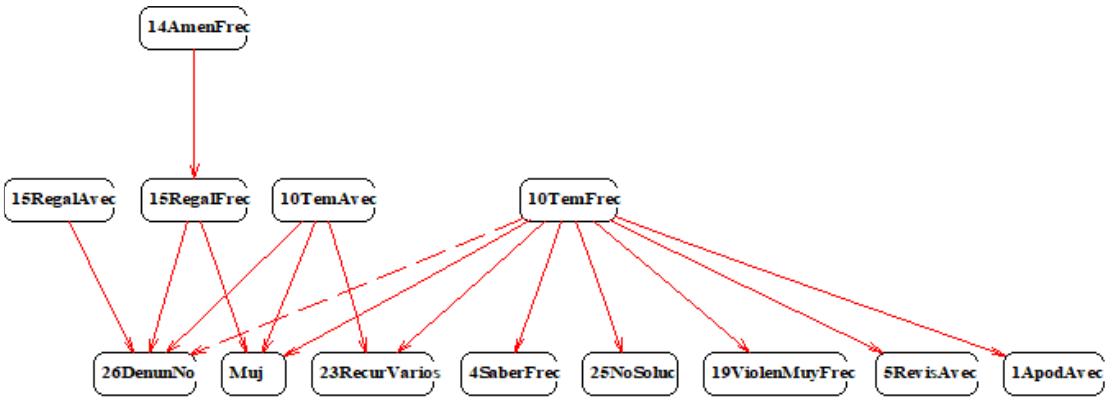


Figura 4. Título: Ciclo de violencia y violencia sexual. Nivel de significación 0.90. Fuente: Elaboración Propia

Como se aprecia en la Figura 4, las amenazas de muerte frecuentes ante una discusión seguidas de regalos y promesas de cambio frecuentes (Ciclo de la violencia - Fase Luna de miel) le sucede a las mujeres. Del mismo modo, quienes han cedido a sus deseos sexuales por temor o presión, tanto a veces como frecuentemente también son mujeres. Sin embargo, quienes lo han hecho frecuentemente y opinan que la violencia en las parejas jóvenes es muy frecuente o que a veces le revisan el celular o le asignan apodos que no le agradan son tanto varones como mujeres, es decir esta tendencia escapa a la distinción de género.

Las reglas anteriores representan un salto informacional significativo respecto de la estadística descriptiva univariada disponible anteriormente. No solo ello, la herramienta de CHIC llamada tipicalidad de los individuos permite identificar, en este caso las personas que responden a la regla. A los efectos solo ilustrativos, a continuación se transcriben los individuos que responden a la regla: quien recibe a veces amenazas que su pareja quiere matarle o matarse cuando se enoja o cuando le ha dicho que quieres terminar entonces “es mujer” (14AmenAvec→Muj):

Grupo optimal : cardinal 30

i554 i553 i402 i842 i759 i352 i351 i846 i493 i345 i194

i143 i847 i540 i550 i574 i82 i314 i715 i712 i387 i788

i502 i645 i371 i700 i48 i60 i818 i786

Este retorno a los individuos fue realizado sobre cada regla de interés a los equipos de trabajo, aquí sólo el grupo optimal de una de ellas. Esta identificación de individuos, que podrían estar en riesgo o que podrían poner en riesgo a otros, resulta de una importancia singular por las posibilidades de prevención que se podrían desarrollar a partir del conocimiento obtenido sobre la información recolectada.

7 Reflexiones acerca de los Resultados

Las reglas anteriormente presentadas, constituyen un insumo sumamente valioso para iniciar una caracterización y dar lugar a una serie de reflexiones, en torno a cómo vivencian las diversas situaciones de violencia los y las jóvenes encuestados/as. En primer lugar y en términos generales, se observa la presencia de conductas violentas, tanto físicas, como psicológicas y sexuales en las relaciones sexo-afectivas de estas parejas. Según lo refieren los y las jóvenes que respondieron la encuesta, esta violencia es ejercida y sufrida tanto por varón es como por mujeres, aunque no en la misma magnitud. Es decir que se podría referir a primera vista la existencia de una violencia ejercida bidireccionalmente. Sin embargo, si se analizan con detenimiento las características de estas formas del ejercicio de la violencia, se deduce que son las mujeres las que padecen situaciones de mayor gravedad, incluida la sensación de que su vida puede estar en peligro. Del mismo modo, son las que han realizado denuncias formales. Autoras como Mariana Palumbo sostienen que en la juventud los vínculos sexo-afectivos poseen características que los diferencian de los que se dan en la etapa adulta de la vida. El amor y las relaciones son entendidas y vivenciadas desde otros lugares. Sus representaciones y prácticas sobre lo amoroso muchas veces se encuentran atravesadas por discusiones, celos y peleas las cuales forman parte de un juego de seducción dentro de los noviazgos. De este modo, "hay contextos de sentido y motivaciones que incluyen a ambos miembros de la pareja que permiten entender el amor vinculado a la violencia. Esto no implica justificar la violencia como una provocación de una parte hacia la otra, sino que permite, por un lado, comprender el sentido por el cual se dan esos juegos y conductas de placer y agresión en un contexto específico; y por el otro, analizar qué pautas de cortejo son consideradas como aceptables para estos jóvenes dentro de los umbrales de la relación y cuáles son percibidas como violentas" (Palumbo, 2017). Esto conlleva a la reflexión en torno a cómo algunas formas de vincularse pueden resultar violentas desde una mirada adulta, pero que en muchas parejas jóvenes, son vivenciadas como parte de un juego de seducción y por lo tanto consideradas aceptables para ellos y ellas mismos/as. Por lo tanto, la dimensión de la violencia física, referida en la pregunta acerca de "si alguna vez te ha golpeado, empujado, pellizcado o tirado de los pelos" (pregunta 13) debe ser interpretada de manera contextual, teniendo en cuenta que en algunos de los casos pudo haber sido respondida desde esta lógica. En términos más estructurales, en los casos de estas parejas que aún no conviven, que no dependen económicamente del otro/a y que no tienen hijos/as en común se observa cierta simetría de poder en la relación. Estas condiciones observables en los vínculos de pareja que se dan en la juventud, se modifican en el transcurso hacia la vida adulta y podrían generar una relación asimétrica de poder, ejercicio de la violencia y sometimiento, en detrimento de las mujeres. Por otro lado, se observa que en las primeras relaciones sexo-afectivas se da un fenómeno que también está presente en las relaciones de pareja de la vida adulta y que tiene que ver con la aparición de las diferentes fases del ciclo de la violencia. Es posible encontrar esto reflejado en la pregunta acerca de "si después de una discusión fuerte se muestra cariñoso/a y atento/a, si te regala cosas y te promete que va a cambiar" (pregunta 15). En situaciones en las cuales las amenazas de muerte han sido frecuentes (Ciclo de la violencia - Estallido de violencia) y posteriormente, de manera frecuente la persona agresora regala cosas y/o se muestra cariñoso/a y promete cambiar (Ciclo de la violencia - Luna de miel), se aprecia la reiteración del ciclo de la violencia en más de una oportunidad y posiblemente el incremento de la intensidad de las conductas violentas. Paralelamente, en

casos con riesgo de vida por las amenazas de muerte y los golpes frecuentemente recibidos, las mujeres no han solicitado asistencia gubernamental a través de denuncias policiales en el último año. Es posible que esto se deba principalmente a uno o a varios de los siguientes motivos:

- Minimizar y/o justificar la situación
- Sentimientos de culpa, vergüenza y/o temor a no sercreídas
- Expectativas que el agresor cumplirá con las promesas de cambiar
- Temor al cese de la relación y/o perjudicar al agresor
- Miedo a que el agresor ejecute la amenaza de terminar con su propia vida o la de ella.

En otros casos, quienes han realizado denuncias policiales en el último año por situaciones de violencia han sido mujeres. Posiblemente los varones no realizan denuncias debido a que se trata de situaciones de menor gravedad y/o a la existencia de creencias limitantes basadas en el machismo, tales como que los varones no sufren violencia ya que deben ser fuertes y viriles, o el temor a cuestionamientos sobre su masculinidad, entre otras. Por otro lado, se apreció que los varones recurrieron a su grupo de amigos en busca de ayuda, probablemente porque reconocen la violencia como una experiencia individual vinculada a cuestiones personales o de salud mental, a diferencia de las mujeres que la consideran como un problema combinado de causas psicológicas, educacionales y familiares, es decir, causas históricas y estructurales. Asimismo, en otras de las reglas presentadas, se observa que los varones justifican la violencia, creen que los celos son una demostración de amor y sostienen que la violencia en parejas jóvenes no existe. En estos vínculos suelen relacionarse los celos a una expresión de amor, interés o cuidado. Esto contribuye con la naturalización de acciones violentas en nombre del amor, invisibilizando prácticas de control. La construcción de una idea de amor romántico ha llevado en muchos casos a justificar o a no reconocer conductas violentas, al mismo tiempo que vulnera las posibilidades de relaciones saludables. De esta manera, se instalan representaciones tales como “por amor, todo se perdona”, “te cela porque te ama” (Arévalo et al., 2018) Retomando el análisis de las reglas, se detectó que en el caso de las mujeres las respuestas fueron más homogéneas, dando cuenta de un mayor conocimiento y reconocimiento de las situaciones de violencia, mientras que, en el caso de los varones las respuestas fueron variadas, con lo cual se produjeron menor cantidad de reglas al mismo nivel de significación (0.95 y 0.90).

La mayor conciencia social acerca de esta problemática es posible de ser interpretada a la luz de diversos acontecimientos ocurridos en nuestro país en los últimos años, entre los cuales podemos mencionar:

- Ni Una Menos, movimiento que comenzó en junio del 2015 a partir del femicidio de una joven embarazada de 14 años, el cual interpeló a la sociedad en general y convocó a una participación masiva de jóvenes a las movilizaciones desatadas.
- Programa Nacional de Educación Sexual Integral aplicado en todos los niveles de las escuelas provinciales.
- Campaña Nacional por el Derecho a la Interrupción Legal del Embarazo (2005) .

La trascendencia de estos y otros movimientos sociales, sumado a la implementación de políticas públicas tendientes a promover la equidad de género, dieron lugar a una

creciente consciencia sobre el impacto de la violencia de género y consecuentemente una mayor participación y compromiso de la ciudadanía y de los y las jóvenes con la temática.

8 Perspectivas

La aplicación del método ASI a la temática de violencia en las relaciones de pareja y en particular contra las mujeres tiene una serie de potencialidades que es posible referir a partir de cuatro aspectos: conocimiento específico, detección, prevención e intervención. En primer lugar, este trabajo permitió profundizar el conocimiento específico acerca de la problemática, contribuyendo a que las intervenciones no sean intuitivas sino basadas en información concreta surgida en este caso, de las respuestas de jóvenes que pueden estar padeciendo o ejerciendo una situación de violencia en algunas localidades de la provincia de Neuquén, Argentina. El conocimiento específico se enriqueció también en otra dirección: la observación de características particulares en las relaciones de parejas de jóvenes demandó la lectura de bibliografía no prevista en el inicio de este estudio, que enriqueció los marcos teóricos utilizados por los equipos. En lo que respecta a la detección, el estudio permitió identificar asociaciones de conductas relativas a la violencia, incluso en un contexto de fenómenos estadísticamente raros. Esto resulta de fundamental importancia dada la complejidad y la gravedad de la problemática, especialmente en estas primeras relaciones de pareja, procurando la detección temprana de estas situaciones y fomentando vínculos saludables en la vida adulta. En lo referido a la intervención, el método permitió identificar a jóvenes en situación de violencia con riesgo alto. A futuro esto resultaría un insumo fundamental para los equipos técnicos de abordaje de estas situaciones en terreno. Como resultado de todo lo expresado anteriormente, se considera que la aplicación de ASI contribuiría al desarrollo de políticas públicas tendientes a prevenir y erradicar la violencia contra las mujeres, logrando resultados más acertivos al contar con información precisa y confiable.

Referencias

[1] Arévalo, L., Romero, G., &Ferreyra, A. (2018). *Violencia en el Noviazgo*. Instituto de Estadísticas del Defensor del Pueblo.

[2] Ley de ProtecciónIntegral a las Mujeres (2009).

[3] Censos, D. P. d. E. y. (2020). *AnuarioEstadístico de la Provincia del Neuquén*. Gobierno de la Provincia del Neuquén.

[4] Gras, R., Ag Almouloud, S., Bailleul, M., Larher, A., Polo, M., Ratsimba-Rajohn, H., &Totohasina, A. (1996). *L'Implication Statistique*: La Penséée Sauvage.

[5] Gras, R., & Bailleul, M. (Eds.). (2000). *La Fouille dans les données par la méthode d'analyse statistique implicative*. Caen: ARDM.

[6] Gras, R., Couturier, R., Guillet, F., &Spagnolo, F. (2005, 30 mai - 1er juin). *Extraction de règles en incertain par la méthode statistique implicative.* Paper presented at the 12èmes Rencontres de la Société Francophone de Classification, Montréal, Canada.

[7] Gras, R., & Régnier, J.-C. (2009). Analyse implicative des variables binaires. Intensité implicative. Intensité entropique *Analyse Statistique Implicative Une méthode d'analyse de données pour la recherche de causalités* (pp. 17-129). Toulouse: Cépaudès Editions.

[8] Mujeres, O. d. V. c. l. (2020). *Informe Anual de Violencia contra las Mujeres*. Gobierno de la Provincia del Neuquén.

[9] Reglamento de Ley 2786 (2015).

[10] Palumbo, M. (2017). *Las dinámicas de la violencia contra las mujeres y el amor en los jóvenes*:Teseo.

MODÉLISATION DU DÉVELOPPEMENT COGNITIF MISE À L'ÉPREUVE PAR L'ASI : APPROCHE GÉNÉRALE, APPROCHE DIFFÉRENTIELLE ET QUESTIONS MÉTHODOLOGIQUES

Christian Pellois[1]

MODELING COGNITIVE DEVELOPMENT, PUT TO THE TEST BY THE ASI: A GENERAL APPROACH, DIFFERENTIAL APPROACH AND METHODOLOGICAL QUESTIONS

RÉSUMÉ

Les multiples approches du développement du raisonnement de l'enfance à l'âge adulte amènent à se poser des questions sur une modélisation de ce développement. L'ASI peut constituer un outil pour traiter de la confirmation ou non, de cette modélisation hypothétique, portant sur des cheminements pluriels, et aboutissant à sa maîtrise. Cette utilisation de l'ASI soulève des questions de fonds concernant la réalité des différences et pose quelques questions d'ordre méthodologique.

Mots-clés : *développement cognitif, modèle, différences, méthodologie*

ABSTRACT

The multiple approaches to the development of reasoning from childhood to adulthood raise questions about a modeling of this development. The ASI can be a tool to deal with the confirmation or not, of this hypothetical modeling, dealing with plural paths, and leading to its mastery. This use of the ASI raises substantive questions about the reality of the differences and raises some methodological questions.

Keywords: *cognitive development, model, differences, methodology.*

1 Introduction

La culture scientifique constitue, c'est l'évidence, une expression parmi d'autres de ce que l'on appelle couramment *« l'intelligence »* humaine. Il est tout aussi évident que cette expression culturelle, se développe, et prend sa place à côté, mais aussi en interaction, avec d'autres aspects de la culture humaine. Il en est ainsi de la communication et de l'expression des organisations et conceptions politiques et sociétales, de la sensibilité et de la compréhension de ce qui se passe dans les interrelations entre personnes, les expressions émotionnelles, affectives, sociales. Il en est également, de même de la créativité, de l'esthétique, des arts, en général, comme de la réflexion morale, éthique, philosophique, etc.

Ces différentes formes d'expression de l'*intelligence* humaine peuvent présenter, pour partie, des aspects communs, comme le fait de s'appuyer sur la science pour traiter des organisations sociétales. Mais ces différentes formes peuvent, aussi, relever d'aspects distincts, voire d'aspects, au moins apparemment, *contradictoires*[2], entre eux. Il en serait ainsi d'un imaginaire débridé et s'exprimant sans contrainte, dans la créativité ou l'art, ceci au regard de la prise en compte des réalités et des contraintes de ce monde comme

[1] 7 rue des pins, Courseulles-sur-Mer, 14470, pellois.christian@wanadoo.fr

[2] Ce qui renverrait, par exemple, à la question : *« Que perdons nous en apprenant à lire ? »* (Dehaene, 2007, p. 280).

c'est le cas dans des approches à caractère scientifique, qui n'en nécessitent pas moins, quelquefois, de faire appel à certaines formes ciblées de cet imaginaire là ...

La culture scientifique fait donc partie d'un ensemble plus vaste constitutif de la culture humaine. Mais, cette culture scientifique, utile à cet ensemble plus vaste, est, elle-même, faite de disciplines, de domaines, également distincts et plus ou moins en interaction, en interpellation, voire en confrontation, entre eux. Il en est ainsi, notamment, du domaine des mathématiques (toutefois dans une position qui lui est plus spécifique), de la physique, de la biologie, etc., et des sciences humaines.

A des degrés divers, ces différents domaines de la culture scientifique, recourent, dans le cadre de ce que peut étudier la psychologie générale, à des comportements, des ensembles d'*habiletés*, de *capacités* diverses, et variées, notamment *cognitives*, mais pas seulement. Du point de vue d'une psychologie davantage appliquée, par exemple au travail, la culture scientifique, mais aussi technologique, pourra relever davantage, de *compétences*, à acquérir, à faire évoluer et même à créer puis développer, plus ou moins communes ou différentes, et plus ou moins mobilisées selon les domaines scientifiques ou technoscientifiques considérés.

Il en est ainsi des savoirs et savoir-faire, voire de *« savoir-être »*, relevant de la maîtrise des conceptions théoriques du domaine considéré, comme des capacités d'observations -à partir d'appareillages plus ou moins sophistiqués-, de maîtrise des mesures et des résultats d'expériences, faites avec des protocoles et des instruments plus ou moins complexes, et plus ou moins sensibles aux conditions contextuelles ou spécifiques à ces expériences.

Il en est de-même des qualités et de la pertinence : à en tirer des explications, des interprétations plus ou moins délicates à formuler, et plus ou moins fiables et avérées, et à articuler ces interprétations avec les conceptions scientifiques du domaine considéré.

Il en est également ainsi des capacités à concevoir, et ceci de manière judicieuse, ce que Einstein appelle des *« expériences de pensée »* (Klein, 2016, p. 168), de la faculté, plus ou moins développée, à contourner, dépasser des obstacles à caractère épistémologique, tel que G. Bachelard (Bachelard, 1987, par exemple, p. 24 et suivantes, p. 46, p. 119, p. 126, p. 178) a pu en rendre compte. Comme il en est ainsi, des capacités, à compiler, ordonner, calculer, et -ce dont il va être question maintenant-, *raisonner* et donc mettre en œuvre des capacités à utiliser la logique inhérente aux formes de raisonnement développées.

2 Brefs rappels théoriques des cadres conceptuels sous-jacents aux données soumises aux analyses statistiques réalisées au moyen de l'ASI, modélisation et variables prises en compte

2.1 Le cadre conceptuel

Le maîtrise du *raisonnement*, tout comme la maîtrise du langage écrit au regard de la compréhension et de l'appréciation littéraire, l'éducation esthétique dans l'art, etc., constitue bien, une dimension centrale du développement et de la mise en œuvre de la culture scientifique telle qu'évoquée précédemment. Mais les différentes capacités acquises en ce domaine donneraient lieu, semble-t-il, à des apprentissages très

progressifs, notamment scolaires, ainsi qu'à un développement de la personne, de l'enfance à l'adolescence. Ces capacités de raisonnement ne seraient donc pas, d'emblée, disponibles, et, par ailleurs, référencées aux expressions culturelles d'une société donnée. Elles devraient bien attendre *« maturité »* et acquisition de connaissances pour pouvoir être pleinement mises en œuvre.

Pour autant, ces *« capacités »* du raisonnement peuvent-elles relever de ce que R. B. Cattell (Cattell, 1971) a appelé, dans ses conceptions factorielles de l'intelligence humaine, d'une intelligence dite *« cristallisée »*, dont la maîtrise des opérations logiques sous-jacentes ferait partie ? Il semblerait qu'il s'agisse plutôt, là, de ce que cet auteur a appelé intelligence *« fluide »*, conception reprise auprès de Cattell, dans des modèles factoriels plus élaborés (Carroll, 1980, 1993), et, même, par les neurosciences (Dehaene, 2018, p.230).

Une structure formelle de ces opérations logiques évoquées ci-dessus, appelée alors, *pensée opératoire*, a pu être dégagée et formalisée (Piaget, 1972). Cette structure est apparue constituée, à la base, d'opérations logiques élémentaires, puis de sous-structures constitutives d'ensembles de plus en plus larges (ensembles d'opérations de classes et de relations, puis groupements de classes et de relations, et groupes dits de *« quaternalité »*), aboutissant à cette structure d'ensemble, organisant formellement le raisonnement (combinatoire, logique propositionnelle, groupe INRC). Cette organisation par étapes a été supposée s'inscrire dans un développement *logique* de stade en stade, de l'enfance à l'adolescence ce que des travaux, tels ceux de J. Piaget et collaborateurs (Inhelder, Piaget, 1970) ont voulu étayer, notamment, dans la seconde partie du vingtième siècle.

Il semblerait donc bien qu'une organisation logique du sujet se développe progressivement et existe, in fine, donnant lieu à l'expression d'habiletés de *compétences* et de *performances* spécifiques en ce domaine du raisonnement scientifique, appelé, selon les manières d'en rendre compte, *« logico-mathématique »* ou*« hypothético-déductif »*.

Cependant, une chose est de rendre compte de cette organisation structurelle formelle disponible dans le raisonnement, et vue essentiellement sous un angle *épistémique*, une autre est de saisir, au plan *psychologique*, toute la complexité de son acquisition (développement), puis de son utilisation (expression) par l'humain.

En ce sens une observation fondamentale doit être faite. Dans ce développement des capacités logiques, durant l'enfance puis l'adolescence comme dans leur expression à l'âge adulte, existe une grande disparité, une grande variabilité des formes prises par le raisonnement et plus précisément la *pensée opératoire* du sujet. Ces formes peuvent être *« hiérarchiques »* car plus ou moins complètes d'un sujet à un autre. Mais ces formes peuvent être, aussi, différenciées, au plan *« qualitatif »*, en fonction, des différents sujets considérés (variabilité interindividuelle), par exemple, sous l'angle de leurs *styles cognitifs* dominants, mais aussi, pour un même sujet, en fonction du moment et des différentes situations, contextes, et contenus traités (variabilité intra-individuelle).

Ces remarques faites sur la genèse du développement ont amené, alors, chercheurs et théoriciens à se pencher sur les *procédures* utilisées par les sujets aux prises avec le raisonnement, ceci tout autant qu'à se focaliser uniquement sur les étapes de l'organisation de la *structure* logique formelle. Cette préoccupation déjà présente dans les travaux de collaborateurs de Piaget (Inhelder, Cellerier, 1992), prend toute son ampleur lorsque la recherche se penche, notamment par le passage par le *traitement de l'information*, sur l'analyse proprement cognitive de l'expression du raisonnement, que

celle-ci soit généraliste (Richard, 1990) ou différentielle (Lautrey, Chartier, 1987). Il s'agit alors, d'aborder, là, les modalités, tout à la fois, tactiques et stratégiques, de ce traitement de l'information, modalités mises en œuvre par le sujet aux prises avec des conduites adaptatives, et plus précisément, la résolution de problèmes. Cette manière de voir l'expression cognitive par le biais *« du traitement de l'information »,* a trouvé ultérieurement toute sa place dans l'approche *composantielle « dont le but est d'isoler les processus communs aux différentes tâches saturées par un même facteur d'intelligence »* (Huteau, Lautrey, 1999).

Mais, si l'approche par les *procédures*, ou les *processus*, a pu mettre en évidence des différences entre humains, et des différences de mise en œuvre de procédures, pour le même sujet humain, entre différentes situations, et contenus considérés, elle a aussi amené à formuler des conceptions mixtes procédure/structure, complexes, traitant, à la fois, de ces deux formes de différenciation (Lautrey, 1990).

Ces nouvelles perspectives nécessitent, alors, de prendre en compte des dimensions cognitives, des habiletés, des *capacités* plus spécifiques à cette approche particulière qu'est le traitement de l'information[3], dont ne tenait pas nécessairement, ou pleinement compte le modèle piagétien.

Et ces dimensions cognitives sont susceptibles d'interférer, de manière différentielle, par facilitation ou perturbation, sur le développement et l'expression, au sens classique piagétien, des structures opératoires *logiques* mobilisées dans le raisonnement.

Une précédente publication au colloque ASI 10 (Pellois, 2019) suggère que, si la structure d'ensemble s'acquière bien progressivement et conformément à la théorie, par étapes successives (stade concret, intermédiaire, formel A, formel B), il existerait un rôle possible des *« procédures »* dans le traitement de contenus (et donc de l'information soumise au sujet) et des formes et modalités d'expression logique, dont il conviendrait de saisir les formes propres. Et ces*« traitements »* et*« formes »* seraient susceptibles d'interférer avec la structuration progressive, puis l'expression structurale des capacités logiques acquises (Pellois, 2019).

Ces perspectives procédurales, ou processuelles, évoquées ci-dessus, ne seraient pas très éloignées de ce qui est exploré, et ceci déjà depuis maintenant quelques décennies, dans le domaine des neurosciences. Pour les neurosciences, des potentialités *« constitutionnelles »*, des ressources précoces du sujet, seraient à disposition, afin de permettre, au-delà du développement du langage oral chez l'humain, l'expression de la maîtrise et de la compréhension du langage écrit. Ceci au même titre que la maîtrise basique des quantités et de leurs possibles comparaisons, l'installation dans le cerveau de *« cartes de localisation »*, etc., aboutiraient à des compétences mathématiques.

Ces processus et les procédures fonctionnelles qui en découlent, opèreraient soit par des *dispositions* -mais aussi des *contraintes* relevant de l'évolution proprement humaine-, (Dehaene, 1997), soit par des ressources issues d'un *recyclage neuronal* (Dehaene, 2007), exercés par la culture et dans l'école, et tôt installés dans le développement humain. L'établissement de la preuve passe alors, à la différence des perspectives précédentes s'appuyant seulement sur l'expression de performances et leur traitement

[3] Telles la discrimination et la clôture perceptive, les styles perceptifs/cognitifs, les capacités attentionnelles, la maîtrise du langage, les capacités de représentation spatiale, différents niveaux mnésiques du traitement de l'information (mémoire de travail, mémoire à court terme ou MCT, etc.).

statistique, par l'imagerie médicale des zones activées au regard de la prise de conscience cognitive et de l'exercice dûment répertorié dans le fonctionnement cérébral observé, des différentes habiletés mises en œuvre par le sujet.

Mais, cette nouvelle perspective ouvre aussi la porte à un regard différentialiste (Jobard, 2002) et ceci même reconnu par des auteurs (Dehaene, 2007) affichant, assez régulièrement, leur position *« généraliste »* dominante (Dehaene, 2018). Il serait alors possible de concevoir deux voies *« support »* de l'expression cognitive, une voie propre au langage, une voie propre au traitement spatial des données de l'information manipulée par le sujet *« pensant »*. Ces deux voies rejoindraient les conceptions hiérarchiques les plus élaborées et les plus avérées, de l'approche factorielle de l'intelligence humaine distinguant deux grands facteurs de groupe, l'un verbal/numérique, l'autre spatial/représentatif, tous les deux permettant l'expression et le développement de la pensée. Mais le modèle factoriel s'est trouvé être largement complexifié, par la suite, au regard des travaux de R. B. Cattell et surtout, de ceux déjà précédemment évoqués, de J. B. Carroll (Carroll, 1980, 1993). Et, par ailleurs, dans ce domaine de l'intelligence humaine, plus particulièrement de l'expression du raisonnement, il y aurait aussi des *processus*, relevant globalement de la personne, et, au-delà des styles cognitifs, déjà évoqués (Huteau, 1982), de la personnalité (Huteau, 1985), de capacités d'inhibition (Houdé, 1995), et même du plan affectif et émotionnel (Damasio, 1995), l'ensemble mettant en œuvre la pensée *éduquée* de l'adulte, celle de l'intelligence *« cristallisée »* bien sûr, mais aussi de l'intelligence *« fluide »* de R. B. Cattell (Cattell, 1971), et, du côté du raisonnement scientifique, de la *pensée opératoire* piagétienne.

Ces dimensions seraient-elles, pour partie, à rapprocher des ressources *recyclées*, ou très tôt *mises à disposition* par ce qui serait le *« jeu »*[4] de l'évolution, dans l'expression des interactions homme/milieu évoquées par les neurosciences ?

La question, quelque peu d'actualité, pourrait être, alors, la suivante : n'existe-t-il pas, à côté de ces *dispositions* permettant la lecture l'écriture, le calcul, d'autres dispositions, d'autres *habiletés initiales*, comme la discrimination et la clôture perceptive, des dispositions de mémoire à court terme, etc., d'autres *« intuitions*[5] *profondes »* tels, selon S. Dehaene, *« une physique naïve des objets et de leurs comportements, un sens des cartes et de la navigation dans l'espace, des intuitions du nombre, de la probabilité et des mathématiques, …, un véritable génie pour les langues. »* (Dehaene, 2018, p. 200) ? N'en serait-il pas, alors, ainsi de ce qui constitue, dans le modèle factoriel, l'expression d'un *facteur général* d'intelligence, antérieurement évoqué (Pellois, 2019), et donc de certaines formes d'expression de l'*« intelligence fluide »* selon Cattell (Cattell, 1971) ? Il se pourrait bien que ces questions deviennent particulièrement centrale et d'actualité au regard même des neurosciences pour lesquelles il existerait, là, au travers, notamment, de l'entraînement du contrôle exécutif de l'activité cognitive, un enjeu de développement de l'enfance à la vie adulte, puisque selon ce même Dehaene (Dehaene, 2018, p.230),

[4] Soit la part de liberté d'expression et d'émergence tenant aux mutations, mais aussi à l'expression de l'épi génétique, aussi bien que des parts de libertés que se donneraient les cultures au regard de contraintes sur lesquelles il est possible de jouer aux marges des réalités constitutionnelles de l'humain.

[5] Ce terme est, quelque part, pour le moins, malheureux au regard de sa connotation habituelle, référant au jaillissement de l'esprit, à une émergence inattendue, originale et possiblement, éminemment pertinente, alors, qu'ici, il réfère à des ressources initiales basiques, et, surtout ressources aux contraintes étroites tenant à l'expression du génotype/phénotype, même sous des angles conceptuels plus ouverts, tels que l'épi génétique ou le tout nouveau concept de *« recyclage neuronal »* …

« L'entraînement du contrôle exécutif modifie même le quotient intellectuel. Ce qu'on appelle l'intelligence fluide -la capacité à raisonner et à résoudre des problèmes nouveaux- fait massivement appel aux systèmes cérébraux de contrôle exécutif : tous deux mobilisent un réseau similaire d'aires cérébrales, … les tests standardisés qui mesure l'intelligence fluide ressemblent à ceux qui évaluent le contrôle exécutif : tous deux privilégient l'attention, la concentration, la capacité de passer rapidement d'une activité à l'autre, sans perdre de vue le but global de la tâche. ».

2.2 La modélisation

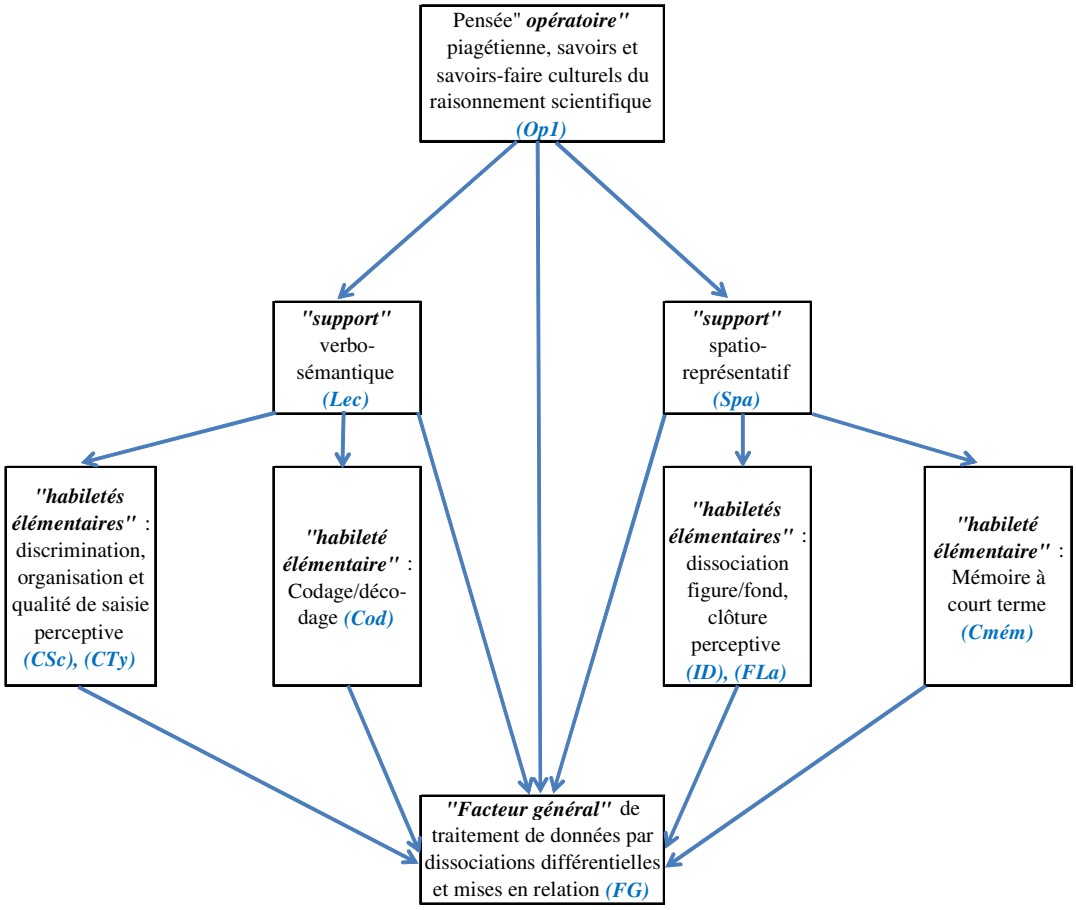
Les conceptions développées ci-dessus nous amènent, inévitablement ici, et ceci pour clarifier les rapports entre toutes ces dimensions considérées, à tenter une modélisation large, faisant l'hypothèse de relations de *« cause-à-effet »* entre ces *habiletés initiales*, dont un *facteur général* d'intelligence, les capacités dites *« support »* et des formes culturellement plus élaborées de l'intelligence *fluide,* dont le raisonnement abouti de l'adulte, la pensée *opératoire* piagétienne, ferait partie, et, in fine, dans son expression culturelle, de l'intelligence *cristallisée*, davantage en référence à un développement dit *« verbal/numérique »*, et les multiples expressions culturelles telles celle de la pensée scientifique, etc.

Pour être plus précis, la *« construction »* culturelle du raisonnement humain ainsi réalisée, pourrait se faire par l'action intermédiaire, tout d'abord, de variables dites *supports*, telles la maîtrise du langage écrit et de sa compréhension –avec la difficulté conceptuelle que cette variable serait, aussi, possiblement, selon Carroll (Carroll, 1980, 1993), l'expression, elle-même, d'une forme d'intelligence *cristallisée*-, et des capacités de représentation spatiale. Mais cette construction et l'expression fonctionnelle des variables *« support »* dépendrait, elles-mêmes, de l'expression d'*habiletés cognitives élémentaires,* acquises précocement, et exploitées au service du traitement de l'information nécessaire à un fonctionnement cognitif du sujet, culturellement organisé.

Cet ensemble de relations fonctionnelles, supposées de *« cause-à-effet »*, et modélisées, relèverait, alors, de l'observation, après traitement par l'ASI, de relations implicatives entre les variables considérées. A savoir une relation de *« cause-à-effet »* peut être envisagée entre la variable A et la variable B si l'on peut observer, au plan statistique, que *B implique A*[6], et pas l'inverse. Toutefois, comme cela sera évoqué par la suite, cette condition, bien que nécessaire, n'est pas obligatoirement suffisante pour certifier cette relation de cause-à-effet possible de A par rapport à B.

La modélisation proposée est la suivante :

[6] Tel qu'exprimé par le sens donné ci-dessous aux flèches représentées dans la modélisation.



Graphe 1 – Modélisation, hypothétique, dynamique et « *implicative* » du développement

2.3 Source des dimensions et variables de la modélisation

Il s'agit d'anciennes données (Pellois, 1986, Annexes II et IV). Le recueil de données a été réalisé sur une population de 175 élèves scolarisés en classes de 6ème. Seize dimensions ont été retenues, renvoyant à autant de variables définies sous deux statuts différents, soit : les variables considérées comme pertinentes et valides selon le modèle, et des variables considérées comme variables *« témoins »,* présentes, elles, pour valider leur absence d'interaction avec les variables jugées pertinentes. Ce nombre passe à dix-neuf dimensions, lorsque la variable *Op1* est subdivisée en ses sous-parties (variables *Crois1, Comb1, Dess1, Lamp1*).

2.3.1 Variables considérées comme entrant effectivement en interaction dans la modélisation

Les dimensions et épreuves de performances proprement dites, numérotées de 1 à 10, sont les suivantes :

a) Les capacités *opératoires,* au sens de la théorie piagétienne, relèvent du recueil des performances à l'Echelle Collective du Développement Logique (ECDL) mise au point par J. Hornemann à partir des travaux de F. Longeot (Pellois, 1986). Une première analyse de la cohérence interne, et de la cohérence avec la théorie opératoire piagétienne, de

l'épreuve a donné lieu à une présentation dans une publication antérieure (Pellois, 2019), publication dans laquelle cette épreuve est présentée dans le détail. Cette capacité *opératoire,* au sens piagétien, renverrait bien, chez l'adulte, à une certaine forme d'expression aboutie du raisonnement scientifique (variable *Op1*(1)[7]), mais aussi à une certaine forme, culturellement *« sophistiquée »,* de l'intelligence *« fluide »* (Carroll, 1980, 1993).

b) La capacité générale en termes de *facteur G* des conceptions factorielles (au sens de Spearman) relève du recueil des performances à l'épreuve D 48. Cette capacité, renverrait à l'expression, très tôt installée dans le développement de l'enfant, d'un *facteur général* de développement cognitif (Carroll, 1980, 1993) et/ou de l'intelligence *« fluide »* (Cattell, 1971) plus précocement disponible que celle issue du développement *opératoire* piagétien. Il s'agit, ici, d'une succession de dominos, dont les contenus sont ordonnés selon une, ou plusieurs règles. Le dernier domino de la série étant vide, il convient de trouver la ou les règles d'organisation des contenus des dominos qui précèdent pour trouver les valeurs des deux cases de ce dernier domino. Cette variable (*FG* (2)) a aussi déjà été prise en compte dans la précédente publication évoquée ci-dessus (Pellois, 2019).

c) La capacité de lecture et de compréhension du langage écrit est évaluée à l'aide des feuilles *D, E, F* (Pellois, 1986) de l'épreuve de *« Lecture silencieuse »* (variable *Lec*(3)). Cette capacité constitue une des deux variables dites *« support »,* mais aussi l'expression d'une forme d'intelligence *« cristallisée »* (Cattell, 1971, Carroll, 1980,1993). L'épreuve est composée d'une succession de textes assez courts. A la suite de chacun des textes, cinq questions précises sont posées, questions permettant de vérifier la bonne compréhension du texte qui les précède.

d) La capacité à se représenter des objets dans l'espace est évaluée à l'aide de l'épreuve *« Briques »* (Pellois, 1986). Elle constitue la seconde variable dite *« support »*. Cette épreuve est composée d'une succession de tas de briques schématisées, il convient de trouver, parmi des propositions de schémas de réponse, la configuration correspondant à la vue indiquée par une flèche. Les tas de briques sont, bien évidemment, toujours disposés, de manière différente (variable *Spa* (4)) pour chaque exercice proposé.

e) La capacité consiste ici à trouver, dans une série de dessins complexes, disposés en ligne, deux dessins, contenant un dessin plus simple, qui y est caché et dont la reproduction est présentée en début de ligne. L'épreuve utilisée est le MC13-T de Kouteinikoff (Pellois, 1986). Cette épreuve a été choisie pour traiter d'une dimension qui paraît assimilable à la capacité de *dépendance/indépendance* à l'égard du champ perceptif (variable *I/D* (5)).

f) La capacité recherchée ici relève d'une *clôture perceptive signifiante* réalisée à partir de dessins incomplets. Cette capacité permet de saisir de quel objet schématique il s'agit (un chat, un banc, une cafetière, une fleur, etc.) et de le désigner par écrit. L'épreuve utilisée est celle des Figures Lacunaires (Pellois, 1986) de R. Zazzo (variable *FLa* (6)).

g) La capacité consiste à reproduire des symboles schématiques associés à des chiffres. Le sujet doit, dans les cases vierges, sous les chiffres, reporter le symbole inscrit

[7] Un second recueil de données portant sur la dimension opératoire, et dont il sera brièvement question ultérieurement, a été de nouveau réalisé sur la même population, un an plus tard, en classe de 5ème, ce second recueil de données a été dénommé (*Op2)*.

sous chaque chiffre, en tête de l'exercice, (Cf., par exemple, Hauser, V., 2014[8]). Il s'agit de l'épreuve de Code B extraite de l'Échelle d'intelligence de Wechsler pour enfants (variable *Cod* (7))

h) Trois capacités sont évaluées avec l'aide de ce que l'on appelle la *« Figure de Rey »*. Il s'agit d'un dessin schématique abstrait[9], à reproduire à main levée. Cette reproduction se fait par étapes successives, avec l'utilisation de 4 couleurs différentes, étapes de 45 secondes pour les deux premières, puis d'une minute 20 secondes pour les deux suivantes. Ces étapes sont suivies d'un temps supplémentaire de 3 à 4 minutes. Ce temps est ajusté en fonction de la réalisation de la copie par la plupart des élèves considérés. Puis, il est demandé de faire le même exercice, de mémoire, et ceci après un temps de pause de trois minutes. Les trois capacités retenues sont : la capacité à reproduire, le plus exactement possible, la figure considérée, la capacité à la reproduire, de mémoire, et enfin la capacité à faire valoir, dans quel type de structuration du dessin s'est faite la copie de cette même figure (respectivement, variables *CSc, Cmém, CTy* (8) (9) (10)). Le type de structuration est évalué à partir des parties du dessin de la Figure de Rey réalisés successivement à l'aide de couleurs différentes (cf. consignes ci-dessus).

2.3.2 Variables considérées comme *« témoins »*

Les capacités retenues, numérotées de 11 à 16, sont les suivantes :

a) Il s'agit de la capacité à produire, en un temps donné, de manière divergente, le plus possible de dessins originaux figuratifs, à partir de motifs simplifiés, et sans signification particulière, de taches d'encre informelles, présentés en début de feuille, ou de mots à partir d'une consigne simple, tel que : *« les choses rondes »*, *« ce qu'on peut manger »* (variable *Flui* (11)). L'épreuve considérée est tirée du test de facteur *F* de R. B. Cattell[10].

b) Le but, ici, est d'aborder les capacités : à développer une vitesse grapho-motrice, et exprimer une souplesse/rigidité adaptative grapho-motrice. Il s'agit, d'une épreuve tirée également des travaux de R. B. Cattell. L'objectif est de reproduire, le plus rapidement possible, des séries de nombres de lettres ou de mots comme habituellement dans l'écriture courante, puis, alternativement dans la succession des exercices, de les reproduire à l'envers. La vitesse grapho-motrice (VM) est mesurée en établissant le score du nombre de reproduction. L'indice de persévération (IP) relève de la différence de score entre le nombre de signes réalisés à l'endroit et le nombre de signes réalisés à l'envers (Pellois, 1986). (variables *VM, IP* (12) (13))

c) La capacité abordée consiste à discriminer un motif parmi d'autres. Il s'agit de l'épreuve de barrage de R. Zazzo (Pellois, 1986). Dans un premier exercice, le sujet doit toujours discriminer le même motif parmi d'autres motifs presque semblables et le barrer. Dans un deuxième exercice le motif à discriminer change régulièrement ce qui amène le

---

[8] Hauser, V., 2014. Les fluctuations dans les épreuves de vitesse de traitement du WISC-IV sont-elles liées aux capacités cognitives ? Master, Université de Genève.

[9] extraite de :
http://www.rqrv.com/fr/document/outils_accueil/Fiche%20descriptive%20de%20la%20Figure%20de%20Rey%20Final.pdf

[10] Bénassy , M., Chauchard, C., (1942). Le test F de Cattell est-il un test objectif de tempérament ? *L'Année psychologique*, 1942, 43-44, *200-230.*
https://www.persee.fr/doc/psy_0003-5033_1942_num_43_1_7876

sujet à devoir tout le temps adapter sa capacité discriminative au nouveau motif à barrer (variable *Disc* (14) et *Adap* (15)).

d) Cette dernière variable *« témoin »*, traite de l'expression d'un trait de la personnalité portant sur l'introversion *vs* l'extraversion (variable *I/E* (16)). Il s'agit, dans cette épreuve faite d'une série de propositions, d'établir, en fonction des réponses cochées, si la personne est plus centrée sur le monde qui s'exprime en soi-même (introversion *I*) ou sur le monde à l'extérieur de soi (extraversion *E*). (Cf. Pellois, 1986).

3 Les résultats obtenus après traitement des données par l'ASI

3.1 Les résultats de l'analyse implicative au regard du modèle

Les données recueillies, pour les 175 élèves considérés, ont été soumises au logiciel CHIC (version 4.2[11]). Dans le graphe implicatif obtenu, toutes les relations directes (traits pleins), ainsi que transitives (traits pointillés), ont été recherchées. Point important à souligner : les indices choisis pour traiter de l'intensité d'implication, associés aux différentes couleurs des flèches ci-dessous n'ont pas été choisis au hasard. Ils tiennent compte de deux préoccupations (Cf. critères a) et b) ci-dessous), dans une situation complexe où les fluctuations possibles dans le recueil des données, ne permettent pas d'espérer des valeurs implicatives très élevées.

a) Il n'a pas été retenu des indices d'implication très élevés (*« exigeants »*) tels 0.99, 0.95, 0.90, indices qui entraineraient, de facto, le constat d'absence de relations implicatives entre la plupart des variables considérées.

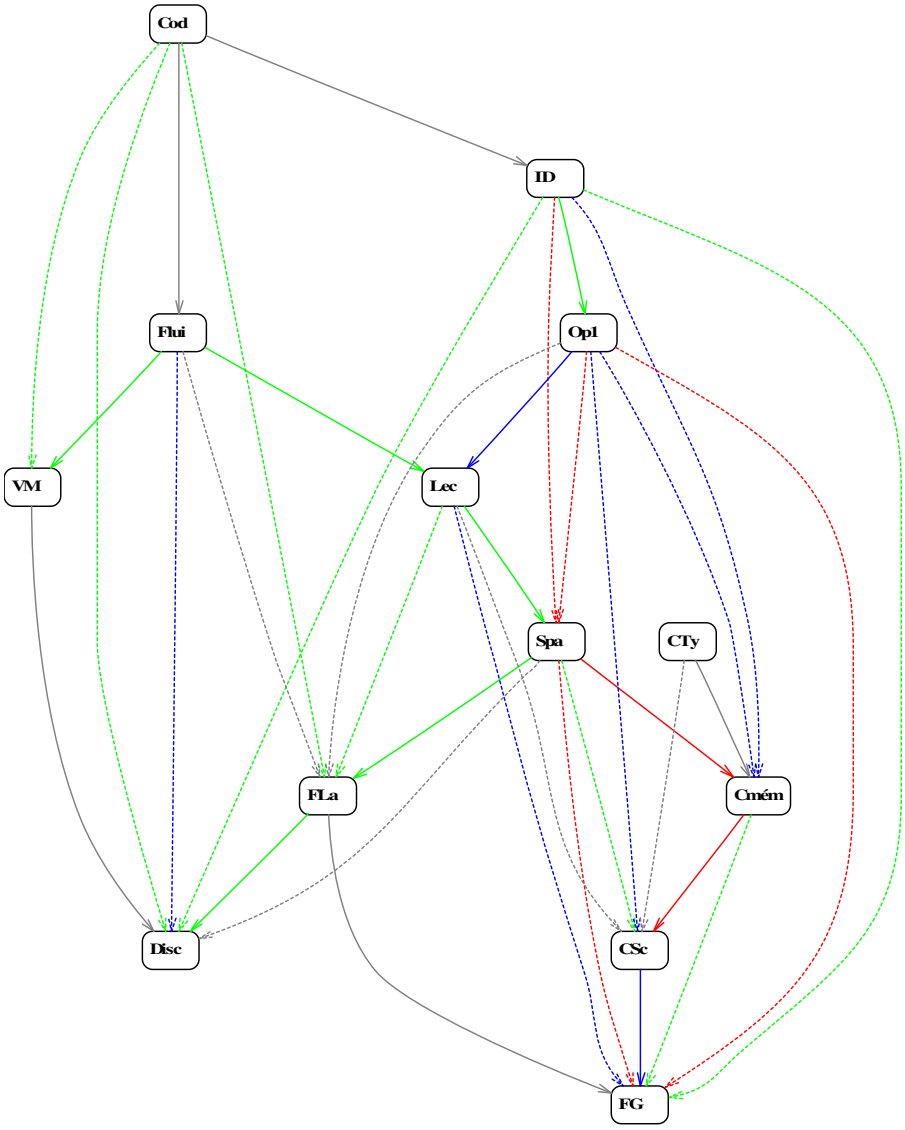
b) il a été retenu des valeurs d'indices d'implication que l'on peut considérer encore comme pertinents au regard de *« l'étonnement statistique »* que l'on pourrait raisonnablement, avoir compte- tenu du faible nombre, ici relatif, de contre-exemples (Gras, 2018, notamment, p. 18 et suivante).

Ce second critère, vise également à ne pas tomber dans un ensemble implicatif de relations à des valeurs d'indices jugés trop *« peu exigeantes »* en termes d'intensité d'implication, ensemble se traduisant, souvent, par une chaîne davantage *« linéaire »* de relations implicatives possiblement *« douteuses »*[12] et moins propices, par l'enchaînement alors constaté, au repérage clair d'ensembles différentiels constitués, plus partiels, et de formes *« vicariantes »* ou « *équifinalisées »* dans le *« système »* (Cf. Von Bertalanffy, 1980) telles que la modélisation du point 2.2 le formule implicitement.

Pour cette raison, les indices choisis et retenus, ceci après analyse empirique des données sous différentes intensités d'implication, sont : 0.80 (couleur rouge), 0.75 (couleur bleue), 0.70 (couleur verte) et 0.65 (couleur grise).

[11] Sur le plan de l'analyse *implicative*, le traitement des données n'a pu se faire sous CHIC 7b, pas opératoire au plan de cette analyse sur l'ordinateur utilisé, alors qu'il a pu se faire sans difficulté pour les analyses en termes de *similarité* ou *cohésitives*, d'où son utilisation par la suite.

[12] La position *« d'étonnement statistique »*, au regard du nombre des contre-exemples possibles ne tenant plus vraiment. (Résultats non publiés)



Graphe 2 – Ensemble des relations *implicatives* entre variables après traitement par l'ASI (CHIC 4.2)

Si nous comparons l'ensemble du réseau des relations implicatives, dégagé par CHIC, et l'ensemble du réseau des relations implicatives du modèle, nous pouvons faire les constatations suivantes.

Il apparait bien que la variable traitant des capacités *opératoires* piagétiennes et relevant de l'intelligence dite *« fluide »* (*Op1*), mais sous l'angle culturellement plus *« sophistiqué »* de l'expression culturelle du raisonnement dit *« hypothético-déductif »*, se situe effectivement dans la position, conforme au modèle. Les performances en ce domaine impliquent bien tout un ensemble de capacités plus *« initiales »*, et qui pourraient être reconnues comme « *intuitives* » au sens des neurosciences. Il en est ainsi, particulièrement, du facteur général de l'intelligence *(FG)* qui paraît bien constituer une capacité d'intelligence *« fluide »*, tôt installée, traitant, au sens des neurosciences, de *l'attention/concentration* comme des *systèmes cérébraux du contrôle exécutif* (Cf. point 2.1). De cette capacité à caractère plus *« générique »*, dépendrait bien d'autres capacités (soit, ici, 7), dont les capacités dites *« supports »* (*Lec, Spa*) dans la modélisation, mais

aussi des variables plus *« élémentaires »*, comme les capacités à reproduire une figure complexe (*Csc*) et à organiser la reproduction de la copie de la Figure de Rey (*CTy*), ou celles faisant jouer efficacement la mémoire à court terme (*CMém*), ainsi que des capacités de clôture perceptive *« signifiante »(Flac),* et, in fine, bien sûr, la capacité de l'intelligence *opératoire* piagétienne*(Op1)*.

Mais quelques *« surprises »* apparaissent, ou tout au moins, quelques résultats inattendus qui posent question.

La première *« surprise »*, relève de positions inattendues, au regard du modèle, de capacités considérées comme *nécessaires* au développement préalable des capacités supports (*Lec* et *Spa*), comme de l'intelligence *fluide*, mais sous une forme élaborée (*Op1*). Il s'agit des variables portant sur les capacités d'encodage *(Cod)* et sur la capacité à distinguer une figure simple dans une figure complexe, capacité ici assimilée au style *« dépendance/indépendance »* à l'égard du champ perceptif *(ID)*. Or, il en va différemment dans le graphe implicatif obtenu, après traitement des données, ces deux dimensions apparaissant tout en haut de ce graphe. La dimension *(ID) « implique »* même la dimension *(Op1)* ! La dimension du *codage* (*Cod*) relève, elle d'une position un peu différente, et peut-être plus ambigüe (Cf. ci-dessous).

La seconde *« surprise »*, celle-là possiblement attendue, est l'expression de deux ensembles *« d'organisations »* implicatives, partiellement dissociées et partiellement associées.

Le premier ensemble, globalement -à l'exception de deux variables-, plutôt conforme aux hypothèses, lie de manière implicative, les capacités opératoires *(Op1)*aux capacités spatiales *(Spa)* -ceci dans une position à la fois *charnière* entre les deux ensemble, mais surtout *centrale* concernant cet ensemble-, ainsi que les capacités de compréhension du langage écrit *(Lec)*, les capacités de mémoire à court terme *(Cmém)* comme les capacités de copie *(CSc)* et d'organisation de la copie *(CTy)* de la Figure de Rey, ainsi que des capacités d'intelligence générale *(FG)*, *« fluide »,* au sens de R. B. Cattell, et, dans une position non attendue (Cf. ci-dessus), les capacités assimilées à la *dépendance/indépendance* à l'égard du champ *(ID)*.

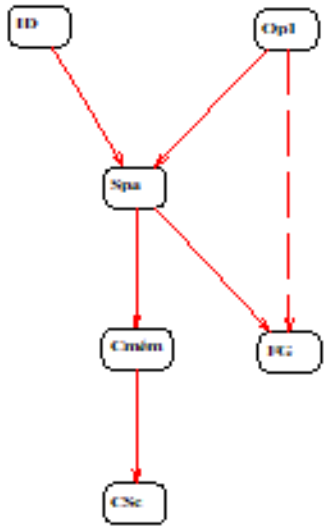
Le second ensemble, plus inattendu, regroupe une variable dite *« témoin »*, les capacités relevant de l'épreuve de *Fluidité* (flexibilité mentale) de R. B. Cattell *(Flui)*, l'épreuve de codage *(Cod)*, l'épreuve de clôture perceptive[13]*(Fla)*, ainsi que l'épreuve de compréhension du langage écrit *(Lec)*[14] -située en position charnière entre les deux ensembles considérés- et deux autres variables, dites *« témoins »* la vitesse graphomotrice *(VM)* et la capacité de discrimination perceptive *(Disc)*. Y aurait-il, là, deux configurations renvoyant, à deux typologies possibles du développement cognitif (?) de l'enfant, pris au sens large ? L'une serait plus centrée sur le langage et la communication, le *« sens »,* culturellement défini, l'expression d'une certaine fantaisie et d'une pensée plus *« divergente »*. L'autre serait davantage centrée sur le développement du raisonnement, d'une pensée plus *« convergente »* et de capacités plus spatiales. Et, même, pour certains élèves au spectre cognitif large, il pourrait exister une configuration d'ensemble étendue, regroupant ces deux configurations partielles … Cela rejoindrait la

[13] Permettant, rappelons-le, de traiter de la *signification* de figures rendues difficilement perceptibles.
[14] Ici, dans une position *centrale* globale, et ceci, peut-être davantage que la dimension spatial *(Spa)*.

littérature à ce sujet, concernant les *« hauts potentiels »* associant convergence de l'esprit et … créativité !

Par ailleurs, nous voyons bien, également, que des variables supposées ne pas s'exprimer dans le modèle, telles les variables *« témoins »* citées ci-dessus et initialement considérées comme telles, s'expriment bien, pour certaines d'entre elles, non pas dans la configuration stricte et réduite du développement du raisonnement dit *« opératoire »*, mais, au moins à la marge de celui-ci, dans le second ensemble. Les autres variables témoins n'apparaissent, elles, pas du tout, tel qu'attendu. Il en est ainsi de*(IP)*, *(Adap)*, et *(I/E)*.

Enfin, la troisième *« surprise »*, certes moins inattendue que la première, relève des liaisons, à indices d'implication élevés, et spécifiques de la dimension opératoire avec des dimensions figuratives/schématiques : celle référant au *« support »* spatiale (*Spa)* et celle référant à la MLT[15] (*Cmém*), alors que cette liaison implicative parait moindre, avec le *« support »* compréhension du langage écrit *(Lec)*. Cela apparait particulièrement lorsque l'on retient seulement la valeur de l'indice d'implication de 0.80 (Cf. Graphe 3 ci-dessous)



Graphe 3 – Ensemble des relations *implicative*s après traitement par l'ASI (CHIC 4.2) à la valeur de l'indice d'implication de 0.80

A cette valeur d'indice, traduisant une plus grande qualité d'implication (Gras, 2018, p. 20), le rôle *« central »* et même *« charnière »*, de la dimension référant aux capacités de représentation spatiale *(Spa)* comme *« support »* au développement *opératoire(Op1)* du sujet, est particulièrement mise en évidence par le graphe, alors que le rôle de la dimension de compréhension du langage écrit (*Lec)* n'apparaît plus. Autre singularité de cette présentation, cette dimension des capacités spatiales semble, elle-même, dépendre à la fois du facteur général d'intelligence et des capacités de mémorisation à court terme *(Cmém)* de la Figure de Rey, ces dernières dépendant, cela paraît *« logique »*, des capacités à reproduire cette même figure *(CSc)*.

[15] La mémoire à court terme, ceci, cependant, sous la forme d'une relation transitive.

Du point de vue méthodologique, deux raisons explicatives de cette prévalence *fonctionnelle* apparente, dans le développement *opératoire(Op1)*, de la capacité spatiale *(Spa)* sur la capacité verbale *(Lec)* peuvent être avancées. L'une ou l'autre de ces raisons, ou les deux conjointement, pourraient expliquer cette prévalence renvoyant à un cheminement organisationnel dominant du développement opératoire, passant, ici, par la capacité spatiale.

Tout d'abord, il y aurait le caractère plus schématique et figuratif des exercices de l'épreuve opératoire (*Op1*), présentant, souvent, tout comme l'épreuve spatiale, à côté de consignes verbales relativement concises, de nombreux schémas et dessins.

Mais il pourrait y avoir aussi[16] les effets de biais liés à un échantillonnage non nécessairement représentatif de la population générale des élèves de 6ème alors considérée, effet de bais, tendant à exclure toute possibilité de généralisation des résultats obtenus à l'ensemble de la population de ces élèves de classe de 6ème.

En effet, si la population retenue a été, par construction, voulue comme représentative de la population du secteur considéré, ce secteur relève, lui, d'une représentation plus rurale, dans une région, par ailleurs *« réputée »* moins performante sur des questions de maîtrise du langage écrit. Dans une logique de développement opératoire tenant à deux *« supports »* l'un verbale *(Lec)* l'autre spatial-représentatif *(Spa)*, ou, de développements selon des cheminements alternatifs, voire compensatoires de l'un par rapport à l'autre, si un *support* fait plus ou moins défaut, le *biais* de représentativité de la population d'élèves considérés a des conséquences, sur les résultats obtenus. Ceci rend, alors, toute interprétation *généraliste* abusive et donc erronée, validant indirectement une approche plus *différentielle*.

Ce biais de différence de performances concernant les deux supports est effectivement bien présent, quoique modéré, puisque la moyenne des performances spatiales (*Spa*, m=0,51) est clairement supérieure[17] à celle des performances en termes de compréhension verbale écrite (*Lec*, m=0,45). Ceci devient, d'ailleurs, davantage évident lorsque l'on traite des évolutions dynamiques du développement *opératoire*, de sous-populations issues du groupe de référence, sous-populations maitrisant conjointement les deux dimensions support, ou n'en maîtrisant qu'une seule[18]

Et donc, le rôle, ici *« central »,* de la dimension spatiale pourrait tenir, simplement, à des capacités moindres dans le domaine de la compréhension du langage écrit *(Lec)*, et, comme évoqué ci-dessus, à des capacités spatiales *(Spa),* en moyenne, *« supérieures »*.

3.2 La question de méthodologie posée par les résultats obtenus : point de vue d'ensemble

Outre le fait, possible, que des effets d'une construction *« biaisée »* de l'épreuve opératoire, ou de *« biais »* d'échantillonnage, amènent, dans une généralisation hautement abusive, à privilégier le poids d'un des supports *(Spa)* nécessaire à la construction opératoire du sujet, au regard de l'autre support *(Lec)*, un autre aspect, tenant

[16] Et nous persistons et insistons, là, sur ce fait, et ceci malgré l'invitation faite, durant l'intervention dans ASI 10, à ne pas considérer cet argument comme possiblement pertinent.

[17] Ceci dans une échelle *relative* au groupe et non pas référencé à la réussite théorique maximale, comme évoqué à un autre moment dans ce texte.

[18] Pellois, 1986, p. 390 et suivantes, et résultats actuellement non publiés.

à la méthodologie, interfère avec des interprétations qui pourraient être trop hâtivement formulées en termes *« implicatifs »*.

Cet aspect relève du *sens* que l'on peut donner au *« jeu de mécano »* à la base de l'expression du raisonnement logique achevé et à son développement.

Dans une compréhension *constructiviste* du développement cognitif, en particulier du développement logique[19], aboutissant au raisonnement, dit scientifique, *hypothético-déductif,* la logique interprétative voudrait que l'on considère, dans la recherche de possibles *« causes-à-effets »* en cascade, comme pressenti dans le modèle précédemment présenté, la maîtrise de certaines capacités comme *nécessaire* à l'expression de la pensée opératoire mobilisée dans le raisonnement. Ces capacités de raisonnement *impliqueraient*, alors, au sens strict, des capacités initialement développées (acquises), et/ou disponibles (plus ou moins innées), ceci, certes, dans un emboîtement rigoureux mais *« psychologiquement discutable »* (Pellois, 2013). Or, le sens de l'*« implication »* observée peut recouvrir d'autres interprétations (Cf., par exemple, Pellois, 2007).

La relation implicative peut ne pas traduire une possible relation de *« cause-à- effet »*, ou simplement de *« nécessité »*, sous deux aspects.

Le premier aspect relève de l'absence de rapport, et donc de cohérence à accorder au raisonnement mis en œuvre, raisonnement reliant des dimensions pour lesquels il n'existe aucun lien entre eux, aucune relation *fonctionnelle* évidente, décrite, directe ou plus indirecte, même lointaine. Cela peut être, par exemple, à la petite enfance, mettre en rapport : le fait de communiquer oralement et l'aisance à marcher en équilibre sur une poutre, ou, maîtriser les relations sociales entre pairs et savoir rouler à bicyclette. Pourtant, il peut être facilement observé que les capacités à maîtriser l'un et l'autre s'échelonnent dans le temps. Et, là où l'analyse implicative relèvera, pour une population d'âge hétérogène, une relation possiblement implicative *« évidente »*, l'interprétation des résultats, jusqu'à preuve du contraire[20], ne pourra conclure qu'à une simple *« précédence »* de l'acquisition d'une compétence par rapport à l'autre, et ceci même en présence d'une corrélation non négligeable entre les dimensions considérées. Il est clair que ce ne peut être que la consistance théorique, la cohérence *fonctionnelle* évidente ou étayée par d'autres observations ou expériences, qui peuvent permettre de trancher à ce sujet-et toujours sur ce plan *fonctionnel*- entre une interprétation en termes de *précédence* et une interprétation en termes d'*implication*, au sens, a minima, d'une condition nécessaire.

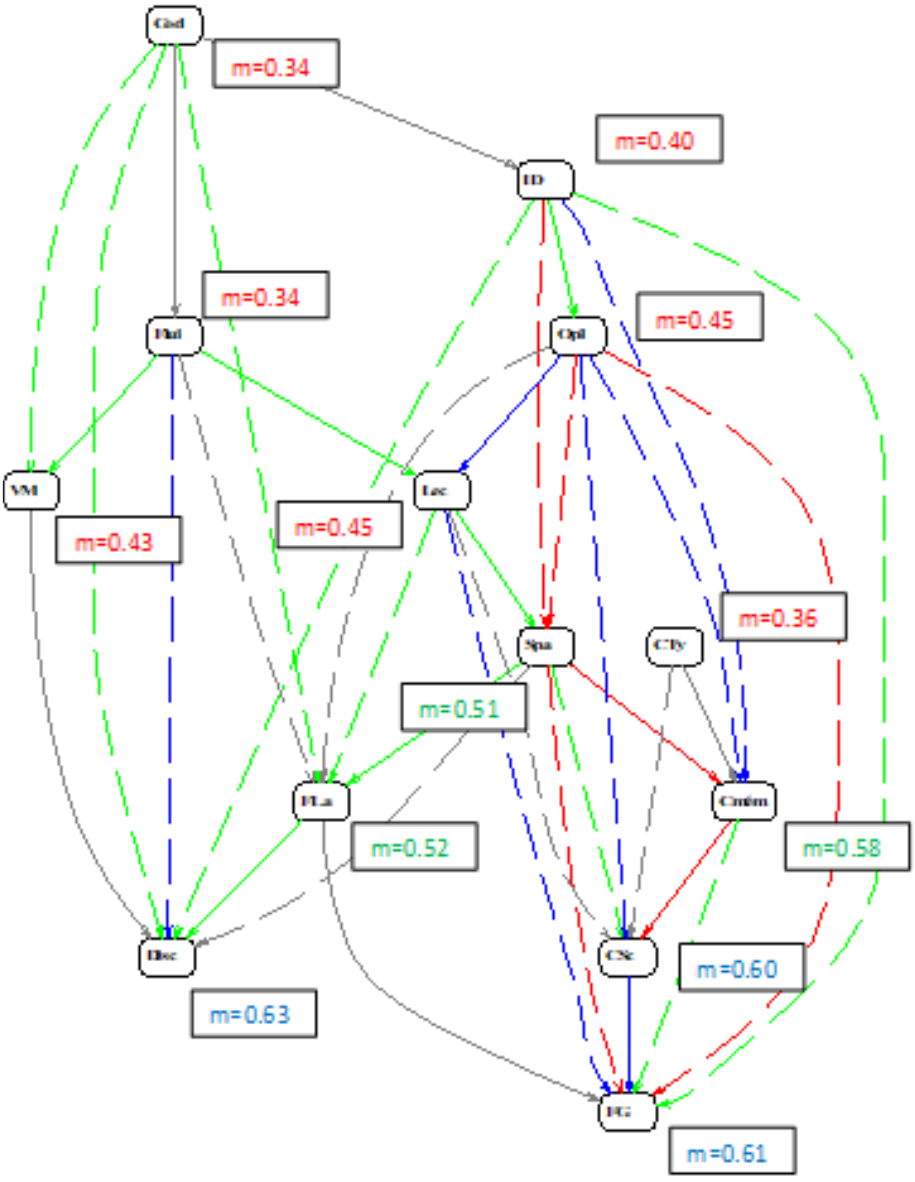
Le second aspect renvoie aux différences observées entre les moyennes des performances aux différentes dimensions retenues, amenant à considérer que des épreuves apparaissent plus *« difficiles »* que d'autres. Dans ce cas encore, une relation de type *implicatif* peut être observée, relevant d'une autre interprétation.

Pour traiter de ce dernier aspect, le graphe 2 a été repris mais en associant aux étiquettes des variables, la moyenne des performances du groupe pour chaque dimension considérée. Cette moyenne de performances apparait encadrée de la façon suivante : par exemple pour l'épreuve *« code »* : (*Cod*), m=0,34

[19] Et ceci même en tenant compte de l'approche différentielle et/ou relevant du courant de psychologie cognitive dit *« du traitement de l'information »*, comme abordé au début de ce texte.

[20] La découverte ultérieure (peu probable ?) d'un possible lien *fonctionnel*, très indirect, entre les dimensions ici évoquées …

Le graphe 3 ci-dessous permet un constat assez *« saisissant »* au regard de l'hypothèse suggérant que la différence entre les moyennes de performances, traiterait aussi bien de rapports de *difficultés* différentes entre les épreuves considérées, que d'une relation dite *« implicative »*. En effet, comme attendu par cette hypothèse, les capacités sensées en *impliquer* d'autres sont caractérisées par des moyennes de performances, dans l'ensemble, plus faibles. Réussir les épreuves associées serait donc plus *« difficile »,* comme, par exemple, pour l'épreuve *« code »* ci-dessus évoquée ((*Cod*), m=0,34). Alors que les capacités considérées comme *nécessaires* à l'expression d'autres capacités sont caractérisées par des moyennes de performances plus élevées, et réussir les épreuves associées serait plus *« facile »,* comme, par exemple en ce qui concerne l'épreuve *« Discrimination perceptive »* ((*Disc*), m=0,63).



Graphe 4 – Ensemble des relations *implicatives* entre variables après traitement par l'ASI avec indication des moyennes de performances pour chaque dimension considérée (CHIC 4.2)

L'objection d'un rapport entre variables dont le sens serait, non pas l'implication, mais une relation de type *« plus difficile/plus facile »* paraît bien exister, là, et telle qu'évoqué antérieurement (Pellois, 2007). Elle serait même confirmée, et ceci de manière plus méthodique, par des analyses plus récentes[21] de plusieurs distributions théoriques différentes de données traitées ensuite par l'ASI.

Pourtant, ici, et du fait des *intuitions* antérieurement évoquées sur cette question, des précautions avaient été prises, ceci afin de limiter l'effet possible de la différence entre les moyennes des performances correspondant aux dimensions considérées. D'une part, les épreuves retenues ont été soigneusement choisies, ceci afin de correspondre au mieux à la tranche d'âge des élèves considérés. D'autre part, le calcul des pourcentages de réussite ne s'est pas fait à partir des scores théoriques minimum ou maximum, mais, pour chaque épreuve, à partir de l'éventail des scores allant de la performance observée la plus faible (affectée du pourcentage 0%, soit 0,00) à la performance la plus forte (affectée du pourcentage 100%, soit 1,00).

Cependant, si le constat global paraît assez clair dans le détail, la question semble plus complexe qu'il n'y paraît en première analyse, et les résultats obtenus après traitement par CHIC, *« résistent »*, au moins pour partie, à cet argument du simple rapport *« facile/difficile »*.

Il y a, tout d'abord, le fait que des rapports *implicatifs* allant, pour une part importante des variables considérées, dans le sens retenu dans le modèle initial, semblent bel et bien exister (notamment entre *Op1, Lec, Spa, Cmém, FLa, FG, CSc*).

Il y a surtout, ensuite, le fait que ces rapports *implicatifs*, conformes au modèle, mais aussi plus *« inattendus »*, apparaissent, et ceci même lorsque les performances moyennes du groupe sont voisines, quasi identiques, voire identiques (*Op1* et *Lec,* m=0,45 et m=0,45, *Spa* et *FLa*, m=0,51 et m=0,52, *CSc* et *FG*, m=0,60 et m=0,61,*Cmém* et *CSc,* m=0,58 et m=0,60, *Cmém* et *FG,* m=0,58 et m=0,61, *Cod* et*Flui,* m=0,34 et m=0,34).

Par ailleurs, des indices d'*implication* peuvent être jugés peu élevés, du fait de cette hypothèse d'interprétation en termes *facile/difficile* venant *« parasiter »,* en quelque sorte, une interprétation implicative, alors que la *« distance »* en termes de moyennes de performances est grande. C'est le cas, par exemple, de (*CTy*) et (*Cmém*) qui présentent un écart important entre leurs moyennes (m=0,36 et m=0,58) écart associé à une intensité d'implication *« réduite »* (valeur entre 0.65 et 0.70). Et des exigences en termes d'intensité d'implication peuvent être, même, ici, jugées *« élevées »* (valeur supérieure à 0.80) alors que la *« distance »* entre les moyennes de performances, est, elle, nettement plus réduite, comme pour (*Spa*) et (*Cmém*) (m=0,51 et m=0,58).

Enfin, d'autres différences de moyennes de performances sont très importantes entre des variables, ici, dites *« témoins »*, sans qu'il soit relevé de relations implicatives entre elles, ni avec d'autres dimensions, ce qui, confirme aussi, à l'occasion, le modèle initial, par l'absence de relations constatées. Il en est ainsi, notamment, des capacités d'adaptation (*Adap*) et de l'indice de persévération (*IP*), dont les moyennes de performances sont respectivement de 0,44 et 0,70.

Il n'en reste pas moins vrai que des différences de moyennes de performances rendent cette interprétation de relations *implicatives*, en particulier dans le sens fonctionnel

[21] Résultats, correspondant à des distributions formelles, non publiés.

précédemment évoqué, possiblement, complexe à établir. Ces différences de moyennes constatées pourraient renvoyer, ici, les interprétations avancées en termes *implicatifs,* à des rapports de *difficultés* plus ou moins grandes entre les épreuves considérées, ou à une interprétation davantage *« métissée »*.

Concernant l'interprétation plus ou moins *« parasite »*, traitant de la *facilité* ou de la *difficulté*, il pourrait en être ainsi pour des variables comme *Flui* (0,34) et *Disc* (0,63). Se pourrait-il, également que cette interprétation traite, mais plus ou moins *partiellement*, des positions non *« attendues »* de *Cod* (0,34) et *ID* (0,40) au regard de *Op1* (0,45), *Lec* (0,45), *Spa* (0,51) ?

Enfin, et malgré l'objection des différences de niveau de difficulté, la dimension du facteur général d'intelligence (*FG*) présente bien la position *« singulière »* attendue dans le modèle, et bien dans des rapports effectivement *implicatifs* qui ne laissent guère de doute, les moyennes de performances étant voisines au moins de celles des variables concernées les plus proches *(CSc, Cmém)* (m=0,61 et 0,60, 0,58). Il paraît clair, en particulier, que dans le premier ensemble (dénommé précédemment raisonnement convergent) la réussite en ce domaine *(FG)* serait *« nécessaire »* à la réussite concernant de nombreuses autres dimensions cognitives. Cependant, elle ne semblerait pas être la seule dans ce rapport *« singulier »*. La dimension de discrimination perceptive (*Disc*), quoique sous une forme plus modeste, l'est au regard du second ensemble, considéré comme plus verbal/divergent.

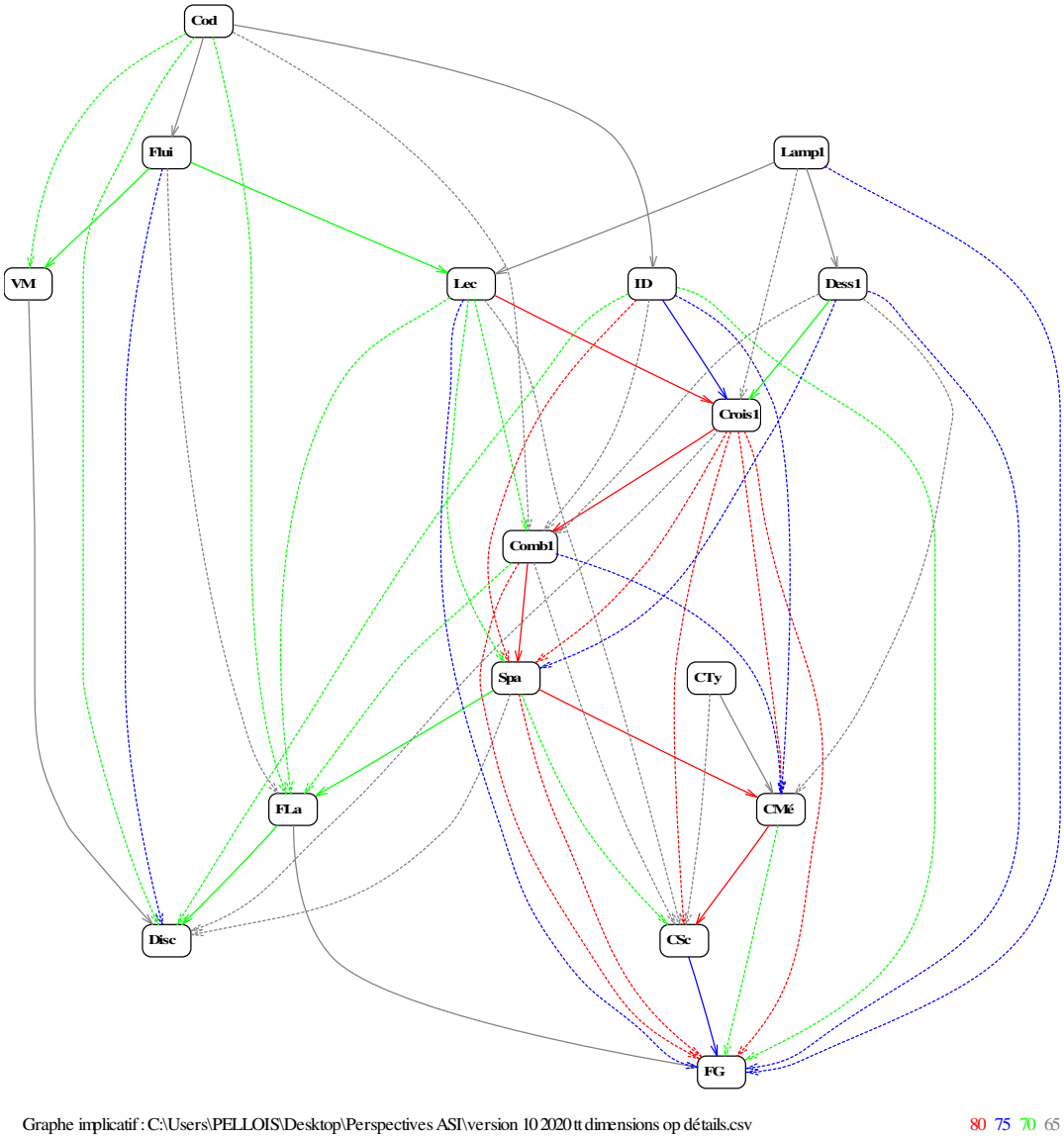
3.3 Les résultats obtenus lorsque la dimension opératoire est subdivisée en ses différentes sous-parties

Les résultats précédemment présentés, peuvent-ils être davantage éclairés lorsque l'on introduit dans l'analyse implicative réalisée par CHIC, en lieu et place du résultat global (*Op1*), les résultats de ses sous-parties (*Crois1, Com1, Lamp1, Dess1)* ?

Certes, cette manière de faire n'est pas sans poser quelques délicates questions méthodologiques telles qu'évoquées dans la précédente publication sur ce sujet (Pellois, 2019). Cependant, en prenant quelques précautions dans les interprétations, cette position est apparue comme acceptable, et, peut-être, paradoxalement, du fait même que les différentes sous-parties considérées, traitent aussi de sous-stades différents.

En effet, si les réussites sont partiellement sensibles, à l'expression des différences interindividuelles *qualitatives* d'efficacité dans l'épreuve, les réussites sont aussi plus ou moins sensibles à la fourchette d'âge des élèves considérés, et donc aux stades du développement, momentanément acquis, ce qui différencie, aussi, mais seulement partiellement, les sous-parties de l'épreuve opératoire. Certaines relèvent strictement du stade dit *« intermédiaire »* (*Crois1*), d'autres s'étalent du stade concret au stade formel A, comme *(Lamp1)*, d'autres traitent également du stade formel B comme *(Com1)* et *(Dess1)*. Il est clair, notamment aux âges ici considérés, en classe de 6ème, que les réussites sont abondantes, au stade concret, et, communes, au stade dit *« intermédiaire »*. Elles sont assez courantes au stade formel A pour l'épreuve de combinatoire *« Jeu de lettres » (Comb1)*, mais plutôt rares pour l'épreuve *« Lampes » (Lamp1)*, nettement plus rares au stade formel B pour l'épreuve *« Dessins » (Dess1),* qui ne possède pas d'items formel A, et rarissime, à ce stade formel B, pour l'épreuve de combinatoire *« Jeu de lettres » (Comb1)*.

Les résultats du traitement par CHIC 4.2 (Cf. ci-dessous) montrent que l'ensemble des interactions renvoie bien à un possible jeu différentiel des relations implicatives en termes qualitatifs traitant des contenus, mais aussi, et peut-être surtout, à un jeu différentiel de relations implicatives de l'ordre du *« hiérarchique »*, en rapport avec les stades considérés, tel que mis en évidence antérieurement (Pellois, 2019).



Graphe 5 – Ensemble des relations *implicatives* entre variables après traitement par l'ASI (CHIC 4.2), la variable (*Op1*) étant subdivisée en ses sous-parties (*Lamp1, Dess1, Crois1, Comb1*)

a) Observations plus approfondies, sur le plan global

Tout d'abord, nous retrouvons bien la position des deux ensembles initialement considérés, ensembles à la fois partiellement disjoints et partiellement liés. Mais le fait de dissocier la dimension opératoire en ses sous-parties amène à compléter et affiner les constats qui ont été précédemment faits sur les deux ensembles considérés.

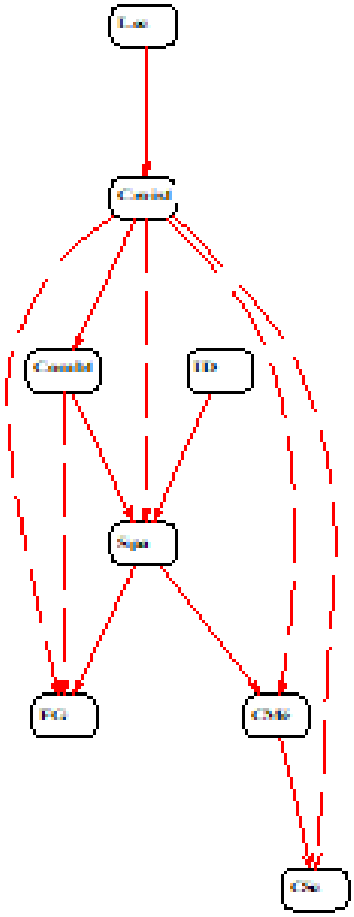
Ensuite, tel qu'observé antérieurement (Pellois, 2019), nous retrouvons bien la *« hiérarchisation »,* au moins partielle, en termes de stades, évoquées ci-dessus. En effet, ces sous-parties de l'épreuve opératoire *(Op1)*, sont en rapports implicatifs entre elles et

ceci de manière partiellement conforme au modèle théorique du développement de la pensée opératoire piagétienne. Ensuite, les positions des sous-parties *(Lamp1)* et *(Dess1)* ne sont pas les mêmes que celles des sous-parties *(Comb1)* et *(Crois1)*. Si les deux premières sont clairement au sommet du premier ensemble (raisonnement convergent abouti), et en relations implicatives avec d'autres dimensions de ce premier ensemble, les deux autres paraissent avoir des rapports plus complexes tant avec le premier qu'avec le second ensemble (verbal/divergent). Plus précisément, une de ces deux dimensions, celle portant sur la combinatoire (*comb1*), apparait en position *« charnière »* à côté des dimensions *« support »* portant sur la compréhension du langage écrit *(Lec)* et la dimension spatiale/représentative *(Spa)*, L'autre, portant sur la maîtrise de l'intersection de classes *(Crois1),* se place, elle, en position à la fois *« centrale »* et *« charnière ».*

Le fait que ces deux dimensions apparaissent plus *« charnières »* ou plus *« centrales »*, serait totalement conforme aux différentes problématiques *qualitatives* du développement des enfants, pris dans une perspective plus large, non seulement momentanément, et particulièrement à cet âge du développement de la pensée opératoire piagétienne (Longeot, 1969, 1978), mais aussi, peut-être plus durablement (Laroche, 1966) et sur un plan que l'on peut considérer, dans un sens strictement piagétien, comme *« hiérarchique ».* Ces développements différentiels s'appliqueraient également à l'ensemble des dimensions cognitives qui y sont, ou y seraient associées. La position de dépendance des deux autres sous-parties *(Lamp1)* et *(Dess1)* au regard d'un premier niveau de combinatoire *(Comb1)* et de l'intersection de classe *(Crois1)* paraît alors *« logique »*, au sens propre du terme, sur un plan généraliste[22]. Mais il l'est aussi, au sens figuré, au regard des développements différentiels *« qualitatifs »* entre les enfants, de la pensée opératoire, ceci en fonction des contenus traités (Cf. Pellois, 2019). Et, si différence *« hiérarchique »* il y a bel et bien, au moins momentanément, elle tient, ici, à ce que de rares élèves accèdent précocement, au stade formel A, voire au stade formel B, ceci, bien évidemment, associé à des intensités d'implication peu élevées, tenant vraisemblablement, au nombre réduit d'enfants, ici, concernés.

b) Observations sur un plan strictement opératoire et avec une exigence d'intensité d'implication *« élevée »*

[22] Début de *« combinaisons »*, logique propositionnelle *« élémentaire ».*



Graphe 6 – Ensemble des relations *implicatives* après traitement par l'ASI (CHIC 4.2) à l'indice d'implication de 0.80, la variable (*Op1*) étant subdivisée en ses sous-parties (*Lamp1, Dess1, Crois1, Comb1*)

Seuls apparaissent, ici, dans des positions *« centrales »,* les sous-parties *« Croisements »* (*Crois1*) et *« Jeu de lettres » (Com1),* tout comme la dimension *« support »* spatiale/représentative *(Spa)*. Mais, à côté de ces dimensions réapparait bien la dimension *« support »* compréhension du langage écrit *(Lec)*. Cependant, sa position, quelque peu *inattendue*, éclaire alors d'une autre manière, le statut que pourrait (aussi ?) avoir cette variable dont la position la ferait dépendre de l'ensemble des autres comme expression d'une capacité *« sophistiquée »* de l'intelligence *« cristallisée »* selon l'appellation de R. B. Cattell. Et ceci suggère, alors, une autre interprétation possible. A ce niveau de l'enseignement, si la dimension *opératoire (Op1)* se développe pour traiter du raisonnement hypothéticodéductif, il semblerait bien que des sous-parties de ce développement, correspondant aux variables *(Crois1)* et *(Comb1)*, au moins, permettent aussi des améliorations d'efficacité dans la compréhension verbale à l'écrit.

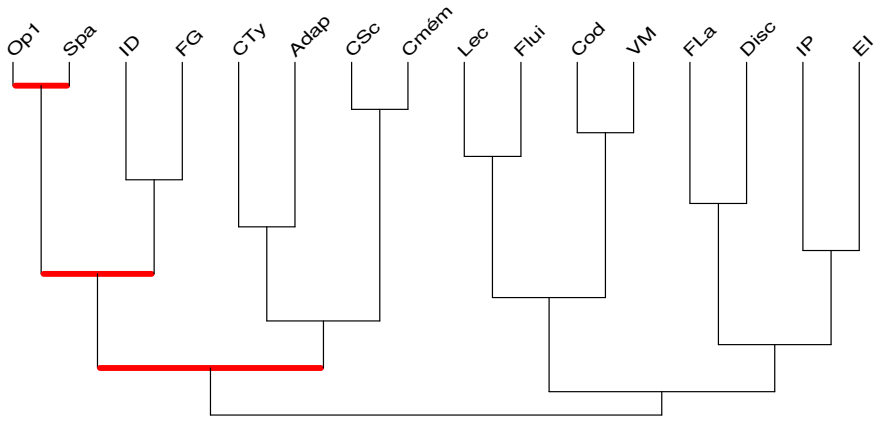
Enfin, à cet indice d'implication de 0.80, une *« chaîne implicative »* particulièrement marquée, avec relations transitives, attire l'attention, au regard du modèle *dynamique* initialement posé. Les capacités de mémorisation (*Cmém*) à court terme (MCT) de la figure de Rey impliquent les capacités de sa copie (*CSc*), l'activité efficace de la copie favorisant la mémorisation au sens, par exemple, de Dehaene (Dehaene, 2018, p. 246). Puis, les capacités de mémorisation à court terme *(Cmém)* et portant sur le facteur général

de l'intelligence (*FG*) *« favoriseraient »*, à leur tour, l'expression des capacités spatiales/représentatives (*Spa*). Ensuite, ces dernières *favoriseraient* l'expression de la combinatoire *(Comb1)*, au moins aux stades concrets et intermédiaires, et accessoirement l'expression de la dépendance indépendance à l'égard du champ perceptif *(ID)*. Puis, les capacités de *combinatoire(Comb1) favoriseraient* à leur tour -ainsi que, directement ou indirectement, les capacités spatiales (*Spa*)-l'expression des capacités de maîtrise de l'intersection de classe *(Crois1)*. Ceci ferait entrer les *« sujets »* dans une étape précédant la maîtrise de la *logique propositionnelle*, utile au raisonnement hypothético-déductif, et que les élèves des classes de 6ème ici considérées utiliseraient in fine, pour permettre que se développent, aussi, les capacités de compréhension du langage écrit *(Lec)*. Mais tout cela n'aurait véritablement de sens que confronté aux évolutions *effectives* du développement opératoire des élèves, évolutions observées à partir d'une nouvelle prise d'information[23] sur cette dimension en classe de 5ème. Maintenant, qu'en est-il au regard des autres traitements opérés par CHIC ? Y-a-t-il une cohérence d'ensemble incluant le *« jeu » implicatif* ci-dessus interprété ?

3.4 Les résultats sous l'angle de l'analyse des *similarités* et de l'analyse *cohésitive* sous CHIC 7b[24]

Remarque importante, ici, toutes les variables apparaissent, et pas seulement celles présentant des relations implicatives aux intensités d'implication précédemment considérées.

a) Analyse selon l'arbre hiérarchique des *similarités*

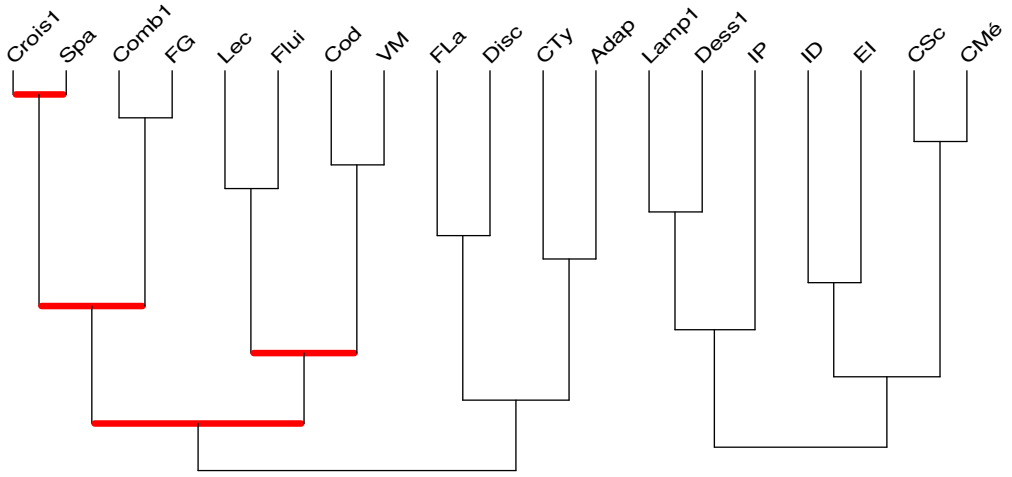


Arbre des similarités : C:\Users\PELLOIS\Desktop\Perspectives ASI\version 10 2020 tt dimensions.csv

Graphe 7 – Arbre hiérarchique des similarités, la dimension opératoire étant considérée globalement (vers. CHIC 7b)

[23] Données traitées en 2021, non publiées.

[24] Cependant, sur ces différents points, les résultats ne sont pas tout à fait identiques sous CHIC 4.2.



Arbre des similarités : C:\Users\PELLOIS\Desktop\Perspectives ASI\version 10 2020 tt dimensions op détails.csv

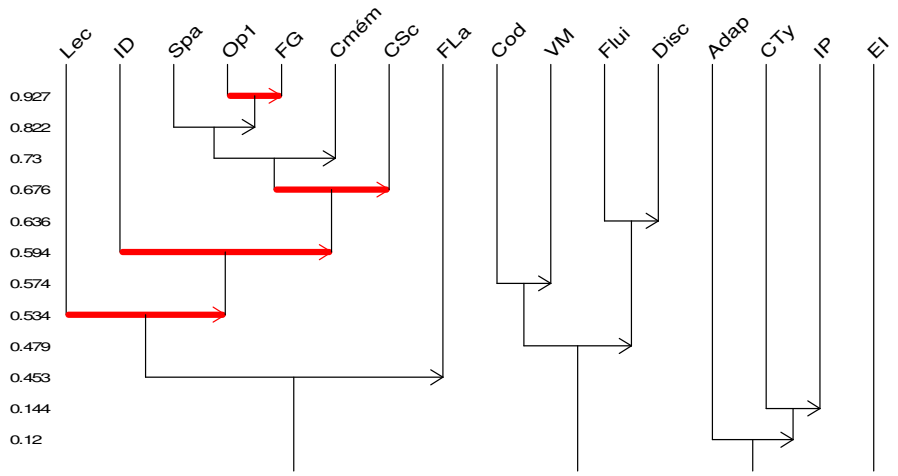
Graphe 8 – Arbre hiérarchique des similarités, la dimension opératoire étant considérée en ses sous-parties (vers. CHIC 7b)

Lorsque l'épreuve opératoire est prise dans son ensemble, apparaissent bien les deux ensembles considérés précédemment. L'un tient bien à l'expression de la pensée dite *« convergente »* et du raisonnement, associant *(Op1), (Spa), (ID)* et *(FG)*, ce qui est conforme à la littérature sur ces questions (Huteau, 1980a, 1980b, Longeot 1969, 1978, Carroll 1980, 1993). L'autre porte davantage sur la compréhension du langage écrit *(Lec)*, la communication et le sens de ce qui est *« signifié »* (*FLa*)), les capacités de discrimination perceptive *(Disc)*, de codage *(Cod)* et même la dimension de la personnalité extraversion/introversion *vs* extraversion, et une forme de pensée dite *« divergente » (Flui)*. Par ailleurs, ces deux ensembles sont liés.

Lorsque l'épreuve opératoire est subdivisée en ses sous-parties, l'interprétation ne peut plus être la même. Deux groupes de variables contenant chacun deux sous-parties de l'épreuve opératoire suggèrent cette fois-ci de manière non plus transversale mais longitudinale, deux étapes du développement opératoire *(Crois1)* et *(Comb1)* d'un côté, *(Dess1)* et *(Lamp1)* de l'autre. Ce second groupe associe à ces deux dernières sous-parties de l'épreuve opératoire, d'autres dimensions *« témoins »* ou non. Il en est ainsi de l'épreuve dépendance/indépendance à l'égard du champ perceptif *(ID)* et de personnalité introversion/extraversion *(E/I)*, et aussi de l'indice de persévération *(IP)*, des capacités de copie et de mémorisation à court terme*(CSc), (Cmém)*.

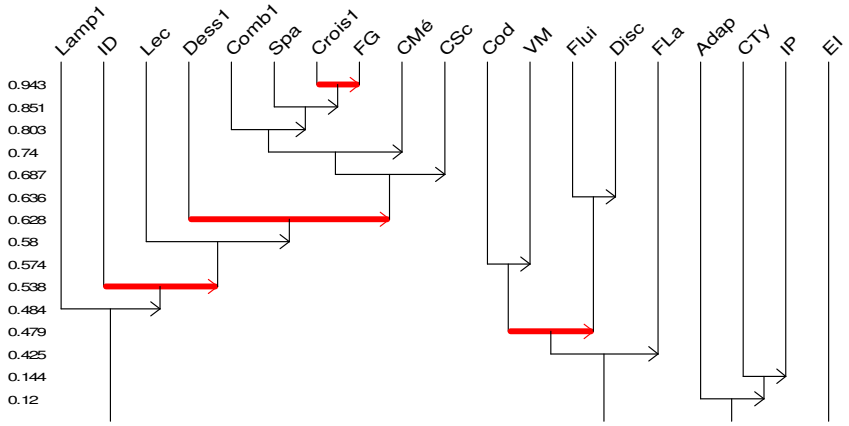
Mais, dans le premier groupe, se retrouve bien deux sous-groupes de variables associées. L'un rassemble les deux autres sous-parties du domaine opératoire, la dimension spatiale et la dimension du facteur général d'intelligence (*Crois1, Comb1, FG, Spa*), rappelant le premier ensemble initialement repéré. L'autre traite du langage et de la flexibilité/créativité, notamment (*Lec, Flui, Cod*), rappelant le second ensemble, associant aussi la variable *« témoin »(VM)*. Ce qui apparait comme nouveau, ici, c'est un troisième sous-groupe, associé aux deux précédents, traitant de qualités perceptives renvoyant, ou non, au sens des objets représentés (*Fla, Disc, CTy, Adap*).

b) Analyse cohésitive



Arbre cohésitif : C:\Users\PELLOIS\Desktop\Perspectives ASI\version 10 2020 tt dimensions.csv

Graphe 7 – Schéma *cohésitif,* la dimension opératoire étant prise globalement (vers. CHIC 4.2)



Arbre cohésitif : C:\Users\PELLOIS\Desktop\Perspectives ASI\version 10 2020 tt dimensions op détails.csv

Graphe 8 – Schéma *cohésitif,* la dimension opératoire étant considérée en ses sous-parties (vers. CHIC 4.2)

La logique de ce qui est représenté en termes *« cohésitif »* serait-elle, ici, sensiblement différente ? Et dans la complexité interprétative inhérente au traitement *cohésitif*, relèverait-elle, non pas, comme en ce qui concerne la similarité, d'une approche à caractère *« typologique »,* mais d'une approche davantage *« dynamique »* ?

L'observation des graphes obtenus fait penser, effectivement, qu'il y aurait bien une logique (dynamique ?) du développement opératoire associant *« fortement »(Op1)* au facteur général *« d'intelligence »* (*FG*) (Graphe 7), en particulier par le biais de la relation *cohésitive* entre une sous-partie de la dimension opératoire (*Crois1)* et (*FG*) (Graphe 8). Cette relation entrainerait des effets *« en cascade »* reliant d'autres dimensions, ceci au moyen de relations *cohésitives* emboîtées. Comme l'indique le second graphe *cohésitif*. En effet, derrière la relation *cohésitive (Crois1)*/(*FG*) apparait la relation à la dimension spatiale (*Spa*) puis la relation de celle-ci à la combinatoire (*Comb1*) dont *« la relation de relation »* composée avec les précédentes *« dépendrait »* de (*Cmém*) et (*Csc*). Puis la dimension opératoire figurative/schématique (*Dess1*) *« dépendrait »* de ce nouvel ensemble plus vaste *« de relations de relations »*. Viendraient ensuite les dimensions

(*Lec*) et (*ID*), confirmant la position, moins conforme, ou non conforme au modèle du développement théorique et hypothétique du fait du *« rôles »* attribués à ces deux dimensions dans le modèle, rôle différant de celui retrouvé dans les résultats.

Enfin, au sommet de ce jeu de *« lego »* de *« relations de relations »*, viendrait celle de la relation de la dimension (*Lamp1*), traitant d'une logique propositionnelle complète (et complètement acquise ?), avec l'ensemble des autres dimensions, ci-dessus évoquées. L'emboîtement considéré traiterait bien d'une logique d'investissement que, au regard des hypothèses et des conceptions théoriques, l'on peut considérer comme *« dynamique »* de la pensée opératoire piagétienne, dite, plus généralement, *convergente*, ceci à côté d'un autre emboîtement, d'une autre logique possible, celle autour de la dimension *fluidité* (*Flui*), et relevant d'une forme plus *divergente* de la pensée.

4 Que conclure ?

4.1 Sur le fond de la recherche considérée ici

a) Les éléments tendant à confirmer la modélisation proposée

Au sens, notamment de l'analyse implicative, les résultats obtenus, au-delà des aspects qualifiés de *« surprises »* ou d'*« inattendus »* semblent bien avoir du sens et, assez largement, le sens que la modélisation permettait d'attendre (Cf. point 3.1). *(Op1)* implique bien : *(Lec)* et *(Spa)* qui impliquent à leur tour, pour *(Lec)*, *(FLa)*, et pour *(Spa), (CMém)* et *(FLa), (CMém)* impliquant, ensuite, *(Csc)*. Puis, nombre de ces variables impliquent directement ou indirectement *(FG)*. Certes, deux variables fonctionnelles ne sont pas, ici, dans des positions attendues *(Cod)* et *(ID)* -mais il y aurait, sur ce résultat (Cf. point 3.2), une possible interférence tenant aux moyennes de performances, ici, nettement plus faibles-. Enfin, les variables dites *« témoins »* sont absentes, comme attendu, sauf *(Disc)*. Aussi, même s'il existe, comme évoqué dans l'interprétation des résultats obtenus, quelques divergences, l'ensemble constitué par les résultats obtenus peut être considéré comme remarquable, cela dans un contexte de prise d'informations *« fragile »*, fragilité inhérente à toute prise d'information en sciences humaines, et dans le domaine de la psychologie en particulier. En effet, l'information recueillie, est toujours suspecte d'aléas, de jeux de fluctuations, fluctuations qui, par l'effet de *« désordre »* possible[25], de *« flou »* de la mesure, amènerait à ne rien observer comme tendance, plutôt que d'observer des résultats considérés, ici, comme *« significatifs »*. Pas d'intensités d'implication très exigeantes, certes, mais des indices d'implication (0.80, 0.75, 0.70, et même 0.65) qui sont considérés, là, et compte-tenu des remarques ci-dessus, et précédemment formulées (point 3.1), dans une tolérance acceptée d'un nombre de contre-exemples non négligeable, comme des indications *« raisonnables »* de tendances effectives. De plus, au-delà des indices d'implication observés, ce qui parle clairement c'est la cohérence de l'ensemble des résultats et, donc, des interprétations qui peuvent être émises, liant analyse *implicative*, analyse en termes de *similarités*, et analyse *cohésitive*.

[25] Plusieurs performances différentes, de type test/retest, peuvent être recueillies sur la même dimension concernant le même élève, en fonction du moment, de la standardisation rigoureuse des consignes, de l'effet de groupe, de l'état de concentration très labile des enfants, etc. (Cf., par exemple, Huteau, Lautrey, 1999, p. 129-130, Pellois, 2003, 266-267 et 275-281)

L'expression du développement *opératoire* dépendrait bien de l'expression de plusieurs dimensions, à la fois des variables dites *« support »,* mais également, ceci cependant de manière plus variable, des dimensions traitant de différentes modalités perceptives, de représentation et/ou de mémorisation, et de ce qui pourrait relever, peut-être, d'une dimension *« intuitive »*, au sens des neurosciences, plus fondamentale, à savoir le facteur *G* des conceptions *factorialistes* de l'intelligence. Mais l'analyse implicative montre que cette confirmation au moins partielle du modèle, relève d'une compréhension possiblement *restreinte* du développement et qu'il s'esquisse une autre forme de développement parallèle (associée ?) ou plus large dont la théorie opératoire de l'intelligence ne rendrait pas nécessairement compte, amenant à une vision renouvelée de cette question du développement de l'enfant, puis de l'adolescent, à l'adulte.

b) La singularité de l'épreuve de facteur *G*, ébauche d'analyse interne

Il est clair que la dimension d'un facteur général d'intelligence (*FG*), non nécessairement structuré sous la forme, culturellement acquise, de l'expression *opératoire* piagétienne, mais en rapport *« étroit »* en termes *implicatif*, de *similarité* et de relation *cohésitive* avec elle, semble bien jouer un rôle majeur dans l'initialisation puis le développement culturel/cognitif, même élargi au sens ci-dessus proposé (Cf. point a)). Ce rôle semble bel et bien exister vis-à-vis du premier ensemble, *spatial/représentatif* et *opératoire/convergent*, du raisonnement scientifique appliqué aux données et aux expériences. Mais ce rôle, certes, moins attendu, semble exister aussi vis-à-vis du second ensemble, plus *verbal/divergent*. Et, au travers de ces ensembles, elle joue, aux âges considérés ici, sur une organisation dynamique plus complète et plus large du développement de la personne.

Mais, de quoi s'agit-il lorsque l'on évoque cette dimension dite d'un *« facteur général de l'intelligence » (FG)* ? La réponse à cette question peut, éventuellement, passer par une autre question. Que se passe-t-il, concrètement, lorsque l'on tente de résoudre les exercices proposés par cette épreuve dite de facteur G ? La démarche envisagée est, alors, éminemment périlleuse, tout psychologue le sait, *« l'introspection »* sans des éléments objectivés pour vérifier sa valeur explicative scientifiquement validée pose problème. En neurosciences, celle-ci peut être avantageusement couplée à ce que donne l'analyse des résultats de l'imagerie médicale. Cependant, même cela n'est pas sans poser question au plan méthodologique (par exemple, Joliot, in Houdé et Coll., 2002, 2ème partie chapitre 6, et Mazoyer, Joliot, en particulier p. 202, in Houdé et Coll., 2002).

Concernant la tentative très directe, tentée ci-dessous, de saisir comment la pensée a procédé, de manière dynamique, pour résoudre un des exercices considérés dans l'épreuve D48 traitant de ce facteur général *(FG)*, il y aurait plusieurs difficultés conjointes à surmonter. Il y a, d'une part, le risque de rationalisation a postériori et de conscientisation et explicitation d'une démarche de résolution de problème, sous une forme *« séquentielle et successive »*, qui peut *fonctionner* de manière implicite et a-consciente, et sous une forme, plus ou moins formalisée, *« parallèles et simultanés »,* qui plus est, sans rationalité évidente, voire de manière plus ou moins chaotique ... D'autre part, il y a le risque de généraliser une telle démarche recomposée par une personne donnée traitant d'un problème particulier, à toute démarche, ceci quelque que soit la personne considérée et le problème posé. A ce sujet chacun pourra juger ici *« sur pièce »*, à partir des propositions de l'auteur de ces lignes, de l'illusion possible d'une

généralisation à toute personne s'attelant à résoudre l'exercice considéré ! Passons, pourtant, à cet exercice très concret.

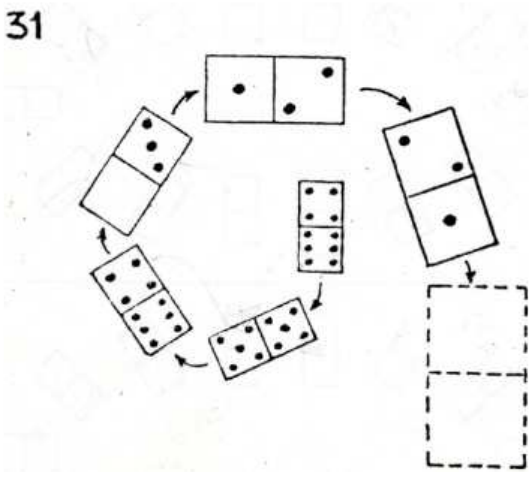


Figure 1 –Exemple d'exercice de l'épreuve D 48

L'épreuve D 48 propose des séries de dominos, tel que, par exemple, l'exercice 31 représenté, sur la figure 1 ci-dessus. Il apparaît, tout d'abord, que les *« objets »* ordinaires d'un jeu connu, les dominos, doivent être mentalement *transformés*. Il convient à la fois de traiter d'*« objets »* intérieurs à celui-ci, *« objets »* correspondant aux différentes parties de ces dominos, à savoir les deux cases du domino dont le nombre de points va de 0 à 6, et d'*« objets »* extérieur à chaque domino, correspondant à plusieurs dominos considérés conjointement. Le percept/concept (?) domino est alors *« cassé »*, et supplanté par un traitement d'*« objets »* non totalement et directement *« présents »* mais retenus comme tel et sur lesquels l'attention du sujet se focalise alors.

Il y a, ensuite, et à partir de là, le fait, plus ou moins spontanément ressenti, pour la personne qui résout l'exercice, de ne pas se sentir trop contraint par ce qui est perçu, et donc libre de traiter l'information *perceptive* disponible, comme elle l'entend, mais ceci dans le cadre des règles posées dans cette épreuve du D 48. La personne se donne donc le pouvoir de manipuler cette information *(ID ?)* dans cette liberté partielle, mais cette fois-ci, sur un plan r*eprésentatif (Spa ?), (Fla ?)* plus que directement *perceptif*. Encore faut-il percevoir exactement *(Csc ?) (Disc ?)*, au départ, l'exactitude des informations à traiter. Cela passe donc par la compréhension claire et exacte des consignes *(Lec)* permettant de saisir de quoi il s'agit, ceci en vue de répondre correctement à ce qui est demandé.

Mais, si, à partir de ce qui est perçu, il y a bien, désorganisation, puis réorganisation, représentative (déstructuration/restructuration) active des *« objets »* considérés, cela peut-il être accompagné de l'expression d'un langage intériorisé (soliloque), comme pour *(Lec)*, ou de l'évocation d'images mentales comme pour *(Spa)* ? Deux formes de traitement de l'information seraient alors possibles. Que ce soit sous une forme ou sous une autre, ou les deux conjointement, pour l'exercice 31 ci-dessus, peuvent être examinées les cases de dominos, en ignorant la case qui suit immédiatement celle observée, ce qui permet de voir que 4 points dans la première case sont suivis de 5 points dans la troisième, puis de 6 dans la cinquième. Ces observations permettraient d'en déduire la règle suivante : le nombre de points de ces cases prises en sautant à chaque fois une case, croissent d'un point à chaque fois. Ensuite, le fait de constater qu'il n'y a, dans

une case de domino, jamais aucune valeur supérieure à 6, permet de comprendre pourquoi la case suivante concernée, présente 0 point. L'observation des cases qui suivent, confirmant la *règle* initialement trouvée, amène à contrôler et vérifier que la série croissante se continue avec des valeurs de 1, puis 2 points, ce qui permet, in fine, de déduire que la case suivante dont il convient de trouver la valeur, doit (pensée convergente),au regard de la règle émise, contenir 3 points. La comparaison des autres cases non encore prises en compte amène, en procédant de la même façon, à trouver la seconde *règle*, celle d'une série décroissante, et la valeur de la dernière case, soit 0 point. Notons que les opérations mentales, apparemment très simples, nécessaires à la résolution de l'exercice ici considéré, et ceci comme reconnu dans l'approche « *composantielle* » (Huteau, Lautrey, 1999, p. 212 et suivantes), se succèdent selon un schéma de résolution présent à l'esprit de manière latente, et que celui qui le résout garde en tête, tout en traitant chaque étape de raisonnement *(Cmém ?).* Par ailleurs ce schéma *« cadre »* (directeur ?) et les opérations en jeu, sont susceptibles de changer d'un exercice à un autre. Arrivé au bout de cette description, rappelons que, comme évoqué au début de cette section, il s'agit ici, *seulement*, de la perception d'une résolution propre à l'auteur de ces lignes ... Il resterait à savoir comment quelqu'un d'autre expliquerait, lui, sa propre manière de résoudre l'exercice. Et, in fine, la manière *« rationnelle »* et explicite décrite est-elle bien celle qui a, effectivement, été utilisée, et ceci de manière plus implicite, par l'auteur pour effectivement trouver la solution ? …

Plus fondamentalement, ce que l'on peut supposer, est, que la résolution des problèmes proposés passe par l'exercice *conjoint et articulé* de plusieurs dimensions, évoquées ci-dessus inscrite en italique et entre parenthèses, dimensions inhérentes au fonctionnement cognitif, et retenues dans le modèle hypothétique initial. Ceci pourrait-il rendre compte du pouvoir *« générique »* de cette capacité mobilisée, plus ou moins *« spontanément »* et dans la vie courante, sans passer nécessairement à un raisonnement logique sophistiqué, tel que c'est le cas dans la pensée opératoire piagétienne, à savoir le raisonnement scientifique ? Mais cet exercice conjoint peut-il rendre compte du fait que cette capacité pourrait être, en elle-même, *nécessaire* et donc quelque part, au moins partiellement, *« antérieure »* à l'expression des autres dimensions mobilisées ? Et, dans ce cas, qu'est-ce que serait, alors, le *« cœur »* de ce qui permet en même temps, cette dynamique efficace de résolution de problème et son pouvoir de permettre tant le développement que l'expression de beaucoup d'autres dimensions de la sphère *cognitive*, qui plus est, prise ici au sens large ? S'agirait-il de l'attention soutenue, des capacités *exécutives* et de contrôle de la tâche, tel que soutenu par les neuroscience ? S'agirait-il : du résultat de l'expression d'un langage intérieur, de la liberté prise d'évoquer des *« objets »* mentaux, et donc, de la capacité à mobiliser des *représentations* spatiales, ou de la capacité à transformer, manipuler ces *« objets »* ? Toutes ces questions, encore au jour d'aujourd'hui, restent, au moins partiellement, ouvertes ![26]

[26] En particulier, peut-on dire que cette dimension paraît si fondamentale qu'elle interviendrait dans tout type de conduites humaines, notamment à vocation cognitive ? Certains contestent, au moins pour partie, cette logique hiérarchique, cette universalité *« basique »* possible renvoyant à des schémas culturels (très tôt installés ?), notamment, occidentaux opposés aux schémas orientaux, larges mais non généralisables à toute culture, tel Richard E. Nisbett (Nisbett, R. E. (2003), *The Geography of Thought,* Ed. Free Press) et défendent la prépondérance des facteurs environnementaux sur les facteurs héréditaires dans l'expression des performances intellectuelles (Nisbett, R. E. (2009), *Intelligence and How to Get It. Why schools and cultures count*, Ed. W. W. Norton & Company).

c) Les éléments *« divergents »* : synthèse rapide et perspectives

Des résultats inattendus existent bien, des aspects méthodologiques peuvent les expliquer partiellement, ceci par un jeu possible d'artefact, comme les différences de moyennes de performances aux épreuves choisies, mais aussi du fait que focalisé sur une approche essentiellement en termes de pensée convergente (l'intelligence *« fluide »*), le modèle initial ne permet pas de reconnaître, à cet âge considéré, et comme évoqué au point a), d'autres formes prises par le développement, tel celui référant davantage au langage et à la pensée dite *« divergente »*. Il en est ainsi d'une forme de *créativité*, abordée incidemment, du fait de sa position de variable *« témoin »*, par la *fluidité* ou *flexibilité mentale (Flui)*[27].

Ces résultats nous disent aussi quelque chose de l'ordre des *différences*, voire des bifurcations de perspectives de développements possibles, comme cette forme plus engagée sur le langage, sur la créativité, la pensée divergente, à côté de formes plus *« classiques »* de la pensée convergente en rapport avec le domaine des mathématiques, des sciences et de la technologie. Cependant, même en ce domaine, des formes différentes d'expression apparaissent, plus verbales/numériques ou plus spatiales/représentatives laissant présager que la population considérée, à l'interne du groupe, n'est pas nécessairement homogène, mais présente une forme d'hétérogénéité possible entre des sous-groupes, tant en termes de typologies exprimées que de dynamiques plurielles possibles du développement, notamment cognitif. Seules des études de sous-groupes caractéristiques d'évolutions opératoires différentes[28] permettraient de rendre compte, a minima, de ces différences-là.

4.2 Au plan des aspects méthodologiques, quelques questions centrales ?

Il y a tout d'abord, à souligner, l'importance ici réaffirmée de l'utilisation de l'ASI dans une position *« confirmatoire »*, en particulier au regard d'un modèle conceptuel conçu en tenant compte de recherches portant sur de nombreuses décennies. Ensuite, apparaît fondamentale la question des « *choix* », tout d'abord, ceux appliqués à la mesure et, ici, le *choix* des variables considérées comme pertinentes, mais aussi, ceux portant sur les *choix* des *modalités* et des *options* du traitement, appliqué aux données. Il y a également, la délicatesse, voire la *« fragilité »* ou la relativité des *interprétations* émises, dépendant en grande partie des *choix* effectués, des modalités et options de traitement par l'ASI, notamment ceux portant ***sur les indices retenus au regard de l'exigence en termes d'intensité d'implication***. En effet, ce problème du *choix*, ici, des indices d'implication considérés, conditionnent clairement les *interprétations* faites. Il en est ainsi de : l'absence de la relation implicative, si l'indice traitant de l'intensité d'implication choisi correspond à une forte exigence, de la linéarité univoque possible des relations *implicatives* si l'indice traitant de l'intensité d'implication choisi correspond à une faible exigence, et de la pluralité des *« cheminements » implicatifs évoqués*, pour des choix d'indices intermédiaires se situant entre ces deux précédentes exigences. Ce problème, très concret évoqué ci-dessus, n'en rejoindrait-il, alors, pas un autre, très général ? Il s'agirait de *choix*, plus larges, que le chercheur fait, de *« modèles explicatifs »*, par

[27] Cet aspect de pensée *divergente*, de la créativité, appelé *fluidité* ou *flexibilité*, ne doit pas être confondu avec celui de la pensée *convergente* exprimée dans l'intelligence dite *« fluide »*, par opposition à l'intelligence *« cristallisée »* selon R. B. Cattell.

[28] Résultats également obtenus en 2021, et non publiés.

exemple univoques ou pluriels, déterminant un cadre mathématique orienté du traitement statistique. C'est ce que l'on retrouve, par exemple, en analyse factorielle, au sujet de la question de la rotation des axes (Lautrey, 1999, p. 288 et suivantes), et pour laquelle il n'y a pas, formellement, de solution unique. Ceci ne rejoindrait-il pas, cette fois-ci sur un plan plus formel (Changeux, Connes, 1989, p.208 et suivantes), la question, abordée par Gödel, de la *décidabilité* ou de *l'indécidabilité,* dans le cadre des mathématiques, ceci en fonction d'une position initiale de choix ?

Pour conclure, contribuer avec rigueur, et sans complexe, à la réflexion méthodologique tout en apportant des éléments de connaissance, dûment questionnés et critiqués, n'est-il pas une manière de contribuer, même modestement, à l'expression d'une science de *qualité*[29] ? Il s'agirait d'une *science* qui serait faite, bien évidemment, de résultats acquis, d'une compréhension claire de phénomènes et des connaissances qui en découlent, mais aussi de regards critiques, comme d'incertitudes, de tâtonnements, de multiplicités d'approches, ainsi que de variabilités, de fluctuations, et de *différences* prises en tant que telles et non de points de vue abusivement généralisés. Ces généralisations potentiellement abusives, seraient susceptibles de relever, alors, sinon d'un possible *« scientisme »*[30] très (trop ?) étroitement défini, et devenu anachronique, d'une *science*, faite d'affirmations qui seraient assénées (Par exemple, Dehaene, 2018, p. 246 et suivantes, 315 et suivante ?), associées à des *« approximations »* de références, *science*, devenues, alors, fort discutable. Il en serait également ainsi des interprétations, parfois délicates, de résultats de traitements statistiques, tels ceux proposés par l'ASI, si ceux-ci n'étaient pas soumis à un questionnement critique, y compris méthodologique, rigoureux, (tel que l'évoque aussi Dehaene, 2018, p. 87) assurant leur qualité véritablement *scientifique*. Cette rigueur, ne permettrait-elle pas de sortir, tout comme celle opposant inné et acquis, d'une opposition, implicitement, quasi idéologique, entre *« généralistes »* et *« différentialistes »* ? Il semble que cette opposition puisse être ici, très concrètement, renvoyée à une articulation possible des deux positions, ceci en lien avec un travail méthodologique, mené à propos de l'outil de traitement statistique utilisé, travail, bien entendu conjoint au travail sur le fond. A ce sujet, la maîtrise progressive du raisonnement scientifique, de l'enfance à l'âge adulte, constituerait un développement parmi d'autres possibles, plus ou moins conjoints, plus ou moins associés, ou carrément indépendants, voire, possiblement, en opposition, et donc quelque part contradictoires, comme semble l'évoquer, elles-mêmes, les neurosciences (Par exemple, Dehaene, 2007, p. 280 et suivantes).

Un modèle initial, tel que présenté dans ce texte crée du sens, ici très largement conceptuel et hypothétique. Du traitement par l'ASI de données issues de réalités effectives, émerge un autre sens *partiellement* confirmatoire mais aussi plus large, plus complexe et, possiblement, différentiel.

Références

[1] Bachelard, G. (1987), *Le nouvel esprit scientifique*, PUF, Paris, (17ème Edition).

[29] Ouvrage en préparation.

[30] « *Opinion philosophique, de la fin du XIXe s., qui affirme que la science nous fait connaître la totalité des choses qui existent et que cette connaissance suffit à satisfaire toutes les aspirations humaines. (C'est une forme du positivisme.)* » https://www.larousse.fr/dictionnaires/francais/scientisme/71476

[2] Carroll, J. B. (1980), *Individual difference relations in psychometric and experimental cognitive tasks,* final report n° 163, LL Thurston, psychometric laboraty university of North Carolina.

[3] Carroll, J. B. (1993), *Human Cognitive Abilities*, Cambridge University Press, Cambridge.

[4] Cattell, R. B. (1971), *Abilities: Their structure, growth, and action*, Houghton Mifflin, New York.

[5] Changeux, J. P., Connes, A. (1989), *Matière à pensée*, Odile Jacob, Paris.

[6] Damasio, A. R. (1995), L'erreur de Descartes, la raison des émotions, Odile Jacob, Paris.

[7] Dehaene, S. (1997), *La bosse des maths*, Odile Jacob, Coll. « Sciences », Paris.

[8] Dehaene, S. (2007), *Les neurones de la lecture*, Odile Jacob, Coll. « Sciences », Paris.

[9] Dehaene, S. (2014), *Le code de la conscience*, Odile Jacob, Coll. « Sciences », Paris.

[10] Dehaene, S. (2018), *Apprendre !,* Odile Jacob, Coll. « Sciences », Paris.

[11] Gras, R. (2018), *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*, Cépaduès-éditions, Toulouse.

[12] Houdé, O. (1995), *Rationalité développement et inhibition. Un nouveau cadre d'analyse*. PUF, Coll. « Psychologie et sciences de la pensée », Paris.

[13] Huteau, M. (1980a), Dépendance indépendance à l'égard du champ et développement de la pensée opératoire, *Archives de psychologie*, XLVIII, 184.

[14] Huteau, M. (1980b), Style cognitif et pensée opératoire, *Bulletin de psychologie*, XXXIII, 345, Mai-Juin.

[15] Huteau, M. (1982), *Cognition et personnalité. La dépendance indépendance à l'égard du champ*, document ronéoté, rapport d'activité 81-82, laboratoire de psychologie différentielle, Paris, services de recherches I.N.O.P., 20-23.

[16] Huteau, M. (1985), *Les conceptions cognitives de la personnalité*, Paris, P.U.F., Coll. « Psychologie d'aujourd'hui ».

[17] Huteau, M. et Lautrey, J. (1999), *Evaluer l'intelligence. Psychométrie cognitive*, PUF, Paris.

[18] Inhelder, B., Piaget, J. (1970), *Essai de logique opératoire*, PUF, Coll. « Bibliothèque philosophique contemporaine », Paris (1ère Ed. 1959).

[19] Inhelder, B., Cellérier, G., et al. (1992), *Le cheminement des découvertes de l'enfant*, Delachaux & Niestle.

[20] Jobard, G. (2002), La lecture de mots, in Houdé, O., B. Mazoyer, et N. Tzourio-Mazoyer, *Cerveau et psychologie. Introduction à l'imagerie cérébrale anatomique et fonctionnelle*, PUF, Paris, 500-520.

[21] Joliot, M. (2002), L'intégration multimodale, in Houdé, O., Mazoyer, B., Tzourio-Mazoyer, N., Cerveau et psychologie. *Introduction à l'imagerie cérébrale anatomique et fonctionnelle*, PUF, Coll. « Premier cycle », Paris, 317-336.

[22] Klein, E. (2016), *Le pays qu'habitait Albert Einstein*, ACTES SUD ? Essai, Coll. « BABEL ».

[23] Laroche, J.L. (1966), *Diagnostic des structures opératoires d'adultes de niveau CEP*, C.P.T.R., Metz bibliothèque, 664, 1-83.

[24] 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.

[25] Lautrey, J. (1990), Esquisse d'un modèle pluraliste du développement cognitif, in Reuchlin, M., Lautrey, J., Marendaz, C., Ohlmann, T., (sous la direction de), *Cognition : l'individuel et l'universel*, PUF, Paris, Coll. « Psychologie d'aujourd'hui », 185-216.

[26] Longeot, F. (1969), *Psychologie différentielle et théorie opératoire de l'intelligence*, Dunod, Coll. « Sciences du comportement », Paris.

[27] Longeot, F. (1978), *Les stades opératoires de Piaget et les facteurs de l'intelligence*, PUF de Grenoble, Paris.

[28] Mazoyer, B., Joliot, M. (2002), La neuroimagerie cognitive : bases physiologiques, in Houdé, O., Mazoyer, B., Tzourio-Mazoyer, N., *Cerveau et psychologie. Introduction à l'imagerie cérébrale anatomique et fonctionnelle*, PUF, Coll. « Premier cycle », Paris, 187-203.

[29] Pellois, C. (1986), *Caractéristiques du sujet psychologique et dynamisme du développement opératoire : le passage du stade des opérations concrètes au stade des opérations formelles*, thèse de doctorat, Université de Caen.

[30] Pellois, C. (2003), *Enseignement et formation : le développement de la personne, tome 2 : Le complexe*, L'Harmattan, Coll. Recherches et Innovations, Formation pour l'enseignant.

[31] Pellois, C. (2007), L'utilisation des modèles structuraux afin d'aborder la complexité du réel en sciences humaines, *Mathématiques et Sciences humaines*, 45ème année, 177(1), 53-85.

[32] 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.

[33] Pellois, C. (2019), Mesure de l'expression et du développement de « l'intelligence » : observation de données et des résultats de traitement par l'ASI, in Régnier, Gras, Couturier, Bodin, *Analyse Statistique Implicative. Cadre théorique en relation étroite et au service de multiples disciplines*. Actes du 10ème colloque d'Analyse Statistique Implicative, Ed. Université Bourgogne Franche-Comté, Besançon, 134-167.

[34] Piaget, J. (1972), *Essais de logique opératoire*, Dunod, Paris, 1ère édition 1949.

[35] Richard, J.-F. (1990), *Les activités mentales, comprendre, raisonner, trouver des solutions*, Armand Colin, Paris.

[36] Von Bertalanffy, L. (1980), *Théorie générale des systèmes*, Dunod, Paris, traduction française, 1973, Nouveau tirage revu et corrigé, 1ère Ed. 1968.

AFFORDANCE DE TWITTER POUR LES USAGES SOCIO-PROFESSIONNELS PAR LES ENSEIGNANTS

Tolga Tekin[1], Jean-Claude Régnier[2]

AFFORDANCE OF TWITTER FOR SOCIO-PROFESSIONAL USES BY TEACHERS

RÉSUMÉ

Bien que les réseaux sociaux n'aient pas été créés initialement à des fins éducatives, ils suscitent un intérêt pour le monde de l'éducation. Les environnements fournis par des réseaux sociaux tels que Facebook et Twitter prennent une importance dans le quotidien des enseignants. Compte tenu de cette réalité, il semble important de se pencher sur les caractéristiques propres de ces outils afin de mieux les comprendre et d'explorer leur potentiel dans l'usage socioprofessionnel des enseignants. En ce sens, cette étude examine comment les affordances de Twitter sont utilisées par les enseignants à des fins professionnelles et comment ceux-ci connaissent les différents éléments de Twitter. Le concept d'affordance permet d'étudier et de comprendre les relations entre Twitter et ses utilisateurs. Nous avons appuyé notre étude sur les analyses qualitatives et quantitatives. Notre échantillon de recherche se compose des enseignants de l'école primaire, du collège, du lycée et aussi de l'enseignement supérieur, ayant utilisé Twitter à des fins professionnelles. Twitter est un outil de réseaux sociaux qui permet de construire un réseau socioprofessionnel entre des personnes qui exercent dans un même groupe professionnel. Les enseignants peuvent profiter de possibilités d'échange, de partage d'information, d'entraide et mettre en œuvre des activités de collaboration. Nous rappelons que cette étude s'inscrit dans la continuité de nos précédents travaux sur l'affordance des réseaux sociaux. Dans ces études précédentes (ASI-9 et ASI-10), nous avons étudié l'impact des affordances de réseaux sociaux dans l'enseignement.

Mots-clés : *affordance, Twitter, réseaux sociaux, médias sociaux, environnement socio-professionnel*

ABSTRACT

Although social media networks were not originally created for educational purposes, they are of interest to the world of education. The environments provided by social media such as Facebook and Twitter are becoming important in the daily life of teachers. Given this reality, it seems important to look at the specific characteristics of those tools in order to better understand them and explore their potential in the professional use of teachers. In this sense, this study examines how the affordances of Twitter are used by teachers for professional purposes and how they know the different features of Twitter. The concept of affordance provides a way to study and understand the relationships between Twitter and their users. We based our study on both qualitative and quantitative analyses. Our research sample consisted of elementary, middle school, high school, and college teachers who used Twitter for professional purposes. Twitter is a social media tool that allows building a socio-professional network between people who practice in the same professional group. Teachers take advantage of the possibilities of exchange, information sharing, mutual support and collaboration. Besides, this study is a continuation of our previous work on the affordance of social networks. In our previous studies (ASI-9 and ASI-10), we have discussed the impact of social network affordances in teaching.

[1] Lyon 2 - ECP, Tolga.Tekin@univ-lyon2.fr

[2] Professeur des universités émérite. UMR 5191 ICAR http://www.icar.cnrs.fr/membre/jcregnier/, Université Lumière Lyon 2. Professeur invité à l'Université d'État de Tomsk (Russie) – National Research Tomsk State University–jean-claude.regnier@univ-lyon2.fr

Keywords: *affordance, Twitter, social networks, social media, socio-professional environment*

1 Introduction

En plus de nombreux usages différents, les réseaux sociaux offrent un environnement dynamique qui « permet mais aussi médiatise les interactions entre personnes » (Henderson, et al. 2013, p.547). Cet environnement favorise également de créer un pool d'informations et de ressources que les enseignants peuvent utiliser pour développer leur pratique professionnelle (Brown et al., 2012 ; Hanraets et al., 2011), enrichir leurs connaissances, élargir leurs perspectives pédagogiques et être facilement informé des actualités. Les mêmes raisons conduisent les enseignants à utiliser Twitter (Carpenter et al, 2018). En plus, à l'aide de ses propriétés, Twitter permet la création et l'échange de contenu généré par les utilisateurs.

Twitter place également des enseignants dans d'autres formes de relation qui permet de créer un environnement socio-professionnel à travers le monde. Au-delà des murs des lieux formels, les twitteurs/twitteuses[3] sont autonomes et actifs. Ils construisent des communautés de connaissances, ouvertes à tous, de partage et d'entraide (McLoughlin, et Lee, 2007). Les affordances particulières de Twitter créent un axe d'attraction pour des usages à des fins professionnelles. Pour mieux comprendre les revendications exprimées à l'égard des affordances de Twitter chez les enseignants, il est important de comprendre certaines des caractéristiques de cet outil (telles que ses éléments) et ses usages professionnels possibles.

Pour examiner les affordances de Twitter par les enseignants, dans ce travail, nous avons profité du cadre théorique de l'approche « Analyse statistiques implicative » (ASI)[4] développée par Régis Gras (1979) et ses collaborateurs (Gras et al. 1996, 2009, 2013, 2017). Cet ensemble de méthodes permet d'étudier les propriétés des relations entre les variables par l'intermédiaire de représentations graphiques telles que l'arbre de similarité, le graphe implicatif et l'arbre cohésitif. Pour le traitement des données, issues d'un questionnaire, nous avons utilisé programme CHIC[5]. Ce dernier *« 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" »* (Régnier et al., 2012).

2 Le concept d'affordance appliqué à l'étude des réseaux sociaux

Le concept d'affordance a été proposé par Gibson (1979) dans une théorie écologique de la perception, et adoptées par la suite dans d'autres domaines. En effet, les affordances sont des propriétés de l'environnement qui permettent ou offrent des actions possibles de l'être vivant qui interagit avec cet environnement. Ce rapport être vivant-environnement,

[3] Personne qui poste régulièrement des Tweets sur Twitter. https://www.larousse.fr/

[4] « [...] 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. » définissent Gras et Régnier (2009, p12).

[5] Classification Hiérarchique Implicative et Cohésitive

souvent étudié à partir d'artefact techniques, concerne aussi les relations directes entre les êtres vivants.

Cependant, l'affordance n'est plus aujourd'hui simplement comprise dans sa dimension perceptivo-comportementale. Elle concerne l'analyse, d'une part, de ce qui relie un sujet (ici les enseignants) à un objet (ici Twitter) ; et, d'autre part, ce que produit la relation à l'objet. En effet, l'affordance qui – au départ – visait l'étude de relation sujet-objet à partir de perception sensori-motrice, est aujourd'hui comprise dans ses dimensions sémiotiques et socioculturelles telles que le précisent Reed (1988), Norman (1988), Turvey (1992), Stroeffegen (2003), Niveleau (2006), Morgagni (2011) ou encore Simonian (2019a ; 2019b ; 2020). Plus spécifiquement, l'affordance est liée à une projection dans la réalisation d'une action de l'agent et aux propriétés offertes par l'objet dans un environnement socioculturel spécifique. Dit autrement, la potentialité d'un objet est dépendante de ce qui pousse un sujet à agir (son intention) dans le but de réaliser quelque chose (efficience).

Les affordances peuvent être conceptualisées à un haut-niveau (abstrait) ou à un bas-niveau (instrumental) (Bucher et Helmond,2017). Ces affordances de bas-niveau sont celles qui se situent généralement dans la matérialité et le design de l'objet. Elles sont des caractéristiques de la conception de la plateforme, des boutons, des écrans et des hashtags qui donnent des indications sur les actions possibles. Dans le sens gibsonien, les caractéristiques techniques - entendues comme les meubles du paysage numérique – de Twitter permettent certaines actions telles que cliquer, partager ou aimer (Bucher et al, 2017).

Les affordances de haut-niveau sont les types de dynamiques et de conditions permises par les dispositifs techniques. En fait, ces affordances de haut-niveaux des réseaux sociaux sont conditionnées par les propriétés des données, qui introduisent essentiellement de nouvelles possibilités d'interaction et de communication (boyd, 2011. p. 39). Ainsi, les relations (partage des valeurs, des pratiques, etc.) sujet-sujet(s) dans un contexte social créent une affordance de haut-niveau[6] qui caractérise l'affordance « socioculturelle » (Simonian, 2015). Par exemple, les affordances de la plateforme de Twitter peuvent être qualifiées de « socioculturelles » pour traduire les multiples possibilités d'interactions mais aussi de partager, commenter, diffuser des informations, expériences, acquérir des savoirs, entre les membres dépendant des rapports aux autres utilisateurs, autrement dit, « *des rapports aux prescrits et, plus largement, à la culture, aux modes d'organisations et de productions sociales* » (Simonian et al., 2016).

3 Affordances de Twitter

Les affordances de la plateforme de Twitter donnent des possibilités aux enseignants de créer un réseau socioprofessionnel personnel qui constitue un environnement d'échanges, de débats et de confrontations avec le monde ou le réseau personnel sélectionné en fonction des tweets, de lancer des débats à partir des tweets postés en fonction des commentaires, de la connectivité momentanée en fonction d'hashtag, de s'adresser à quelqu'un grâce à la fonction *mentionner* et de réagir à des messages. En effet,

[6] « Les affordances les plus riches et les plus élaborés de l'environnement sont fournies par d'autres animaux et, en ce qui nous concerne, par d'autres personnes […] » (Gibson, 1979, p. 135).

ces affordances de Twitter favorisent un processus social et dynamique qui permet d'améliorer, d'acquérir des savoirs et de développer de nouvelles compétences liées aux tâches de l'enseignant, ainsi, elles aident le perfectionnement et le développement de nouvelles pratiques afin de rendre plus efficient l'apprentissage chez les étudiants/élèves.

En tant qu'environnement social, la plateforme de Twitter propose un environnement permanent d'interactions où les twitteurs/twitteuses peuvent communiquer et interagir successivement de manière synchrone ou asynchrone. Ils peuvent, idéalement, partager des informations, faire des débats et ainsi construire des connaissances grâce aux interactions avec d'autres utilisateurs.[7] Dans cette perspective *« le concept d'affordance socioculturelle se rapproche de la théorie de socioconstructivisme de Vygotski »*, en considérant Twitter comme un artefact socioculturel qui médiatise des capacités cognitives entre des utilisateurs et *« permet de donner une signification au but qu'ils poursuivent »* (Simonian, 2015, p. 43).

Les composantes de Twitter permettent d'étayer les interactions entre utilisateurs. L'affordance socioculturelle (Simonian, 2020) cherche à comprendre pour quelles raisons les utilisateurs interagissent jusqu'à se produisent ces interactions dans une dynamique sémiotique (Morgagni, 2011), reposant sur un cadre d'analyse relationnelle sujet-artefact-environnement. Il s'agit donc d'examiner la relation sujet-artefact *« à partir des relations trans-individuelle liées à l'amplification de la relation inter-individuelle»* (Simondon, 1958 cité par Simonian, 2015, p.41).

Plus spécifiquement, la fonction d'hashtag de Twitter avec un mot-clé en compagnie du symbole dièse « # » est une fonctionnalité qui permet de partager des expériences et de regrouper un contenu particulier. Cette fonction a pour fin de permettre aux utilisateurs de retrouver facilement le contenu et de suivre des sujets qui les intéressent. En même temps, le contenu partagé avec l'aide d'un hashtag permet d'assembler les utilisateurs, sans nécessairement être ami ou « follower », autour d'un sujet discuté. Les utilisateurs peuvent aussi effectuer une recherche sur un sujet en utilisant un hashtag pour trouver une communauté spécifique ou un contenu. Les hashtags permettent d'être visible aux internautes dans le réseau.[8] En outre, la fonction du tagging de Twitter permet de marquer d'autres utilisateurs avec le symbole @, mis devant le nom d'utilisateur. Il permet d'identifier d'autres personnes lors de la publication. Twitter met les mentions (@) dans le tweet automatiquement en répondant aux autres. Cela peut provoquer de longues discussions sur la plateforme. Il est possible aussi de retweeter ou de tweeter un contenu en mentionnant quelqu'un.

4 Résultat des traitements et analyses des données dans le cadre de l'ASI

L'analyse statistique implicative (ASI) est, initialement, une méthode d'analyse de données non symétrique, conçue à l'origine par Régis Gras (2018).

[7] En réponse aux besoins de l'activité de l'apprentissage, les affordances de Twitter créent un mode d'apprentissage dans lequel les utilisateurs (ici les enseignants) construisent leurs connaissances en s'engageant, en discutant, en partageant leur expérience et en la verbalisant.

[8] Si un utilisateur souhaite augmenter sa présence dans le réseau, il suffit rédiger sa biographie avec un hashtag (#teamcp) dans sa biographie avec lequel il veut être visible.

La base de données construites à partir d'une enquête par questionnaire en ligne sur le réseau Twitter est constituée de 40 variables binaires issues des vecteurs-variables correspondant aux questions (annexe 1). Elle transcrit les réponses de 104 individus. Pour l'analyse statistique, nous n'avons pris en considération ici que 16 variables avec le statut de variable principale : ce sont celles qui caractérisent le temps d'usage à titre personnel (V10/011) ou à titre professionnel (V012/013), la maîtrise de différents éléments de Twitter et les finalités. Les autres variables qui caractérisent les individus : sexe, âge, pays, cadre de la classe où ils exercent l'enseignement et ancienneté professionnelle, ont un statut de variable supplémentaire.

Les variables numériques : âge, ancienneté professionnelle, temps d'usage de Twitter (depuis combien de temps faites-vous usage de Twitter ?) ont été transformées en variables qualitatives en recourant à la procédure de partitionnement intégrée au logiciel CHIC 7. Chaque partition optimale est formée de quatre segments (intervalle fermé à gauche et ouvert à droite) et rapportée dans le tableau de l'annexe 1

Les composantes des vecteurs-variables « Maitrise des fonctionnalités de Twitter » (Tableau 1) et « Finalités des usages professionnels de Twitter » (Tableau 2) sont des variables qualitatives ordinales à 5 niveaux qui ont été traitées comme des variables modales prenant les valeurs dans {0 ; 0,25 ; 0,50 ; 0,75 ; 1}.

V015 [Écrire un tweet]
V016 [Faire une liste des différents comptes]
V017 [Retweeter avec un commentaire]
V018 [Ajouter un hashtag dans le tweet (#tweetschool)]
V019 [Faire une mention (@un compte ou les organisations)]
V020 [Tweeter avec des emojis-vidéos-images-gifs]
V021 [Faire un sondage]

Tableau 1 : composantes du vecteur-variable Maitrise des fonctionnalités de Twitter

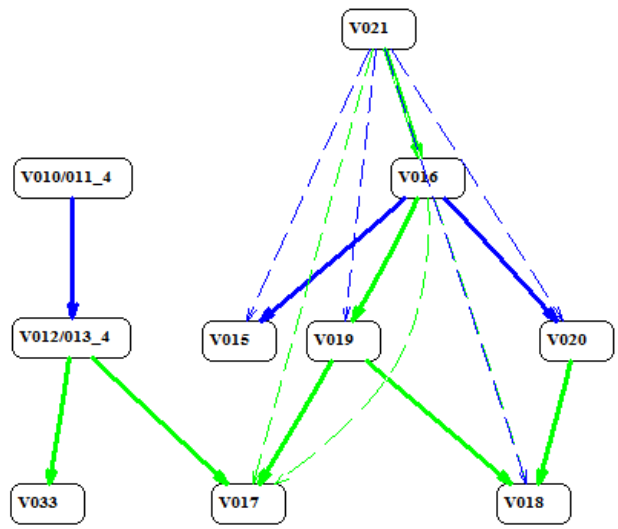
V022 [Partage des ressources]
V023 [Acquisition des ressources]
V024 [Collaboration avec d'autres enseignants]
V025 [Faire des projets avec des étudiants/élèves]
V026 [Communication avec les étudiants]
V027 [Communication avec les parents]
V028 [Activités en classe pour les étudiants/élèves]
V029 [Activités hors classe pour les étudiants]
V030 [Participez aux débats autour de l'éducation]
V031 [Améliorer ou renouveler la pratique]
V032 [Suivre les avis des professionnels/experts]
V033 [Recevoir les dernières nouvelles éducatives]

Tableau 2 : composantes du vecteur-variable Finalités des usages professionnels de Twitter

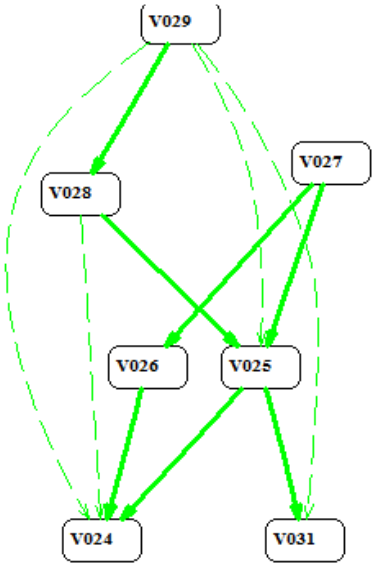
Nous avons procédé à un traitement avec la logiciel CHIC 7 en choisissant le modèle de calcul des intensités d'implication entre variables binaires et des intensités de propension entre les variables modales selon la théorique dite classique avec la loi binomiale (Gras, Régnier, 2017, p.43-61, p.83-88)

Nous avons fixé le niveau des intensités d'implication et de propension à la valeur seuil minimale de 0.95. Ce seuil assure un niveau de confiance que nous jugeons comme satisfaisant pour la validité des relations de quasi-implication entre les variables.

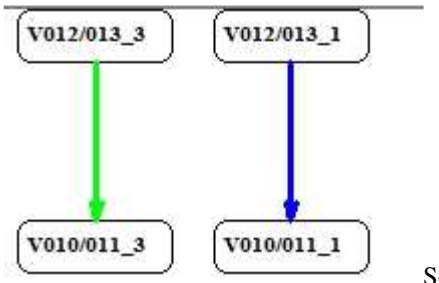
Dans ces conditions nous avons retenu les relations entre les variables qui sont représentées dans le graphe implicatif ci-dessous dans lequel figurent aussi les fermetures transitives (trait en pointillés). Ce graphe implicatif est constitué de quatre sous-réseaux.

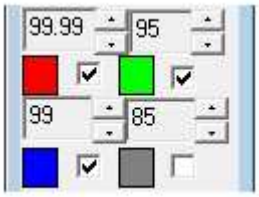


Sous-réseau 1



Sous-réseau 2





Sous-réseau 3 et 4

Figure 1 : Graphe implicatif : réseaux des variables au seuil de 0.95 (vert) et au seuil 0.99 (bleu)

4.1 Rôle de l'ancienneté des usages sur Twitter

Le sous-réseau 1 réunit un groupe de 10 variables qui se décompose de deux sous-groupes. Le premier sous-groupe est constitué par le chemin :

V10/11_4 « Temps d'usage à titre personnel le plus élevé entre 8,5 et 10,9 années »,

↓

V012/013_4 « Temps d'usage à titre professionnel le plus élevé entre 8 et 10,9 années »

↓

V033 « fins professionnelles de l'usage de Twitter [*Recevoir les dernières nouvelles éducatives*] ».

Figure 2

De cette manière, nous observons qu'il y a une tendance parmi les enseignants qui utilisent Twitter depuis le plus longtemps à en faire un usage à des fins professionnelles

et ceci plutôt pour « recevoir les dernières nouvelles éducatives ». Nous interprétons ainsi que ceux qui utilisent Twitter à des fins professionnelles depuis le plus longtemps, l'utilisent plutôt pour être au courant des dernières nouvelles éducatives. L'affordance d'un outil dépend des compétences de l'individu et des objectifs qu'il se fixe. Dans ce cas, il est entendu que si un enseignant utilise Twitter pour être informé des nouveautés sur l'éducation, Twitter peut lui être *affordé* ces dernières nouvelles éducatives. Nous comprenons ici que l'affordance de Twitter traite des possibilités qui établissent un lien entre twitteur/twitteuse et l'outil.

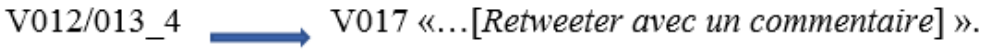


Figure 3

Un autre résultat intéressant est que les enseignants ayant le plus d'ancienneté dans l'usage professionnel de Twitter (parmi le cadre d'activité professionnel des enseignants), ont tendance à ajouter des commentaires lorsqu'ils retweetent. Un Tweet que l'utilisateur partage publiquement avec ses abonnés est appelé un retweet. Le retweet avec commentaire permet alors de citer un texte (ou un tweet) avant de le retweeter. Un retweet commenté permet donc d'apporter un autre contexte au premier tweet. L'affordance socioculturel vise à identifier « *les raisons et conditions favorables* […] *la mise en relation sujet-artefact à partir de la perception du sujet* » (Simonian 2015; 2020). Dans notre cas, nous constatons que les enseignants qui sont sur Twitter depuis le plus longtemps sont plus susceptibles de percevoir la propriété d'ajouter des commentaires lorsqu'ils retweetent. De ce point de vue, le facteur déterminant n'est pas la possibilité de retweeter avec un commentaire, mais la perception de propriété de cet élément de Twitter par l'enseignant.

4.2 Niveau de maitrise de différentes fonctionnalités de Twitter

Deuxième sous-groupe est constitué par les variables concernant le niveau de maitrise de différents éléments de Twitter par les enseignants/enseignantes. Ce que nous cherchons à identifier est dans quelle mesure les enseignants/enseignantes connaissent et savent utiliser les différentes fonctionnalités de Twitter. Les affordances sont des propriétés de l'environnement qui permettent ou offrent des actions possibles de l'être vivant qui interagit dans son environnement. Dans notre cas, ce couplage Twitter-enseignant est étudié à partir des éléments constituant de Twitter. Ce sous-réseau est constitué par les mises en lien entre les 7 variables (Tableau 1) avec des fermetures transitives qui renforcent le caractère implicatif des relations. Ce sous-réseau comporte un chemin principal sur lequel nous portons notre attention.

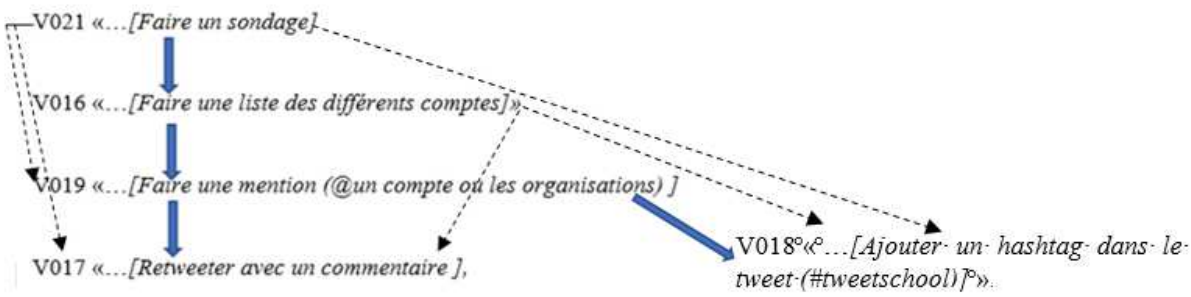


Figure 4

La construction de ce chemin repose principalement sur les contributions d'individus qui ont entre 37 et 44.6 ans, ayant une ancienneté de 2 à 12 ans travaillant hors de la France dans l'enseignement supérieur ou dans d'autres contextes que l'enseignement primaire, le collège ou le lycée. En revanche les femmes et les hommes ont une égale contribution.

Si nous nous centrons sur la variable V016 en utilisant la fonctionnalité du logiciel CHIC 7 : *mode cône*, nous obtenons le graphe suivant :

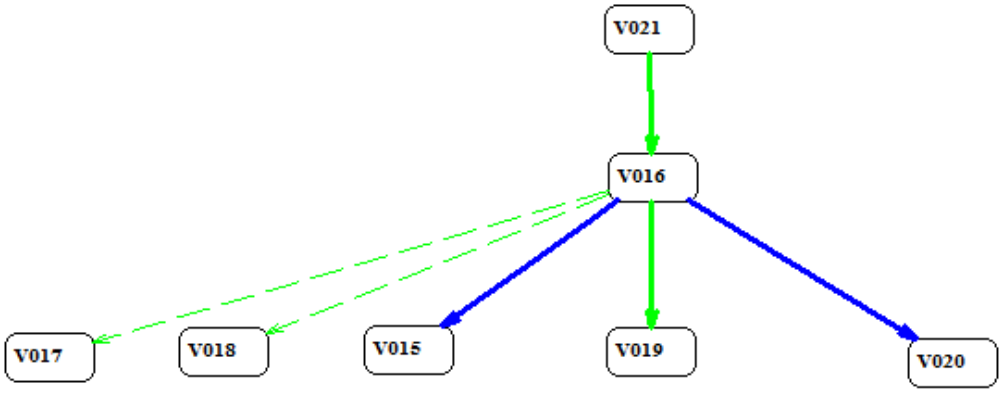


Figure 5

Ainsi la maîtrise de la fonctionnalité V016 «*…[Faire une liste des différents comptes]»* peut être interprétée comme la conséquence de celle V021 «*…[Faire un sondage]* » et conduire à la maitrise de V015 « *…[Écrire un tweet]* », V019 «*…[Faire une mention (@un compte ou les organisations)]* et V20 «*…[Tweeter avec des emojis-vidéos-images-gifs] ».*

Cela montre que les enseignants qui savent mener des enquêtes sur Twitter ont également les connaissances nécessaires pour créer des listes sur différents comptes. De même qu'un enseignant qui peut dresser une liste de différents comptes sur Twitter peut écrire des tweets très facilement et naturellement publier des tweets avec des outils visuels tels qu'avec des vidéos, des images, des émoticons, etc. Nous observons aussi une tendance chez les enseignants ayant les connaissances pour créer des listes avec différents comptes qu'ils peuvent également mentionner un compte ou une personne en tweetant et ajouter un commentaire au retweet.

La relation entre les deux variables V019 et V018 peut être interprétée comme le fait qu'un enseignant qui peut utiliser la fonction @mention en tweetant semble alors en mesure d'ajouter un #hashtag. La connaissance des éléments de Twitter permet d'identifier des affordances perceptives et réelles, en déduisant l'inférence des possibilités d'action de twitteur/twitteuse. En termes de perception/action, la perception des propriétés sémiotique de Twitter par enseignants peut être importantes, car « *une action n'est réalisable que si elle paraît possible à un individu* » (Simonian, 2015, p.22).

4.3 Les buts des usages de Twitter à des fins professionnelles

Le sous-réseau 2 est constitué par les liens entre les variables relatives aux finalités des usages professionnels de Twitter. Le graphe implicatif fait apparaître 7 nœuds qui correspondent à 7 des 12 variables modales présentées dans le tableau 2. D'une certaine manière, l'interprétation de ce sous-réseau nous permet de mieux comprendre ce qui motive les enseignants à utiliser Twitter dans le pratique professionnelle d'enseignant. Rappelons que les données sont issues des réponses à la question : Selon vous, à quelles fins professionnelles Twitter peut être utilisé ? en les confrontant à une liste de 12 usages possibles (Tableau 2) pour lesquels ils étaient invités à formuler un jugement selon l'échelle : 1_jamais, 2_quelquefois, 3_la plupart du temps, 4_souvent, 5_toujours.

Dans ce sous-réseau, nous pouvons identifier un chemin principal qui le structure d'une certaine manière.

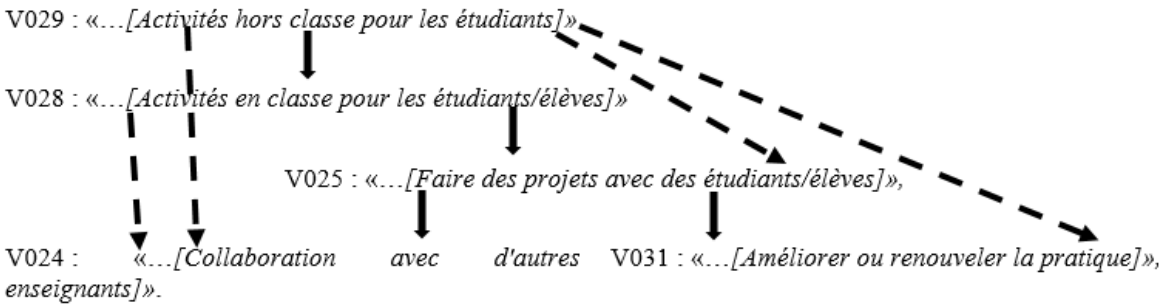


Figure 6

Nous pouvons observer que la variable qui apparaît comme une *cause* initiale, est celle qui considère que l'usage de Twitter est, avant tout, de permettre, le plus souvent possible, des activités hors de la classe. En conséquence de quoi, il en découle des usages en classe qui permettent de faire des projets avec les étudiants générant des possibilités de collaboration avec d'autres enseignants et des possibilités d'amélioration ou de renouvellement des pratiques d'enseignement. Twitter offre aux enseignants divers affordances (parfois chacune liée à l'autre) fonctionnelles pour atteindre leurs objectifs professionnels. Parmi les activités pédagogiques utilisant Twitter, nous avons eu un témoignage portant sur celles d'apprentissage de l'orthographe à l'école primaire et au collège comme le présente Brissaud (Brissaud et al., 2019; reseau-canope,v2016).

La construction de ce chemin repose principalement sur les contributions d'individus masculins qui ont entre 25 et 37 ans, c'est-à-dire les plus jeunes, ayant une ancienneté de 2 à 12 ans travaillant hors de la France dans d'autres contextes que l'enseignement primaire, le collège, le lycée ou l'enseignement supérieur.

Une autre *cause* initiale apparaît au nœud V027 :

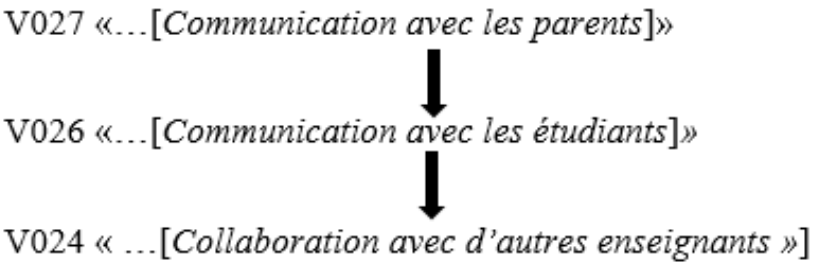


Figure 7

Cela peut signifier qu'un enseignant qui envisage d'utiliser Twitter pour communiquer avec les parents, l'envisagerait aussi pour communiquer avec les étudiants et même pour collaborer avec des enseignants. On constate que les enseignants/enseignantes qui utilisent Twitter perçoivent les possibilités de communiquer et de collaborer sur la plateforme avec les autres. Le concept d'affordance (Gibson, 1979) indique la capacité de l'utilisateur de Twitter à utiliser les possibilités de son environnement pour se communiquer et collaborer. Ces possibilités offertes par l'environnement sont les mêmes pour tous. Cependant, la perception, la capacité potentielle et l'objectif des utilisateurs différencient l'usage de l'artefact.

La construction de l'arc V027⇒V026 repose principalement sur les contributions d'individus féminins qui ont entre 51 et 64 ans, c'est-à-dire les plus âgées, ayant une ancienneté professionnelle de 25 à 40 ans travaillant hors de la France au collège. Il est normal que les enseignants communiquent avec d'autre enseignants avec lesquels ils n'ont pas de relations en face à face (les enseignants hors de la France). Au cours des dernières années, les enseignants ont utilisé Twitter comme un moyen d'établir des liens, d'élargir leurs sphères d'influence et de collaborer avec d'autres personnes (Carpenter et al, 2018, p.52).

5 Conclusion

À partir d'une analyse non symétrique des données d'une enquête sur l'usage professionnel de Twitter et à travers l'analyse et la lecture de graphes implicatifs produits par l'intermédiaire du logiciel CHIC 7, nous avons proposé, dans cet article, des interprétations des relations entre les variables pour une meilleure compréhension des affordances de Twitter au travers des usages par les enseignants/enseignantes à des fins professionnels mais aussi des interprétations relatives à comment ceux-ci connaissent les différents éléments et fonctionnalités de Twitter.

Comme nous l'avons souligné tout au long de notre propos, le concept d'affordance permet d'étudier et de mieux comprendre le sens des relations entre l'être vivant et l'artefact. Dans notre cas, la relation enseignant-Twitter génère la possibilité d'avoir un réseau socioprofessionnel sans frontières pour répondre à des besoins professionnels en fonction de certaines propriétés de l'environnement *affordant*. Les interactions enseignant-Twitter dans un usage socioprofessionnel requièrent des connaissances relatives au fonctionnement, aux fonctionnalités et aux propriétés de l'artefact ainsi qu'une explicitation de la finalité précise avec laquelle l'enseignant se trouve dans cet environnement.

Sur la base de notre échantillon d'usagers de Twitter, nous avons pu confirmer une hypothèse que les enseignants qui ont passé le plus de temps sur Twitter en général, sont aussi ceux qui connaissent le mieux et utilisent, en conséquence, davantage de possibilités de l'artefact. Notons toutefois que la raison la plus déterminante ici ne réside pas directement dans les possibilités qu'offre l'artefact, mais plutôt dans les perceptions que l'utilisateur en a. Il ressort alors que les affordances de Twitter sont plutôt liées à la capacité de l'enseignant à percevoir ses propriétés et fonctionnalités. Toutefois quand elles sont perçues, elles favorisent la réalisation des possibles actions. En outre, le but

d'usage de Twitter par les enseignants façonne leur relation avec l'environnement. Si un enseignant souhaite collaborer avec autres enseignants, Twitter lui permet de réaliser ses objectifs. Notons que la possibilité d'atteindre un objectif est accrue lorsque l'utilisateur a une bonne connaissance des éléments de l'artefact.

Références

[1] Brissaud, C., Viriot-Goeldel, C., et Ponton, C. (2019). Enseigner et apprendre l'orthographe avec la « Twictée » Premiers résultats de l'évaluation d'un dispositif innovant d'enseignement de l'orthographe. *Repères* p.107-130 https://doi.org/10.4000/reperes.2346

[2] Boyd, D. (2011). Social network sites as networked publics: Affordances, dynamics, and implications. In: Papacharissi, Z (ed.) *Networked Self: Identity, Community, and Culture on Social Network sites*. New York, NY: Routledge, pp. 39–58.

[3] Carpenter, J. P., Cook, M. P., Morrison, S. A., et Sams, B. L. (2018). Why Haven't I Tried Twitter Until Now?: Using Twitter in Teacher Education. *LEARNing Landscapes*, *11*(1), 51-64. https://doi.org/10.36510/learnland.v11i1.922

[4] Hanraets, I., Hulsebosch, J. et de Laat, M. (2011). Experiences of pioneers facilitating teacher networks for professional development. *Educational Media International*, 48 (2), 85-99. http://dx.doi.org/10.1080/09523987.2011.576513

[5] Henderson, M., Johnson, N. F., et Auld, G. (2013). Silences of ethical practice: Dilemmas for researchers using social media. *Educational Research and Evaluation, 19*(6), 546-560. doi:10.1080/13803611.2013.805656

[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.

[7] Gras, R., Régnier, JC., Lahanier-Reuter, D. Marinica, C., Guillet, F. (Eds) (2017). Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales. Toulouse : Cépaduès

[8] Gras, R., Régnier, J.C., (2017). Fondements théoriques de l'A.S.I. In R. Gras,J.C. Régnier, D. Lahanier-Reuter, C. Marinica, F. Guillet, (Eds) (2017) *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales.* Toulouse : Cépaduès

[9] Gras, R., (2018) La théorie de l'analyse statistique implicative ou l'invraisemblance du faux. Toulouse : Cépaduès

[10] Gibson, J.J., Crooks, L.E. (1938). A theorical field-analysis of automobile-driving, *American journal of psychology,* 51, 453-471.

[11] Gibson J. J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton-Mifflin.

[12] Gibson, J.J. (1966). The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin.

[13] Gibson, J. J. (1971). The legacies of Koffka's principles. *Journal of the History of the Behavioral Sciences, 7*(1), 3–9. https://doi.org/10.1002/1520-6696(197101)7:1<3::AID-JHBS2300070102>3.0.CO;2-1

[14] Gibson, J.J. (1979). The Theory of Affordances. *The Ecological Approach to Visual Perception.* Boston : Houghton Mifflin. p.127-137

[15] McLoughlin, C., et Lee, M. J. W. (2007). Social software and participatory learning: pedagogical choices with technology affordances in the Web 2.0 era. In R. Atkinson, C. McBeath, S-K. A. Soong, & C. Cheers (Eds.*), ICT: Providing choices for learners and learning* (pp. 664-675). Singapore: Centre for Educational Development, Nanyang Technological University.

[16] Norman, D.A (1988). The Psychology of Everyday Things. New York : Basic Books.

[17] Niveleau, C.E. (2006). Le concept gibsonien d'affordance : entre filiation, rupture et reconstruction conceptuelle, *Intellectica*, 1/43, 159-199.

[18] Reed, E. S. (1988). James J. Gibson and the psychology of perception. New Haven, CT: Yale University Press.

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

[20] Réseau-Canope. (2016). La twictée, dispositif collaboratif d'enseignement de l'apprentissage de l'orthographe. https://www.reseau-canope.fr/notice/la-twictee-dispositif-collaboratif-denseignement-de-lapprentissage-de-lorthographe.html

[21] Simonian, S. (2015). L'affordance socioculturelle une approche éco-anthropocentrée des objets techniques. *Le cas des environnements numériques d'apprentissage*. Habilitation à Diriger les Recherches, Universités Rennes 2

[22] Simonian, S., Quintin J.J., Urbanski S. (2016). La construction des collectifs dans l'apprentissage collaboratif à distance, *Les Sciences de l'Éducation – Pour l'ère nouvelle*, vol. 49, n°1, 57-82.

[23] Simonian, S. (2019). L'affordance, pour comprendre les rapports au numérique, *Education permanente*, 219, 61-70.

[24] Simonian, S. (2020). Approche écologique des environnements instrumentés : comprendre le phénomène d'affordance socioculturelle. *Savoir*, 52(1), 93108. https://doi.org/10.3917/savo.052.0093

[25] Turvey, M. (1992). Affordances and prospective control: An outline of the ontology. *Ecological Psychology*, 4,173–187.

Annexe 1

Vecteurs-Variables	**Code des variables binaires, composantes des vecteurs**	**Statut S = supplémentaire P= principale**
Sexe	1_Femme	S

	2_Je ne me prononce pas 3_Homme	
Age	1_ [25;37[2_ [37;44,6[3_ [44,7;51[4_ [51;64[	S
Pays	1_France 2_Hors de France	S
V006 Cadre de classe enseignée	1_Primaire 2_Collège 3_Lycée 4_EnsSup 5_Autre	S
V008 Ancienneté professionnelle	V008_1 de 2 à 12 V008_2 de 14 à 18 V008_3 de 19 à 24 V008_4 de 25 à 40	S
V10/011 Temps d'usage Twitter à titre personnel	V010/011_1 [0 - 2.5[V010/011_2 [2.8 - 5.4[V010/011_3 [5.5 - 8.2[V010/011_4 [8.5 - 10.9[	P
V012/013 Temps d'usage de Twitter à titre professionnel.	V012/013_1 [0;1.7[V012/013_2 [2;4.2[V012/013_3 [4.5;7.1[V012/013_4 [8;10.9[	P
Maîtrise des fonctionnalités de Twitter	V015 [Écrire un tweet] V016 [Faire une liste des différents comptes] V017 [Retweeter avec un commentaire] V018 [Ajouter un hashtag dans le tweet (#tweetschool)] V019 [Faire une mention (@un compte ou les organisations)] V020 [Tweeter avec des emojis-vidéos-images-gifs] V021 [Faire un sondage]	P
Finalités des usages professionnels de Twitter	V022 [Partage des ressources] V023 [Acquisition des ressources] V024 [Collaboration avec d'autres enseignants] V025 [Faire des projets avec des étudiants/élèves] V026 [Communication avec les étudiants] V027 [Communication avec les parents] V028 [Activités en classe pour les étudiants/élèves] V029 [Activités hors classe pour les étudiants] V030 [Participez aux débats autour de l'éducation] V031 [Améliorer ou renouveler la pratique] V032 [Suivre les avis des professionnels/experts]	

	V033 [Recevoir les dernières nouvelles éducatives]	

Tableau 3: Description des vecteurs-variables codés

Annexe 2

Chemin	Contribution des variables supplémentaires
V021-V016-V019-V017 V021-V016-V019-V018	Age1: 0.537 Age2: 0.25 Age3: 0.464 Age4: 0.565
	V008Anciennete1 : 0.127 V008Anciennete2 : 0.273 V008Anciennete3 : 0.62 V008Anciennete4 : 0.726
	Homme : 0.498 Femme : 0.496
	France : 0.578 HorsFrance : 0.223
	V006Autre : 0.326 V006EnsSup : 0.326 V006Lycée : 0.451 V006LCollège : 0.565 V006Primaire : 0.376

Tableau 11

Chemin	Contribution des variables supplémentaires
V021-V016-V020-V018	Age1: 0.476 Age2: 0.326 Age3: 0.473 Age4: 0.5
	V008Anciennete1 : 0.0896 V008Anciennete2 : 0.368 V008Anciennete3 : 0.794 V008Anciennete4 : 0.503
	Homme : 0.552 Femme : 0.462
	France : 0.6 HorsFrance : 0.182
	V006Autre : 0.283 V006EnsSup : 0.283 V006Lycée : 0.52 V006LCollège : 0.5 V006Primaire : 0.407

Tableau 4

Chemin	Contribution des variables supplémentaires
V027-V026	Age1 : 0.699 Age2 : 0.561

	Age3 : 0.161 Age4 : 0.138
	V008Anciennete1 : 0.885 V008Anciennete2 : 0.488 V008Anciennete3 : 0.196 V008Anciennete4 : 0.0454
	Femme : 0.129 Homme : 0.923
	France : 0.876 HorsFrance : 0.00299
	V006Autre : 0.177 V006EnsSup : 0.177 V006Collège : 0.037 V006Lycée : 0.646 V006Primaire : 0.586

Tableau 512

Chemin	**Contribution des variables supplémentaires**
V029-V028-V025-V24	Age1 : 0.372 Age2 : 0.488 Age3 : 0.375 Age4 : 0.533
	V008Anciennete1 : 0.231 V008Anciennete2 : 0.455 V008Anciennete3 : 0.396 V008Anciennete4 : 0.613
	Femme : 0.595 Homme : 0.246
	France : 0.704 HorsFrance : 0.0532
	V006Autre : 0.0506 V006EnsSup : 0.603 V006Collège : 0.185 V006Lycée : 0.283 V006Primaire : 0.827

Tableau 6

L'INCERTITUDE DES SCHÈMES POURTANT BIEN ORGANISÉS

Jeanne Guiet-Silvain[1]

RÉSUMÉ

A partir d'une recherche dans le cadre de ma thèse sous la direction de Gérard Vergnaud (1994), nous avions noté au début de mes observations, des différences et des ressemblances dans l'exécution de la division chez des élèves de fin d'école élémentaire et du début du collège. Mes analyses ont alors décrit et expliqué leur organisation cognitive grâce à des méthodes statistiques, rendant compte de stratégies comportementales, sur lesquelles se manifestent les conceptions et les pertinences accordées aux situations par le sujet. Certains processus de modélisation, ici étudiés, font appel aux analyses implicatives et à celles des similarités qui permettent, en effet, d'identifier des conceptions cohérentes individuelles et collectives au niveau des groupes étudiés. Nous insisterons ici sur la portée théorique majeure de la notion de schème de Gérard Vergnaud, s'appuyant ici sur des travaux statistiques exprimant en partie l'organisation cognitive du sujet.

***Mots-clefs** : Conceptualisation, schème, chemins cognitifs.*

ABSTRACT

Following research as part of my thesis under the supervision of Gérard Vergnaud (1994), we noted, at the beginning of my observations, various differences and similarities in the way division was assimilated by end-of-year students of elementary school and those in early middle school. My analyses at the time described and explained their cognitive organization using statistical methods, taking into account behavioral strategies, through which different concepts and a certain relevance were applied to the situations by the subject. A number of modeling processes, studied here, call for the analyses of implications and similarities which, in fact, make it possible to identify coherent individual and collective conceptual processes at the level of the groups studied. We will insist here on the major theoretical extent of Gérard Vergnaud's notion of scheme, based on statistical work expressing, in part, the cognitive organization of the subject.

***Keywords**: Conceptualization, scheme, cognitive paths.*

1 Introduction

« Je vous condamne à l'incertitude » Topaze (Marcel Pagnol)

Le présent chapitre rappelle le fil d'Ariane que j'ai déroulé, relatif à la compréhension de la construction du cheminement de l'enfant dans son quotidien, significativement illustré par l'exemple de la résolution de la division. J'avais mis en place des pratiques d'observations attentives afin de comprendre l'organisation, les logiques des procédures des élèves. Ainsi, le regard sur les activités mises en œuvre comprenait un inventaire d'usages fort complexes et variés. Dans le cadre de ma thèse sous la direction de Gérard Vergnaud (1994) et des travaux connexes qui ont poursuivi son élan, j'avais noté au début de mes observations, des différences et des ressemblances dans l'exécution de la division chez des élèves de fin d'école élémentaire et du début du collège. Dès lors, des difficultés

[1] Maîtresse de conférences, Université de Paris, département de Sciences de l'éducation, 45 rue des Saints Pères, 75006 Paris, jeanne.guiet-silvain@u-paris.fr.

d'analyse apparaissaient pour appréhender totalement la compréhension de leur organisation cognitive. A cette période, je me demandais comment les décrire, les expliquer ? Mon approche, influencée par leur prédictibilité, prenait le pas sur toutes les autres.

J'ai voulu apporter une contribution originale en retenant le caractère dynamique que l'on peut prendre en compte dans l'étude des difficultés conceptuelles des élèves. Je vais, ici, apporter une vision générale sur la base des travaux engagés. A cette occasion, je fis appel à Régis Gras[2] et à Saddo Ag Almouloud[3] (Université de Rennes I) afin d'étudier avec leur expertise, certaines analyses statistiques susceptibles de dégager des filiations de comportements des élèves face aux tâches données. Certaines de leurs conceptions correspondent à des convictions, suivant des raisonnements propres éventuellement implicites, dans tous les cas, admises de facto.

Bien que représentant les mêmes structures, la clé de cet écrit est bien de comprendre certains processus de modélisation, rendant compte de stratégies comportementales, sur lesquelles se manifestent les conceptions et les pertinences accordées aux situations par le sujet. La progression cognitive se fait au prix de ces chemins cognitifs, dont l'organisation est significative sur les systèmes explicatifs en présence. Les méthodologies choisies pour interpréter les résultats observés font appel aux analyses implicatives et à celles des similarités qui permettent, en effet, d'identifier des conceptions cohérentes individuelles et collectives au niveau des groupes étudiés.

Enfin, je ne remercierai jamais assez Régis Gras, Gérard Vergnaud et Maryvette Balcou Debbusche[4] pour leur soutien moral dans la réalisation de mon HDR[5] , ainsi que Saddo Ag Almouloud et Nadja Acioly Régnier[6], qui m'ont permis de reprendre ce cheminement de recherche.

2 - Le cheminement des schèmes : des pratiques d'observation à l'utilisation de méthodes de traitement statistique

2.1 La démarche

Tout en faisant porter l'analyse de la structure psychologique des élèves, mon étude a voulu s'étendre -au-delà de mes réflexions initiales relatives à la complexité des outils de pensée utilisés- au support théorique des champs conceptuels de G. Vergnaud (1996). Mon objectif était d'étudier l'analyse du caractère dynamique que l'on peut retenir dans l'étude des difficultés conceptuelles des élèves. L'intérêt de ce positionnement a pu s'affirmer dans l'étude initiale des champs de l'addition et de la multiplication (Vergnaud, 1988). Ces démarches systématiques ont été reconnues par Brun (1993) dans leur capacité à se représenter le « contenus d'enseignement et agir à la formation des concepts chez les

[2] Régis Gras, professeur d'université émérite, Ecole Polytechnique de l'Université de Nantes, Équipe DUKE Data User Knowledge.

[3] Prof. Dr. Saddo Ag Almouloud - Programa de Estudos Pós-Graduados em Educação Matemática PUC/SP - Pesquisador Bolsista do CNPQ - Lider de Grupo de pesquisa PEAMAT - Vice-coordenador do PEPG em Educação Matemática - PUC/SP.

[4] Maryvette Balcou Debussche, professeur de Sciences de l'éducation, Université de la Réunion.

[5] Habilitation à diriger des recherches, Laboratoire Caref, Amiens, Soutenue le 11 janvier 2021.

[6] Nadja Acioly Régnier, professeur de Sciences de l'éducation, Inspe, Université Claude Bernard, Lyon.

élèves. Pour Brousseau (1988), l'ensemble des questions didactiques soumises à des situations, est associé à une circonstance précise, désignable, mettant en jeu la construction du sens. J'avais eu un échange nourri, lors d'une invitation[7] à Bordeaux chez lui, afin de situer sa position de recherche sur ces questionnements.

Le travail de collaboration autour de ce projet de recherche a contribué à une vision dynamique de mon expérience de jeune chercheuse. Soucieuse de relier mon travail à une critique sur ce que je croyais savoir, il m'a fallu réorganiser mes connaissances, l'entrée dans une problématique n'étant pas spontanée. J'avais déjà apporté une base à cet édifice grâce à la collaboration de F. Conne à l'occasion d'une école d'été en 1989[8]:

Mon intervention situait la question de l'entrée des élèves et du maître de la technique de la situation qui supposait de prendre une position déterminée par rapport au savoir : il s'agissait de suivre le calcul pas à pas dans un ordre d'actions déterminé. A l'appui d'études de cas, la réflexion portait sur le problème suivant : comment s'effectuait le passage de la situation mathématique à la situation didactique. J'avais retenu une des caractéristiques essentielles portant sur la théorie des situations (Brousseau, 1998), ce qui correspond à une armature conceptuelle afin de penser la diffusion des connaissances.

Tout en faisant clairement ressortir la difficulté éprouvée par l'élève lors de la résolution de telle ou telle opération, la présence et les formes de cette même difficulté paraissaient peu mesurables. Il me fallait, donc trouver une méthode répondant à l'identification et à la hiérarchisation des erreurs ; cela, alors que mon bagage de connaissances statistiques était, à cette époque-là, limité. Les conséquences des implications pédagogiques et didactiques à tirer de ces observations, étaient, in fine, envisageables pour la suite à donner après la thèse.

Invitée à l'Université de Rennes I, j'ai pu me former et travailler avec une équipe dont la collaboration a été un moteur propice et pour moi novateur. Les conclusions fournies par l'analyse statistique ne pouvant, néanmoins, être dissociées de l'étude des structures et des caractères mêmes des divisions, mon objet d'étude supposait bien mon engagement dans une équipe. Cette chance s'est présentée naturellement.

Être admise, le temps de cette thèse, au sein de l'équipe de Régis Gras, impliquait de maîtriser un large spectre de connaissances mathématiques. Son travail de pédagogue à mon égard, nourri de patience et de dévouement, fut une expérience qui m'a permis d'apprécier la valeur de ce champ de recherche, de l'espace explicatif qu'elle présentait, la portée significative des résultats et des conclusions. Tandis que je recueillais des données afin de dégager les éléments nécessaires à cette étude, je participais à des communications censées parfaire ma formation dans ce domaine. Tout d'abord épaulée par Saddo Ag Almouloud, je fus appelée à présenter les méthodologies implicatives utilisées dans ma thèse lors de deux séminaires. L'un, à l'Université Paris VII, dans le cadre de la présentation de travaux de jeunes chercheurs avec la collaboration de Ag Almouloud S., Dagher A.[9], l'autre, à l'Université de Rennes I, organisé par R. Gras en

---

[7] Je remercie vivement Nadine et Guy Brousseau pour leur invitation généreuse et chaleureuse chez eux à Talence.

[8] Conne F., Guiet J., (1989), "Situations évoquées, situations jouées et structures mathématiques", *Actes de la Vème Ecole d'Eté de Didactique des Mathématiques,* Plestin les Grèves, 51-57.

[9] Guiet, J. (Février 1993). "Présentation de diverses méthodologies (analyses implicatives, factorielles des correspondances appliquées à des recherches), *Séminaire DIDIREM,* Université de PARIS VII."

1993[10]. Un projet de publication (1993) sur ces thèmes fut alors porté par le département de mathématique et informatique de l'Université de Rennes I (IREM).

Ma préoccupation était de trouver un moyen de comprendre les filiations et les ruptures dans les raisonnements des élèves. Dans ce cas, analyser et apprécier les diverses dimensions des difficultés rencontrées par les élèves supposait d'identifier des liaisons dans un ensemble de repères et d'expressions, afin de mettre en évidence une structure cognitive commune chez un ensemble d'élèves de niveau scolaire équivalent. L'évaluation de la stabilité ou la variabilité des erreurs des sujets nécessitait la mise en œuvre d'une méthode appropriée.

Conduire des analyses fines sur l'activité cognitive et gestuelle n'est pas un traitement donné d'emblée. Cela suppose de considérer l'organisation de l'activité non seulement sous l'angle de l'algorithme, mais également sur celui du schème[11]. Le traitement des situations, étudiées ici, a conduit au constat suivant. Je propose de nommer la division « script-algorithme », dans la mesure où le schème porte à la fois sur des signifiants et des objets de pensée. En effet, une représentation symbolique comme celle de la division en colonnes, est essentielle à l'algorithme et aux objets sur lesquels il porte, à chaque étape. Réciproquement, la manipulation des signifiants fait nécessairement appel aux concepts. Tout en suivant Vergnaud (1985), ce sont les invariants opératoires, souvent implicites, reconnus et utilisés par les sujets qui constituent la base conceptuelle des schèmes : tel est le cas des concepts de produit partiel et de dividende d'étape par exemple.

Les erreurs et les raccourcis observés dans la réalisation de la tâche permettent d'indiquer un cheminement. La cohérence ou l'incertitude témoignent de leur effectivité. Quels sont les registres, les rouages en jeu ? Comment accorder une crédibilité aux phénomènes ? Le schème, en qualité d'unité d'analyse fonde nos questions.

Il demeure un repère et me semble permettre la prise en compte d'interrogations directement liées à la conceptualisation (anticipations, retours en arrière…). La manipulation de signifiants nécessite la mise en évidence du schème, sa « saisie ». L'explicitation dans le langage ou dans un symbolisme mathématique a une fonction différente selon les cas. Dans notre étude, la catégorie de dividende partiel pose d'autres problèmes conceptuels que ceux de dividende, de diviseur, de quotient, et de reste. Ma posture de recherche suppose qu'une théorie de la conceptualisation mathématique attribue nécessairement une part à l'épistémologie.

L'intention du sujet qui agit selon certaines formes d'activité, engage des erreurs et des raccourcis pouvant le priver de parvenir immanquablement au résultat espéré. Le gouvernail est l'incertitude, composante des schèmes, dont la conséquence, en termes d'activité, s'engage dans une organisation plus ou moins hiérarchisée, qui va permettre au sujet de croire que sa démarche est couronnée de succès. Le schème a donc pour rôle l'agencement cohérent de l'activité, parfois dans un temps relativement court. En effet, il arrive que l'interprétation confirme une connaissance évidente. L'évaluation de la stabilité ou la variabilité des erreurs des sujets aboutit à l'interprétation suivante : l'absence de

---

[10] Guiet, J. (Mars 1993). « L'algorithme de la division est-il vraiment un algorithme ? », *Séminaire de Didactique des Mathématiques* organisé par R. Gras, I.R.M.A.R., Université de RENNES I, Campus de Beaulieu.

[11] Une première définition de la notion de schème : « Organisation de la conduite dans une classe de situations donnée » 1987. « Le schème selon Gérard Vergnaud » In *Psychologie de l'éducation*, Bréal éditeur, p.79.

contradiction avec le "connu" ou le "vraisemblable", apporte une grande crédibilité, là est son intérêt principal. Les surprises des observations menées mettent en évidence des phénomènes peu connus, voire surprenants. De façon générale, les deux méthodes basées sur la similarité et l'implication, que nous allons présenter, permettent de définir, à partir de classes de procédures, des conceptions stables et consistantes. Les enfants développent des interprétations en fonction des situations qu'ils ont rencontrées et cru comprendre. Mon étude veut rendre compte des "raccourcis", voire des improvisations que le sujet effectue.

Au cours de mes travaux, le préalable à l'interprétation des erreurs des élèves a été la prise en considération des caractéristiques des divisions présentées : les explications susceptibles de m'éclairer étant liées à leurs singularités. Le choix de ces particularités a reposé sur l'hypothèse qu'elles ont éventuellement une influence sur l'utilisation des propriétés mathématiques en jeu dans l'algorithme, et ceci en fonction de l'apprentissage en usage dans notre population d'élèves. Mon champ est réduit au rapport entre sens et algorithme, j'ai situé ma question lors d'un séminaire préalable à la thèse (1992). Ces deux leviers correspondent à l'exposé d'une dynamique, abordant les questions de conception (théorie de Brousseau, 1983), celles de rapport au savoir (Chevallard, 1985), celles de concept, invariants opératoire et schème (Vergnaud,1996).

Mes premières hypothèses avaient fait l'objet d'une présentation, lors d'un séminaire en psychologie cognitive[12], organisé par Ricco[13], en 1992 considérant les aspects parcellaires des savoirs, constitutifs des dispositions des apprenants, en fonction des premières observations de terrain.

Mes résultats rendaient compte de l'organisation de l'activité, qui consistait à faire des arrangements selon les importances accordées à telle ou telle situation. Ricco (1990), abordait à cette occasion la logique positive de la transposition répondant à la question de l'insertion d'un savoir nouveau, en présentant également l'évolution historico-critique des ruptures épistémologiques.

La question du sens est fondamentale, je la situe dans une perspective accordée par Lemoyne *et al.* (1993) à l'aide d'un exemple sur les fractions, qui soutient les positions respectives de Vergnaud (1985), sur la théorie des champs conceptuels et d'Y. Chevallard, (1985), sur le problème du sens dans l'enseignement et l'apprentissage des mathématiques, à partir de différents registres sémiotiques. Ajoutons, avec Conne et Lemoyne, que « […] deux sources de difficulté se conjuguent pour l'enseignante […] d'une part sa conception des rapports numération orale-numération écrite, […] d'autre part son assujettissement à tout un ensemble de la relation didactique. » (1999, p.346).

Il faut garder à l'esprit qu'un schème n'est, en général, pas un algorithme, qui lui, a pour visée d'aboutir en un nombre fini de d'étapes (ou son effectivité) au traitement de toute situation appartenant à la classe considérée. L'incertitude est l'habillement du schème. À partir de classes de procédures, des conceptions stables et consistantes ainsi que les méthodes engagées par le sujet, mettent en évidence des phénomènes peu connus,

---

[12] Je remercie Graciela Rico ainsi que Gérard Vergnaud, pour leur accompagnement en qualité d'allocataire-moniteur puis mon poste d'ATER à l'Université Paris 5 (1990-1995).

[13] Guiet, J., (1992). *Séminaire de D.E.A* : "Psychologie cognitive", organisé par Ricco Graciela (professeur en Sciences de l'éducation", Sens des algorithmes à la fin de l'école élémentaire et au début du Collège ».

parfois assez inattendus[14]. Le schème nous permet d'échapper à une lecture linéaire de l'évolution des connaissances comme le rappelle Flückiger (2000) : « La notion de schème est en dualité totale avec la notion de situation, c'est ce qui en fait, nous semble-t-il, la pertinence dans les analyses didactiques ». Un séminaire, en mars 1998, a porté sur les résultats principaux des difficultés observés des sujets en situation que j'avais observé, en relation avec la notion de schème.

Les normes disciplinaires rendent compte de difficultés chez les élèves, à cause des connaissances établies ou développées. L'individu rend le changement souhaitable, attractif et recherché ou au contraire problématique, néfaste et écarté. Les élèves, dès lors qu'ils subissent des intégrations de réponses automatisées, se soumettent aux lois du hasard, à des généralisations, des contradictions, le tout étant à géométrie variable au sein d'une régulation épistémique. Il me fallait ainsi analyser les activités des apprenants et les obstacles fort nombreux que j'avais observés en situation.

2.2 Description des méthodes

Lors de mon analyse *a priori*, 28 divisions ont été sélectionnées par ordre de difficulté croissante, regroupées en fonction de leurs caractéristiques opératoires. A partir de cette classification, l'exercice périlleux cherchait à faire émerger une relation entre les procédures des élèves et les caractéristiques des divisions données. Il s'est avéré délicat d'observer au long cours, finement et rigoureusement, ces situations. Au cœur de mes investigations, les différents sens alloués au « schème-algorithme »[15], alors que celui-ci est en voie de constitution chez un élève, m'ont imposé d'identifier ses diverses productions.

J'ai constaté que le niveau de difficulté de la description des erreurs observées était étroitement lié à la présence ou l'absence de reste dans une division, effectuée ou non, au quotient décimal ou non décimal, à l'existence ou non de zéros au dividende, au quotient de cette opération et au calcul d'une soustraction posée ou non. La difficulté de cette description était également fonction de la position de la virgule au quotient, de l'ordre de grandeur du dividende et du diviseur et de celui du dividende d'étape. Je reviendrai plus loin sur ces éléments, pour montrer qu'ils se décomposent en paramètres plus fins, qui pèsent d'un poids inégal dans les difficultés des élèves.

Le champ de mon analyse est bien celui des élèves, non celui de l'enseignant. Je me contente de décrire et d'analyser le sens que l'élève attribue à la division. Je fais l'économie de la prise en considération du jeu de l'enseignant avec ses élèves, ainsi que des ruptures de contrat. Dans un contexte proprement didactique, les erreurs liées à l'exécution de l'algorithme n'épuisent pas l'ensemble des erreurs observées. J'ai choisi trois méthodes d'analyse statistique.

- Une classification hiérarchique (selon Lerman),
- Une classification implicative (selon Gras et Larher),
- Une analyse factorielle des correspondances multiples.

Mon but, entre autres, est d'identifier les relations entre les grandes classes de comportements erronés. Je peux regrouper ces erreurs en classes lorsqu'elles admettent

[14] Guiet, J., (Mars 1998). "La conceptualisation : problème central dans l'analyse des démarches", *Séminaire de l'UPRES 1*, (Unité Propre de Recherche de l'Enseignement Supérieur), Université de Paris 8 Saint-Denis.

[15] Expression retenue avec le concours de G. Vergnaud.

entre elles une similarité ou une certaine "cohésion". Je recherche ensuite les implications éventuelles entre les variables retenues (réussite, erreurs de différents types etc.), de manière à préciser certaines hypothèses concernant les procédures erronées, et à dégager des types généraux de comportements. J'ai utilisé la méthode de classification (Lerman, Gras et Rostam, 1981) selon l'algorithme dit de « la vraisemblance du lien ». Ici, les filiations procédurales (Gras, Lahrer 1992), entre les différentes classes de variables chez les élèves représentent un ensemble de caractères et de classes de caractères et mettent en évidence une structure cognitive stable chez cette population[16].

En repérage, j'avais retenu, grâce à l'analyse statistique implicative des données, des structures implicatives communes à des variables ou des classes de variables. A ces relations implicatives, il est possible d'associer la construction d'un graphe orienté et interpréter la structure générale des implications. L'analyse des similarités montre la cohérence des choix des sujets et permet, par conséquent, d'accéder à leurs représentations. Une vision globale des enchevêtrements des liaisons implicatives semble conduire à des prédictions fiables représentées par le graphe (Gras, Fontaine, Larher, Nicolas, Simon, 1994). Cette intuition interprétative des graphes implicatifs est féconde en termes d'analyse. Je vais ici en présenter les aspects majeurs : « La méthode d'analyse statistique implicative (ASI) apparaissait dès lors comme un outil parfaitement adapté pour rendre compte d'un ordre de complexité cognitive sous-jacent à un ensemble de variables mettant en jeu les compétences cognitives d'un ensemble de sujets », (Gras et Bodin, 2017). Cette méthode d'analyse de données non symétrique, à visée prédictive, a été conçue par Régis Gras (1979). Elle m'est apparue nécessaire, afin de modéliser, grâce à une mesure de qualité, les relations entre deux observations comportementales (Gras, Almouloud, Bailleul, Larher, Polo, Ratsimba-Rajohn, Totohasina, 1996). L'ASI est originale quand elle se place sur l'invraisemblance de la présence de ces relations implicatives, celles que l'on ne détecte pas a priori, ou alors de manière hasardeuse. La quasi-implication est significative du point de vue de l'organisation interne des schèmes de l'élève. Par exemple, certains calculs impliquant le positionnement d'une virgule au quotient sont systématiquement échoués par les élèves, de l'école primaire au collège. Ici les erreurs ne sont pas éphémères, elles expriment une cohérence. Ces relations d'implication, mesurées statistiquement de manière significative, m'ont permis d'extraire une structure dynamique des variables en jeu.

2.3 Premiers éléments d'analyse

D'après Gras (1979), la comparaison des résultats de l'analyse implicative avec ceux obtenus par les méthodes classiques d'analyses de données, me convainc de l'originalité de cette méthode pour analyser les liaisons dissymétriques et la hiérarchie des classes qu'elle fait émerger. Cette technique est synthétique. Elle est fondée sur le concept de cohésion (au sein des différentes classes de variables) et sur celui d'entropie, et aboutit à des indices d'implications entre classes. Les questions concernant la complexité progressive des conduites, peuvent y trouver réponse. Les hiérarchies implicatives obtenues m'ont permis de mieux comprendre le sens des procédures des élèves. L'emploi d'une telle démarche suppose le recours à des méthodes ayant des fondements mathématiques différents, comme l'analyse hiérarchique, l'analyse des similarités ou l'analyse factorielle de correspondances, qui conduisent à des résultats se complétant de

[16] Chez un ensemble d'élèves de niveau scolaire équivalent. Ici des élèves de CL1, CM2, 6ème et 5ème.

façon enrichissante. Pour traiter ce problème, une piste consiste à prendre en compte pour l'étude du comportement d'un individu confronté à une valeur manquante sur un attribut les valeurs des autres attributs qui lui sont liés au sens de la quasi-implication. Tel est bien mon souci méthodologique, avant qu'en soit donné les résultats.

3 Les schèmes dans les procédures des élèves en question

3.1 Relations non-symétriques entre variables : relations de quasi-implcation

Ma recherche a permis de valider l'usage de l'outil statistique : erreurs et réussites ont été plus clairement identifiées. Donnons un exemple : l'ordinateur donne la liste des intensités d'implication, à un seuil fixé par exemple .95 dont les valeurs permettent de créditer les hypothèses formulées. A partir de ce graphe d'implication, je peux émettre des conjectures du type : la réussite à une question donnée entraîne la réussite à une autre question, telle erreur implique telle autre erreur. Les mots "entraîne" et "implique" n'ont pas de sémantique causale (Gras., Bodin, 2017, *op.cit.*). C'est l'analyste qui, selon les contextes, pourra faire l'hypothèse de cette causalité, plus particulièrement si l'intensité d'implication est forte (Lerman, Gras, et Rostam*, op.cit.,*1981). La mesure de l'implication statistique d'un attribut a sur un attribut b suit une méthode analogue à celle de Gras (1979) ; Lerman, le confirme par la suite (1981). Plus cette intensité est grande, plus l'affirmation de l'implication est crédible (Gras *et al,* 1989)[17].

Attachons-nous ici un instant à en saisir le procédé (voir l'illustration ci-dessous) : par exemple, considérons l'implication a--b. Elle est admissible au sens suivant : l'effectif de réussite à la division (noté 13), ensemble caractérisé par a, mais ne réussissant pas la division (noté 18), ensemble caractérisé par b (élèves dont les protocoles contredisent donc l'implication), est significativement faible par rapport à celui que l'on pourrait attendre d'une absence de lien unilatéral entre ces deux divisions. En d'autres termes, d'une part, l'effectif de réussite de la division n°13 est plus faible que celui de la division n°18 ; d'autre part, il existe fort peu d'élèves réussissant a et pas b : Ea est presque totalement contenu dans Eb. Ainsi, on dira que réussir la division n°13, implique au sens large, réussir la division n°18 ou bien encore, si les difficultés relatives à la division n°13 sont maîtrisées, alors celles de la division n°18 le sont aussi. Par ailleurs, les élèves qui réussissent la division n°18, ne réussissent pas nécessairement la division n°13, car la liaison est unilatérale.

[17] "Un graphe dit d'implication donne une représentation de l'ensemble des relations implicatives les plus fortes, opératoire pour l'analyse et l'interprétation." ("Le micro-ordinateur, outil de révélation et d'analyse du raisonnement déductif de l'élève", Gras, Fontaine, Larher, Nicolas, Simon, *Publication de l'I.R.M.A.R.*, Rennes I, Fascicule 5, p. 32, Février 1989).

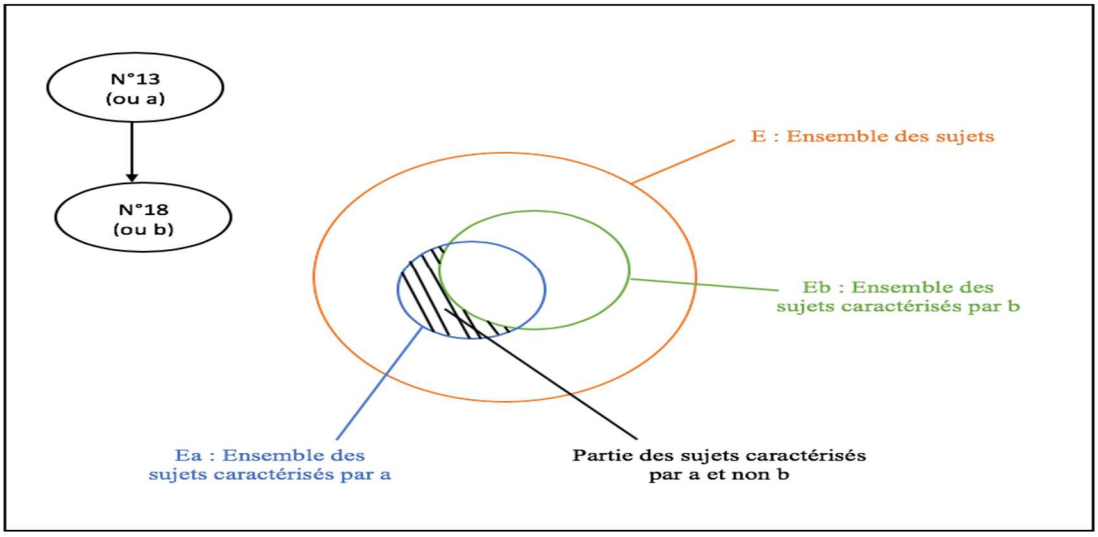


Schéma n°1 : Implication de données

A l'occasion d'un colloque "Vingt ans de Didactique des Mathématiques en France", organisé par l'A.R.D.M (Association pour la Recherche en Didactique des Mathématiques) en 1993[18], j'ai pu concrétiser ma réflexion au sein d'équipes promouvant l'éventail de ce type de recherches[19].

J'ai exposé les questions fondamentales suivantes : comment repérer les régularités, les significations dans la construction des connaissances ? Quelles contraintes peut-on observer dans les processus d'enseignement ? Au sein de la diffusion des connaissances, quels sont les choix sur lesquels l'enseignement peut jouer ? Ces questions ont créé des débats majeurs sur les théories et les méthodes originales répondant à ces questions. Mon approche, résolument cognitive, soutenue par les concepts de schèmes et de champs conceptuels, s'est focalisée sur les régularités et les filiations des procédures d'abstraction.

Une publication d'un chapitre dans un ouvrage scientifique a suivi cette présentation[20]. Ici, mon champ de recherche, identifiant une famille de comportements m'a permis d'inscrire plus largement des développements de recherche notamment ceux, plus approfondis, issus de l'analyse des graphes d'implication et de leur cohésion.

Ces travaux longtemps convergents, se sont progressivement enrichis par des initiatives individuelles, d'où leur inévitable diversification, ce qui correspond à une affirmation autant qu'à une différenciation des travaux en didactique des mathématiques. Les dynamiques individuelles des chercheurs prennent leur essor dans des champs diversifiés, et la collaboration interdisciplinaire de la recherche en didactique et en psychologie est ici favorisée (voir en particulier les travaux de Vergnaud,1994).

---

[18] Cet ouvrage est l'objet d'une recension dans le Bulletin de l'APMEP n° 393.

[19] Guiet, J., (Juin 1993). "L'algorithme de la division est-il un algorithme ?" *Colloque organisé par l'A.R.D.M.*, (Association pour la Recherche en Didactique des Mathématiques), "Vingt ans de Didactique des Mathématiques en France", à l'I.N.R.P., rue d'Ulm, Paris.

[20] Guiet J., (1994), "Algorithmes et schèmes : cas de la division", Artigue, Gras, Laborde, Tavignot (Eds), in "Vingt ans de didactique des mathématiques", *La Pensée Sauvage :* Grenoble, 225-230.

3.2 Les graphes implicatifs

Pour comprendre le sens d'une cohésion à l'intérieur de chaque classe de variables, il m'a fallu considérer un autre niveau d'analyse, situé au-delà de la seule prise en compte de similarités entre classes, afin que le flux implicatif d'une classe sur l'autre soit nourri du flux interne de la première et alimente la seconde. Cette cohésion se traduit par une mesure implicative donc orientée, telle une filiation procédurale ou une genèse. Le calcul de cette mesure rend compte de la faiblesse entropique de la classe de variables considérée. Les contraintes liées à ces calculs et leur expression mathématique sont expliquées par Gras et Lahrer (1992).

"Cette liaison sera, dans le cas des classes de procédures, d'autant plus crédible que cette classe présentera elle-même une cohésion dirigeant les structures internes dans le même sens que celui qu'elles devront exprimer sur le plan externe. En psychologie cognitive, nous dirons que nous cherchons des micro-genèses à travers des liaisons intra-classes se développant et se ramifiant en macro-genèses à travers des liaisons inter-classes. Il est alors évident que si cet objectif est atteint, sa nature se distingue nettement des méthodes traditionnelles d'analyses de données où distance, similarité et variance représentent les critères structurant ces données ou des classes de données de façon symétrique. Ici, c'est au contraire la dissymétrie qui nous intéressera car elle seule restituera la dynamique qui sous-tend une genèse" (Lahrer, 1991).

Le fait que ce réseau soit clairement détaché des autres m'incite à penser que nous avons ici une structure forte. Les chemins des réseaux sont significatifs, et sont ainsi associés à une hiérarchie construite selon les méthodes de construction ascendante classique (Gras, Totohasina, Ag Almouloud, Ratsimba Rajohn, Bailleul, 1994). A certains niveaux, leur impact est étudié selon un critère de classification implicative, ainsi que leurs nœuds significatifs. Soutenue par la dissymétrie des implications, nous lisons la richesse des structures obtenues. Nous approchons ici du caractère modèle épistémique de l'élève. Classant leurs procédures par famille et en fonction de leurs centrations, les différentes résistances à la compréhension de l'algorithme nous sont apparues, reposant sur des conceptions parfois stables, souvent pertinentes. L'ordre des variables prend ici une importance majeure dans l'interprétation de l'implication en cause. Ce graphe permet de représenter synthétiquement des relations entre attributs, en mettant en évidence les relations de constructions des implications, de leur proximité, de leur incrémentation (Gras, Kuntz, Briand, 2001).

Je traduis ci-dessous l'interprétation d'un « chemin » implicatif des structures cognitives des élèves impliqués dans les familles de difficultés identifiées, les numéros étant ceux des divisions. Des familles de comportements se distinguent : la complexité du calcul en fonction de leur nature, le zéro intercalaire au quotient, par exemple suit un « chemin » identifiable : 25---13---21----18--- etc. Ce qui crée une analyse où ce que j'appelle « l'emboîtement » des comportements est nettement marqué.

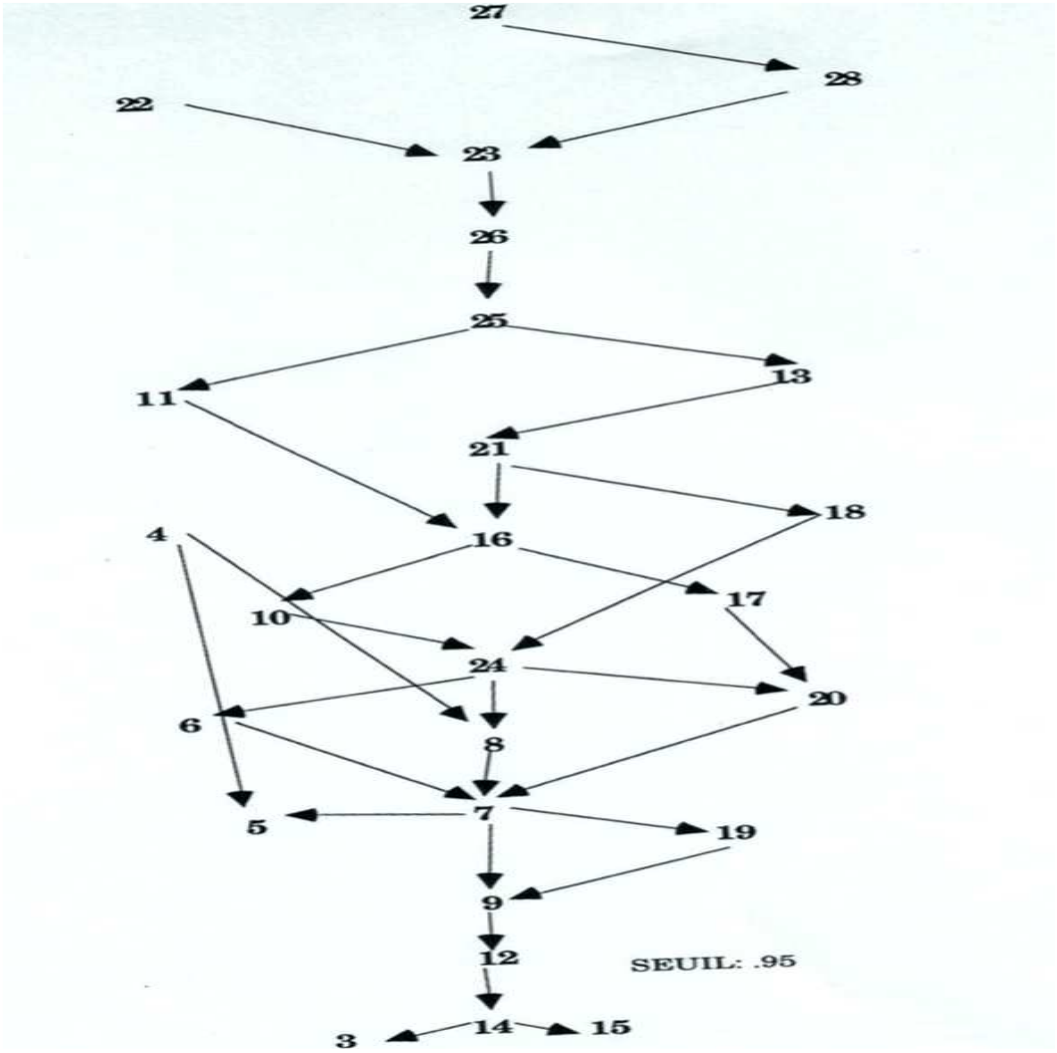


Schéma n°2 : Arbre hiérarchique des implications pour les 4 niveaux scolaires du CM1 à la Sixième : Exemple d'un graphe implicatif, obtenu par CHIC (source Thèse, 1994)

Une représentation de ce graphe a été informatisée par Rostan à Rennes, et Ratsimba Rajohn à Bordeaux par la suite). Au-delà de l'intérêt apporté à l'analyse des graphes implicatifs, il ne me permettait pas de décrire tous les comportements d'une population d'élèves. Une hiérarchie cohésitive est une analyse venant compléter la richesse de l'interprétation en identifiant les familles de règles sous-jacentes révélatrices de la complexité cognitive suivant l'emboitement des procédures des élèves. Mon analyse porte ainsi sur :

- la consistance croissante des niveaux,
- les niveaux d'acquisition cognitive d'un concept et leur généralisation,
- la hiérarchie qui possède une dimension soulevant l'hypothèse d'une équilibration des structures cognitives en jeu,
- le repérage des sauts de continuité ou de discontinuité procédurales qui suit une autre hypothèse liée à celle de l'assimilation et de l'accommodation (Piaget, 1975) puis de l'adaptation, mettant en correspondance les structures des fonctions de l'organisme du sujet accompagnant celles des organisations des schèmes cognitifs.

Des partitions fines sont ainsi construites, à l'aide d'un critère de similarité entre variables, ce qui permet une interprétation en termes de typologie et de ressemblance décroissante des noyaux de variables, constitués significativement à certains niveaux de l'arbre et s'opposant à d'autres à ces mêmes niveaux. A chaque niveau de cet "arbre" de classification, deux modalités sont rassemblées : c'est au dernier niveau que l'indice de similarité est, bien entendu, le plus faible (Gras *et al,* 1981)[21]. L'ordre prend alors une importance majeure dans l'interprétation que l'on fera de l'implication en cause (Bailleul, 1994). J'identifie alors des familles de procédures grâce aux arbres implicatifs dont l'illustration ci-après me permet de faciliter l'interprétation.

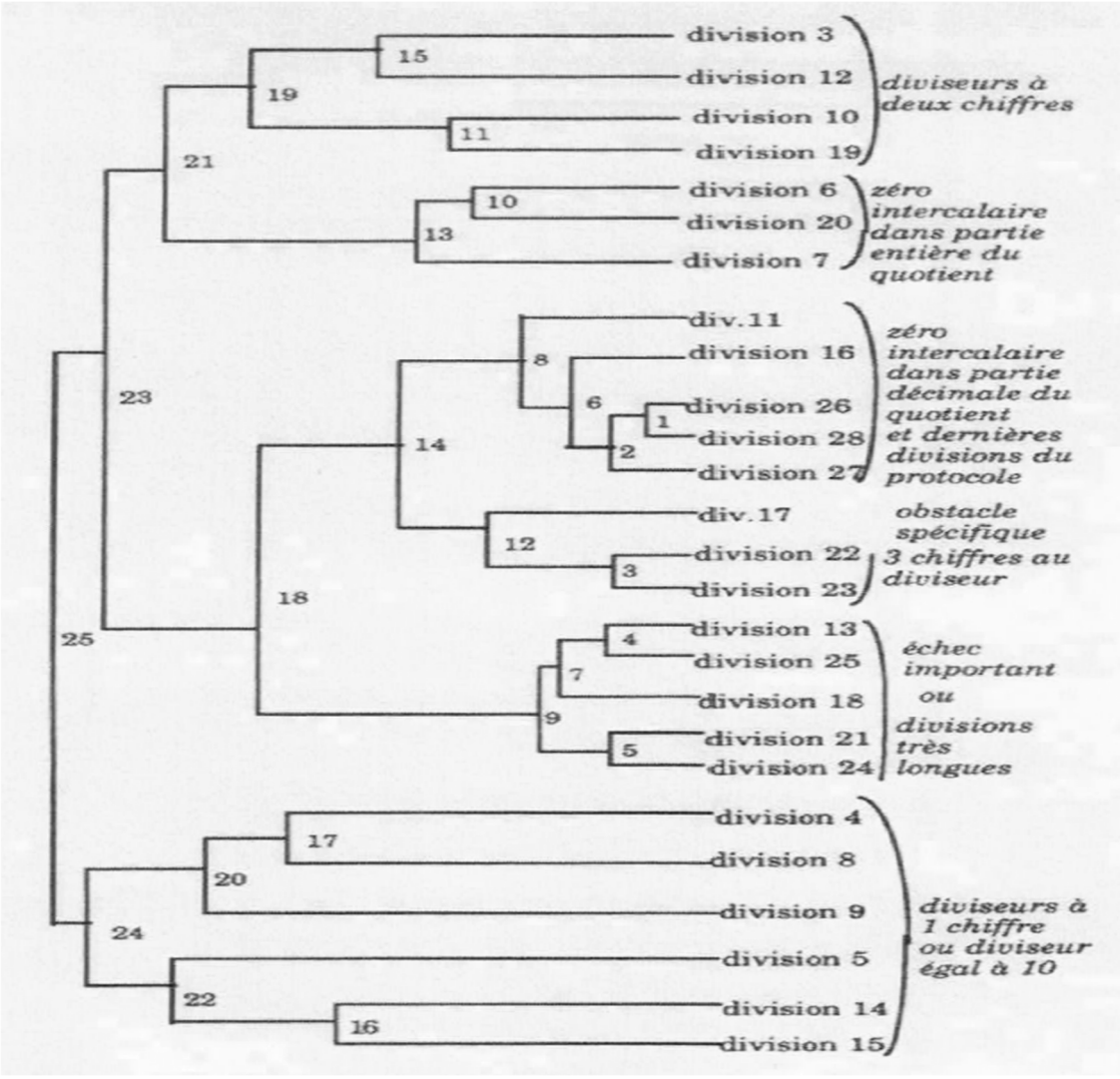


Schéma n° 3 Arbre hiérarchique des similarités pour les 4 niveaux scolaires °Exemple de hiérarchie cohésitive (source : Thèse, 1994)

L'arborescence s'élève de façon significative dès que le nombre de chiffres au diviseur augmente. Certains schèmes révèlent ici leur fragilité pour les divisions dont le dividende

[21] Une illustration intéressante est fournie par l'article rédigé par Gras, Bessot, Comiti, (Novembre 1981) : "Analyse d'opinions d'instituteurs à l'égard de l'appropriation des nombres naturels par les élèves du cycle préparatoire", Publication interne n°155 de l'I.R.I.S.A. (Institut de Recherche en Informatique et Système Aléatoire), Rennes.

d'étape est inférieur au diviseur ; cette caractéristique perturbe les conceptions du sujet. De même, les divisions qui "ne se terminent pas" heurtent l'anticipation que l'enfant peut effectuer pour l'obtention du résultat. En ce cas, le schème "partager" se prête à des inférences : l'enfant "s'arrange" pour terminer le calcul avec un modèle plus simple (qui en ce cas se traduit par une erreur). Le schème consistant à poser un zéro au quotient engendre des actions différentes, en fonction de sa position au quotient : soit au début (premier chiffre du quotient), soit intercalaire (après ou avant une virgule), ou encore terminal. Les décours de la pensée de l'élève sont remarquables, il est passionnant de porter cette classification au niveau de l'analyse.

Dans le cas des divisions qui possèdent un zéro intercalaire après la virgule, les problèmes rencontrés par les élèves sont liés à la conjonction de trois processus indépendants. Il faut "abaisser" un zéro quand les chiffres composant le dividende ont été utilisés, poser une virgule au quotient, puis "abaisser" de nouveau un zéro car le dividende d'étape est inférieur au diviseur. Les schèmes utilisés peuvent alors « dégénérer » de la manière suivante : un ensemble d'élèves suit la procédure suivante : ils abandonnent le zéro intercalaire ou décalent la virgule. Les erreurs sont donc toutes liées au zéro. Le problème du zéro l'emporte sur tous les autres types d'erreurs. Nous avons bien ici une ressource qui permet d'envisager des développements pédagogiques ciblés.

Ma contribution à la recherche, retient les questions des chemins cognitifs des élèves, mais également celle des compétences et entretient un feu roulant sur le regard accordé à la complexité des niveaux observés. Leur activité est organisée et adaptative, d'où l'importance centrale du concept de schème de Vergnaud (Guiet-Silvain, 2007).

Au-delà du schéma socratique des expériences vécues remémorées et combinées en constructions inédites, je suppose en accord avec Brousseau (1986, p. 48) que l'élève apprend en s'adaptant à un milieu où les réponses nouvelles, traversées de contradictions ou de difficultés toujours possibles, sont bien les preuves de ses apprentissages. L'avantage de ces objets d'étude ne me fait pas oublier que la construction d'une méthodologie de l'analyse implicative reste pertinente, mais néanmoins insuffisante, pour mettre en relief tous les processus dissymétriques. En effet, le foisonnement des sous-classes et des classes de l'arbre hiérarchique oblige le chercheur à déterminer ce qui est significatif à proprement parler. Ce n'est pas une tâche aisée de savoir caractériser les structures et la qualification des groupes, ce qui suppose une part de formation essentielle sur la maîtrise de ces outils statistiques, afin de ne pas assister à des phénomènes d'obsolescence dans la recherche. Sa place est possible en formation au sein des Sciences de l'éducation.

J'étais soucieuse ici de comprendre la connaissance sous une forme d'acquisitions plus ou moins partielles, la photographie de ces arbres implicatifs constituant un édifice créateur : une modélisation de descriptions surprenantes des modalités de calcul engagées formant des familles de comportements. Elles ne sont pas explicatives pour indiquer comment faire progresser les apprentissages, mais favorisent des réponses quant à la façon d'effectuer les traitements des stratégies heuristiques, modélisant certains de leurs raisonnements. Ces cas d'étude apportent une réflexion sur l'organisation des tâches, suivant une typologie, plus ou moins ordonnée, relative à une problématique inscrite au cœur de la didactique, mais également situées au cœur de la réflexion psychologique des apprentissages.

Est concernée également une explication des différences résultant d'un manque de structuration des modalités intégratives. En effet, si la connaissance est inaccessible pour certains individus, elle ouvre les voies de l'évaluation des tâches réussies, ou échouées, afin de repérer les indicateurs de la compréhension de la tâche envisagée. L'organisation des tâches, face émergée de l'iceberg de représentations, est constituée de notions à apprendre, s'organisant des plus simples aux plus complexes. Les niveaux, ici situés dans les analyses hiérarchiques, situent l'intégration et modélisent l'organisation de l'appareil cognitif en sein de systèmes parfois indépendants ou autonomes. Les résistances sont forgées d'attitudes pouvant être communes à un groupe. Ces filiations sont une cascade infinie d'informations dont la hauteur et le cadre confirment les phénomènes observés.

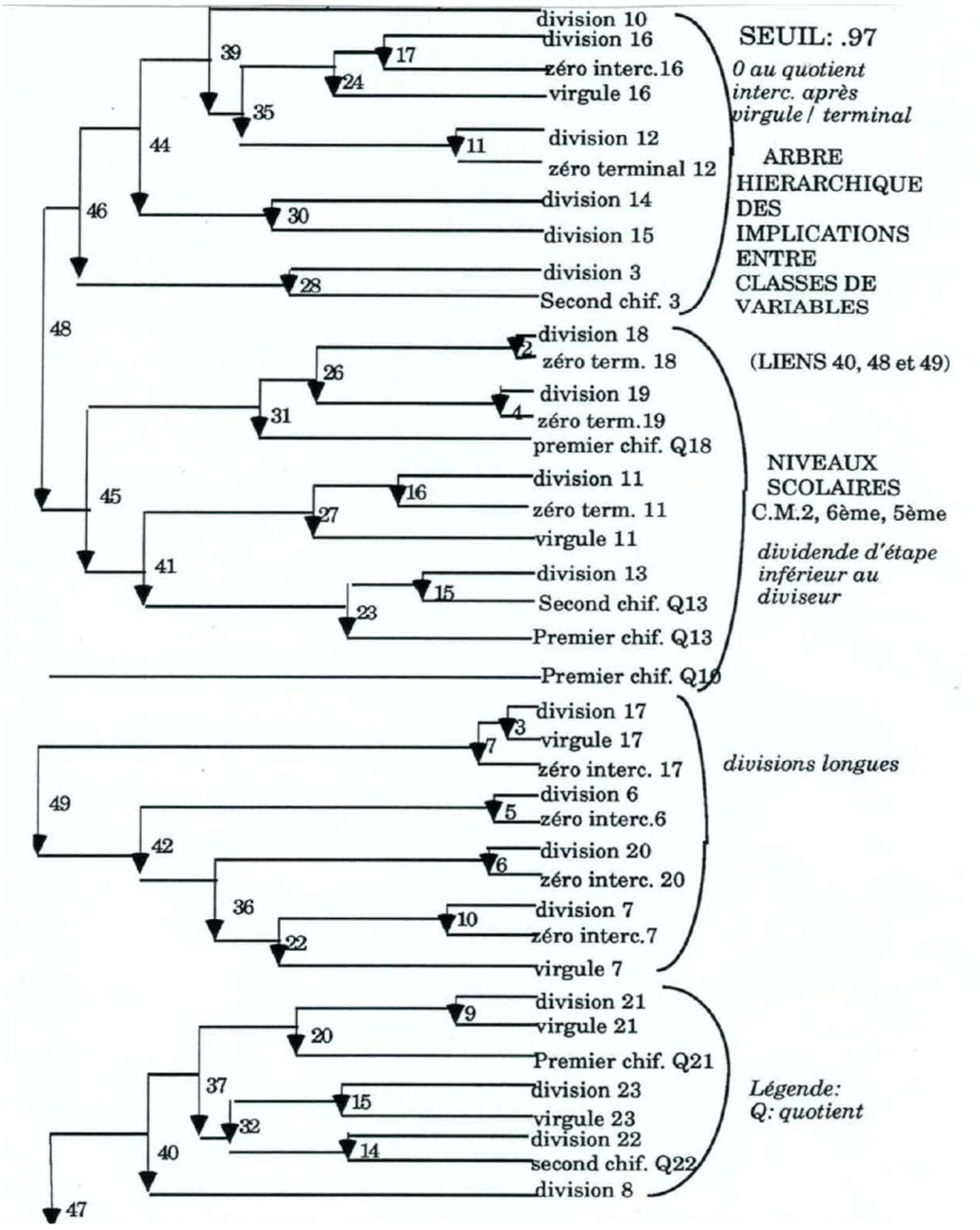


Schéma n° 4 : Exemple d'analyse hiérarchique de classes de variables (Source Thèse 1994)

Je suggère de penser au développement cognitif, constitué de stades successifs, de points d'arrêt, de branches différentielles, décrits par J. Piaget. Je pose une question de recherche avec Gras (2018) qui suit l'hypothèse d'une question piagétienne issue de sa définition théorique de l'équilibration : le développement cognitif piagétien, constitué de stades subissant des discontinuités, dont la schématisation aboutit à des structurations logico-mathématiques, suit-il des niveaux plus généralisables ? Comment comprendre cette question au regard du contrôle organisationnel du sujet ? Et de celui du développement cognitif ? Ce ne sont pas les seules étapes du développement par stades, définies par Piaget, mais bien celles accordées aux structures mentales qui me semblent prépondérantes. Comment un individu intériorise les règles du collectif quand il apprend ? Je retiens l'intérêt de l'état de production de connaissances, enchevêtrée dans le contrôle que le sujet y accorde, ce qui nous souffle des pistes holistiques à explorer.

3.3 L'admissibilité statistique et l'interprétation

La réflexion doit demeurer ouverte dans la recherche du sens qu'il convient d'accorder à la complexité des procédures. Sous son seul aspect premier, cette difficulté peut être considérée comme une attribution négative, renvoyant à une certaine idée de désorganisation, ou d'une cohérence difficile à saisir. Si je suis la position d'Edgar Morin (1990), ce mélange d'ordre et de désordre m'apparaît central : « La pensée complexe vise à la multidimentionnalité », elle va de concert avec la dynamique non linéaire des états possibles d'un système, (diversité, chaos…) et les domaines de connaissance organisationnelles parfois hétérogènes, interdépendantes, antagonistes, paradoxales, ou instables.

Ce modèle peut s'attacher au fonctionnement de l'organisation de l'apprentissage ou de l'éducation, objet d'investigations qui peut faire apparaître la présence ou l'absence de plusieurs systèmes. Leurs caractéristiques peuvent ainsi être appréhendées sous le regard de leur regroupement, de leur dynamique et de leurs niveaux d'échanges d'information, s'appuyant bien sur de multiples réseaux de communication.

Le débat peut être affaibli sous la coupe de positions idéologiques, sa complexité étant liée à la coexistence de plusieurs niveaux d'informations, suivant des chemins croisés et multiples. La complexité du fonctionnement de l'esprit est un terrain propice à l'observation des dimensions cachées du comportement organisationnel du sujet. Cette vision interprétative en Sciences de l'éducation nous apporte toute la créativité des comportements apparemment illogiques. L'enchevêtrement des concepts sert à représenter une part modeste des états et de leur diversité imprévisible. Il existe bien des degrés dans leur interprétation. Soulignons par là-même l'outil entrant dans l'analyse de la construction de la pensée de l'enfant, elle-même en perpétuelle évolution. De plus, la transposition didactique, d'après Brousseau (1986), permet de relever comment les savoirs chaotiques se constituent -surgissant du sens premier de tissus d'activités croisées- et s'adaptent à des circonstances ciblées.

Les chemins significatifs de la structure hiérarchique obtenue semblent montrer le fonctionnement interne du savoir, autrement dit un système structuré de la conduite. Il s'agit ainsi de rechercher des liaisons dans un ensemble de caractères et de classes de caractères, afin de mettre en évidence une structure cognitive stable chez un ensemble

d'élèves de niveau scolaire équivalent (Gras, 1992)[22]. Cette méthode donne des modèles explicatifs aux types d'adaptation du cheminement des processus de l'enfant. Cette activité s'organise par rapport à un but, celui-ci étant jugé utile ou pas. Le choix des caractéristiques des divisions présentées aux élèves repose bien sûr sur l'hypothèse de leur influence sur les propriétés mathématiques en jeu dans l'algorithme.

L'effort de construction scientifique, ici mené, englobe des éléments qui permettent de reconnaître ce sujet inépuisable dont les choix méthodologiques, même les plus féconds, ne peuvent pas rendre compte de l'abondante complexité des hypothèses (Vergnaud, Durand ; 1976). Afin de poursuivre cette réflexion, j'ai pu donner suite à une publication dans un colloque dédié aux analyses implicatives à Caen (1995)[23] et également dans un séminaire organisé par G. Vergnaud à l'Université Paris V (1996)[24]. J'emprunte la notion d'arbre logique à Astolfi (1990) présentant une trame conceptuelle : le caractère intégrateur de notions subordonnées suppose une organisation et une implication logique des contenus. Deux publications, en 1995[25] et 1997, à la suite de communications précédemment citées, ont valorisé mes recherches issues de la soutenance de thèse, l'une dans le cadre des *Actes du Colloque de Caen,* IRMAR (Institut de Recherche Mathématique de Rennes I), l'autre pour la revue *Carrefours de l'Education.* Les divisions que nous supposons les plus complexes sont celles qui comportent un dividende d'étape inférieur au diviseur, un peu moins complexes celles dont le nombre de chiffres du quotient est supérieur à 3, celles enfin qui ont un quotient décimal. Leurs caractéristiques et les calculs engagés, sont arborescents, chaque niveau de fonctionnement étant constitutif d'un ensemble emboîté de morceaux de puzzle. Ces caractéristiques ont permis d'analyser les niveaux de fonctionnement, les procédures utilisées, et les sources de réussite et d'efficacité. Les résultats n'avaient pas été envisagés pour expliquer l'ensemble des comportements observés, mais avaient pour objet de répondre aux assertions suivantes :

- les critères de ressemblance des divisions entre elles éclairent la façon dont les élèves en disposent, avec les techniques de calcul qu'ils utilisent ;
- le fait de contracter dans une même opération multiplication et soustraction nous permettent par extension, de cerner les avantages et les désavantages de l'enseignement lorsque l'apprentissage de l'algorithme se fait directement avec la soustraction non posée;

---

[22] Dans son article "L'analyse des données", dans la revue Recherches en Didactique des Mathématiques (Vol.12, n°1, pp. 59-72, 1992), Gras souligne que l'examen d'une telle relation entre deux classes n'ayant véritablement un sens que dans le cas d'une "bonne fermeture" des classes, le concept de cohésion d'une classe est défini comme antinomique de celui de "désordre implicatif" (au sens de l'entropie dans la théorie de l'information). De là découle la recherche d'implications entre deux classes bien "cohésives", i.e. déjà ordonnées en leur sein.

[23] Guiet J., (Janvier 1995) : *Colloque "Méthodes d'analyses statistiques multidimensionnelles en didactique des mathématiques*", ARDM (Association pour la recherche en didactique des mathématiques), I.U.F.M. de Caen : "Analyses implicatives appliquées à une recherche : cas de la division".

[24] Guiet, J., (1996). "Sens des méthodes statistiques implicatives : étude de cas". *Séminaire de Recherche "Psychologie cognitive et didactique* », organisé par Vergnaud G, Université Paris V.

[25] Guiet, J., (1995), "Analyses implicatives appliquées à une recherche : cas de la division", *Actes du Colloque de Caen,* IRMAR (Institut de Recherche Mathématique de Rennes), Université de Rennes I.

- la compréhension de l'opération n'est pas acquise en fin d'école élémentaire et son extension aux décimaux est problématique, à cause des problèmes de positionnement de la virgule au quotient notamment ;
- enfin, la présence d'un zéro intercalaire ou terminal au quotient apporte des difficultés spécifiques.

3.4 Traitement informatique retenu et dépouillement

En raison de la complexité de la recherche, il m'est apparu plausible de la compléter avec des traitements issus des analyses factorielles de correspondances[26] afin de préciser les résultats obtenus. Un logiciel nommé CHADOC, créé par une équipe de recherche de l'I.U.T de Nice, m'a permis d'effectuer des analyses de données sur des tableaux statistiques importants (plusieurs milliers d'individus et plusieurs centaines de variables). Ce même logiciel a été conçu pour réaliser entre autres, des analyses factorielles des correspondances en composantes simples et en correspondances multiples. Pour l'analyse des composantes simples, les données de base retenues sont les variables concernant les caractéristiques des divisions (virgule, zéro intercalaire, etc.), et les modalités de réussite et d'échec de ces opérations, que j'avais choisies, susceptibles d'être corrélées entre elles.

Les situations comprennent un certain nombre de facteurs et leurs multiples interactions. Je réduis et confirme ici les résultats de mes interprétations.

- La complexité du positionnement du zéro dès que le dividende d'étape est inférieur à un diviseur à deux chiffres.
- Les contraintes supplémentaires lors de la résolution de l'opération, imposées par la soustraction non posée.
- La représentation de la division comme problématique d'une recherche du nombre de parties dans un tout (c'est-à-dire un dividende contenant le diviseur) prédomine dans notre population.

L'interdépendance des phénomènes souligne bien cette capacité d'abstraction, supérieurement délicate, qui n'est réservée en fait qu'à ceux qui maîtrisent de façon éminente les processus requis par la résolution elle-même (par exemple la soustraction, qu'elle soit mentale ou posée, peut difficilement supporter d'autres difficultés, telles les longueurs de certains calculs).

3.5 Deuxièmes éléments d'analyse

Il apparaît ici que les avantages et les inconvénients des approches statistiques sont limités aux départs d'appréciations et de réflexions empiriques immédiatement validées par l'interprétation des intrications des problèmes donnés. Ces méthodes approchent des réalités plus qu'elles ne les expriment, elles ne sont pas simples. Elles subissent des contraintes, qui ne sont pas ici identiques mais elles nous servent de point de départ pour analyser les problèmes qui motivent l'introduction des concepts en cause, et par extension les processus d'enseignement.

[26] Ces résultats ont également été également présentés dans les colloques et séminaires de 1995 et 1996.

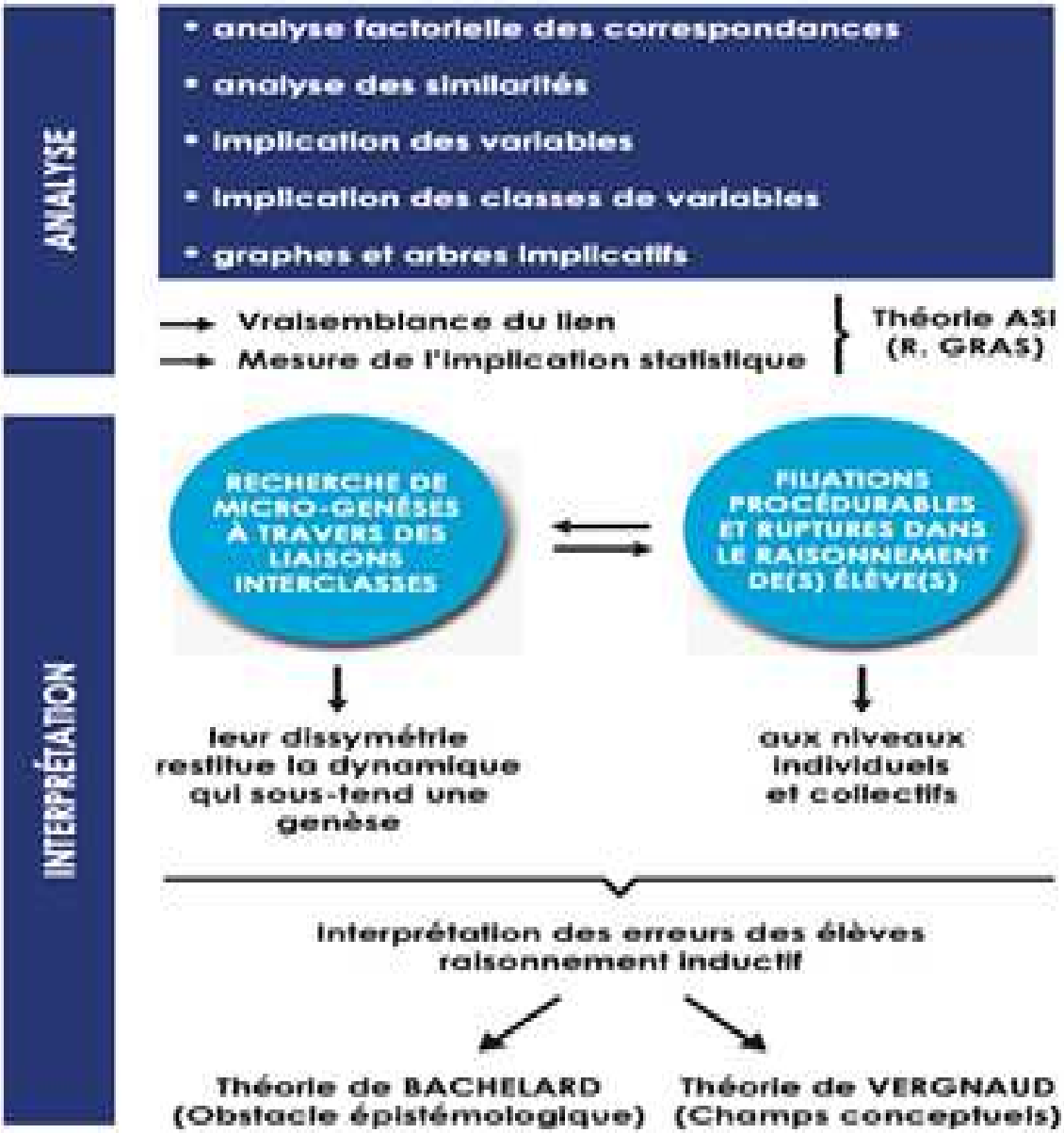


Schéma n°5 : Généralisation du cheminement méthodologique

Le repérage du niveau de complexité des tâches proposées et l'identification de leur spectre permet d'éclairer les difficultés rencontrées par les élèves et les problèmes de conceptualisation (Gras et Acioly-Régnier, 2017). Le fait de poser la soustraction favorise la réussite aux divisions d'une façon générale. Seuls, les enfants qui maîtrisent l'algorithme sont capables d'effectuer une soustraction mentale qui suppose des capacités d'abstraction importantes. Les niveaux d'acquisition d'un concept y sont ainsi ordonnés par la complexité : toute abstraction est ici quantitativement caractérisée comme étant un obstacle. Nos observations sont extensives à d'autres types d'études supposant d'autres échantillons. Nous n'avons pas choisi de les poursuivre après la thèse. Nos résultats montrent que la suite d'opérations que comporte une division est effectuée sans que le sujet comprenne les raisons de ce qu'il fait. Avec une répétition fréquente cependant, on arrive à comprendre, les conditions d'une vraie automatisation étant non réunies.

L'algorithme de la division, s'il se raffine de conditions particulières, comme les zéros au quotient, par exemple, va engendrer des actions spécifiques. L'idée d'obtenir ce qui se « dégénère » dans l'usage, est la preuve que l'algorithme n'est pas acquis, le schème quant à lui, est celui qui réussit. L'ensemble des classes d'erreurs (et des relations dissymétriques entre elles) de l'algorithme de la division écrite, recueillies grâce aux méthodes d'analyse de données telles l'analyse hiérarchique, l'analyse implicative et l'analyse factorielle de données, donne des résultats qui s'approfondissent dans une perspective d'enrichissement didactique.

4 Conclusion

Les progrès de la statistique permettent de travailler sur plusieurs variables, un facteur n'intervient pas de manière isolée (Gras, 1976). Je simplifie ici ma conclusion en disant que ces études statistiques permettent de dépasser les simples constats de relations entre variables pour établir de réels liens de causalité. C'est évidemment leur intérêt principal. Une réflexion sur les méthodes qualitatives n'exclut pas d'appréhender les êtres et les situations dans leur réalité complexe et multiforme que nous ne saurions également ignorer. La complexité humaine (Benkirane, 2002) nous fait comprendre les moments : elle fait système et nous permet d'objectiver la complexité d'un niveau cognitif, au sein dans un champ conceptuel donné. Celui-ci semble profondément attaché aux opérations de pensée dont dispose le sujet interrogé à un moment donné, sa familiarité avec les concepts en jeu, leur intrication, les théorèmes en acte[27] et leur efficience et l'interdépendance de leurs emplois. Les lecteurs me pardonneront de ne pas prendre en compte les aspects nouveaux que les recherches ont apporté dans ce domaine des ASI. Il me semble inutile de reprendre ici, une vingtaine d'années après, l'abondance des articles qui sont parus sur ce sujet. L'enjeu étant seulement ici de faire état de ma recherche en son temps. Encore faut-il imposer une distinction : du point de vue du savoir, la mémoire peut être encombrée d'informations éparses, l'héritage du droit à la réflexion est certainement une revendication en direction de l'exigence du regard pluriel que nous devons y accorder. Entrer dans l'épistémologie des algorithmes complexes, suppose de dire que l'organisation cognitive trouve les principes d'organisateur commun. Les gymnastiques de l'esprit nous imposent une conscience aigüe de leurs dérives, inattendues et parfois éphémères. Elles conduisent des trajets supposant des structurations mentales qui construisent et déconstruisent un savoir (Piaget, 1975, *op.cit.),* permettant de comprendre ce que sont les ruptures et les obstacles épistémologiques (Bachelard, 1934).

Enfin, je marquerai à l'encre rouge l'excellence de la portée de l'œuvre théorique de Régis Gras (1979). Celui-ci a non seulement défini les niveaux d'objectifs hiérarchisés, leur degré et leur intelligibilité, mais aussi les sous-niveaux les constituant et les exercices censés les mettre en œuvre. Je lui renouvelle le remerciement de son aimable attention à mon égard, et plus loin encore ma reconnaissance pour son hospitalité au sein des méthodes implicatives de données. Il me semble indispensable que la recherche en Sciences de l'éducation favorise l'expansion de ces démarches méthodologiques

[27] Les théorèmes-en-acte sont des propositions implicitement tenues pour vraies dans le réel et permettent de les caractériser et de trouver dans les situations rencontrées les conditions de leur fonctionnement (Vergnaud, 1996).

d'implication ou de quasi-implication. Je propose dans cette perspective de reconnaître la place à y accorder. Ceci suppose un regard exercé, une implantation dans les cursus des étudiants, tous deux conjugués en association avec d'autres disciplines. Ces travaux sur l'apprentissage d'une opération abstraite dans le cadre de l'école ont fait naître chez moi le désir de diversifier les investigations. Mon intérêt premier demeurait la volonté d'aborder les résistances à l'apprentissage, les obstacles épistémologiques, les illusions nées des intuitions de pensée ou des certitudes pratiques. Mon désir en revanche était de ne plus me limiter à l'espace et au temps de l'école, d'autant que la confrontation des étudiants me conduisait à d'autres situations pédagogiques et à d'autres objets de savoir. C'est donc vers de nouveaux apprentissages scientifiques, et de nouveaux apprenants que sont allés mes investissements de recherche (Guiet-Silvain, 2007a, 2007b). Ce sont vers de nouveaux types d'obstacles enfin que sont allées mes tentatives de vérification. D'autres travaux, où figurent des problèmes d'adaptation au calcul, sans doute, mais où figurent, quoi qu'il en soit, des apprentissages, où figurent aussi, ce qui n'est pas toujours le cas pour l'élève, des problèmes d'application des savoirs. Ceci a permis des orientations dans d'autres champs, notamment ceux relatifs au domaine de l'éducation à la santé (Guiet-Silvain, 2011a, 2011b, 2020). Les méthodes d'analyse implicatives y trouveraient également des pistes de recherche passionnantes pour l'analyse des comportements cognitifs.

Références

[1] Ag Almouloud. Saddo (1992). « L'ordinateur, outil d'aide à l'apprentissage de la démonstration et de traitement de données didactiques », soutenue sous la direction de R., Gras, *Thèse de doctorat en Mathématiques et Applications*, Université de Rennes I.

[2] Artigue, M., Gras, R., Laborde C., Tavignot P., Balacheff, N., Dir. (1994). *Vingt ans* de didactique des mathématiques en France : *Hommage à Guy Brousseau et Gérard Vergnaud*, La Pensée Sauvage éditions : Grenoble.

[3] Astolfi, J-P. (1990). Les concepts de la didactique des sciences, des outils pour lire et construire des outils d'apprentissage In : Les professions de l'éducation : *Recherche et formation*, n°8, 19-31. https://www.persee.fr/doc/refor_0988-1824_1990_num_8_1_1021

[4] Bachelard, G. (1934). *Le nouvel esprit scientifique*, Paris :PUF.

[5] Benkirane, R. (2002), *La complexité, vertiges et promesses*. Paris : Le Pommier.

[6] Brousseau, G. (1998). *Théorie des situations didactiques*. Grenoble : La Pensée Sauvage Editions.

[7] Brousseau, G., (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en didactique des mathématiques*. 7.2 33-115.

[8] Chevallard, Y. (1985/1991). *La transposition didactique. Du savoir savant au savoir enseigné*. Grenoble : La Pensée Sauvage, Editions.

[9] Conne F., Guiet J. (1989). Situations évoquées, situations jouées et structures mathématiques, *Actes de la Vème Ecole d'Eté de Didactique des Mathématiques,*

Plestin les Grèves, 51-57. http://www.numdam.org/item?id=PSMIR_1989___S6_51_0

[10] Conne F., Lemoyne G. (1999). *Le cognitif en didactique des mathématiques*, Montréal : Presses de l'Université de Montréal.

[11] Flückiger, A. (déc. 2000). Genèse expérimentale d'une notion mathématique : la notion de division comme modèle de connaissances numériques. *Thèse, faculté de Psychologie et de Sciences de l'Éducation*, Université de Genève

[12] Gras, R., (2018), *La théorie de l'analyse statistique implicative, ou l'Invraisemblance du faux*. Toulouse : Cépaduès Ed.

[13] Gras R., Acioly-Régnier J-C. (2017), Hiérarchie de règles en A.S.I. et conceptualisation, in R. Gras et al. (Ed) *L'Analyse Statistique Implicative, des Sciences dures au Sciences Humaines et Sociales*, Toulouse : Cépaduès, , 369-386.

[14] Gras, R., Bodin, A., (4-7 octobre 2017). L'A.S.I., Analyseur et révélateur de la complexité cognitive taxonomique *IX colloque international, ASI Analyse statistique implicative*, Belfort.

[15] Gras, R., Kuntz, P., Briand H. (2001). Les fondements de l'analyse Statistique implicative et quelques prolongements pour la fouille de données », *Mathématiques & Sciences humaines*, 39e année, n° 154-155, 9-29.

[16] Gras, R., Almouloud, S-AG., Bailleul, M., Larher, A., Polo, M., Ratsimba-Rajohn, H., Totohasina, A., (1996). *L'implication statistique, nouvelle méthode exploratoire de données. Applications à la didactique*. Grenoble :La Pensée Sauvage éditions.

[17] Gras, R., Totohasina, A., Ag Almouloud, S., Ratsimba Rajohn, H., Bailleul, M. (1994). La méthode d'analyse implicative en didactique. Applications, In *Vingt ans de didactique des mathématiques en France : Hommage à Guy Brousseau et Gérard Vergnaud.* Grenoble : La Pensée Sauvage éditions.

[18] Gras, R., Lahrer, A., (1992). L'implication statistique, une nouvelle méthode d'analyse de données. *Mathématiques, Informatique et Sciences humaines*, n°120, 1993, 5-31.

[19] Gras, R., Fontaine, M-D., Larher, A., Nicolas, A., Simon, A. (1989). Le micro-ordinateur, outil de révélation et d'analyse du raisonnement déductif de l'élève. *Publications mathématiques et informatique de Rennes*, n°5.

[20] 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'État, Université de Rennes 1.

[21] Gras R. (1976), Recherche d'une taxonomie d'objectifs cognitifs en Mathématiques. *I.R.E.M.* de Rennes.

[22] Guiet-Silvain J. (2020). « La communication des réseaux sociaux des adolescents : une domination de soi ? » Vol. 6-2. *Revue Education, Santé Société.*

[23] Guiet-Silvain, J. *et al.,* (2011a). « Éducation à la santé en milieu scolaire, mise en perspective historique et internationale », *Carrefours de l'éducation,* 32, 105-127.

[24] Guiet-Silvain J. (coord.) (2011b), (Dir.) « Education à la santé : convergences ou émiettements ? », *Carrefours de l'Education*, C.R.D.P. d'Amiens.

[25] Guiet-Silvain, J. (2007a). Gérard Vergnaud (né en 1933), *Psychologie de l'Education,* coord. M. Merri, M. Pichat, 61-63, Bréal.

[26] Guiet-Silvain, J. (2007b). Approches contrastées de l'erreur, *Psychologie de l'Éducation* coord. M. Merri, M. Pichat, 168-171, Bréal.Merri, M. Pichat, 74-78, Bréal.

[27] Guiet, J., (1997). « La conceptualisation : cas d'erreurs dans l'opération de division », *Carrefours de l'Education*, Centre Universitaire de Recherche en Sciences de l'Education de Picardie, C.R.D.P. de l'Académie d'Amiens, 2- 24.

[28] Guiet J., (1994), Algorithmes et schèmes : cas de la division". M. Artigue, R. Gras, C. Laborde, P. Tavignot (Eds), *Vingt ans de didactique des mathématiques*, Grenoble La Pensée Sauvage, 225-230.

[29] Guiet, J. (1994). *La division une longue souffrance* Doctorat de Sciences de l'Education, Université Paris 5.

[30] Guiet, J. (1993). L'algorithme de la division est-il un algorithme ? *Publications de IREM de Rennes*, 1993, fascicule 3 didactique des mathématiques. 97-102.

[31] Larher, A. (1991). *Implication statistique et applications à l'analyse de preuve mathématique*. Thèse de Doctorat

[32] Lemoyne, G., Conne, F., J. Brun (1993). Du traitement des formes à celui des contenus d'écritures littérales : une perspective d'enseignement introductif de l'algèbre. *Recherches En Didactique Des Mathématiques*, 13(3), 333–384.

[33] Lerman, I. C., Gras, R., H. Rostam, H., (1981). Élaboration d'un indice d'implication pour des données binaires I, *Mathématiques et sciences humaines* (74), 5–35.

[34] Piaget, J. (1975). *L'équilibration des structures cognitives*. Paris : PUF.

[35] Vergnaud, G. (2011/1). La pensée est un geste ou comment analyser la forme opératoire de la connaissance. *Enfance*, n° 1, 37-48.

[36] Vergnaud, G. (1996). La théorie des champs conceptuels In J. Brun (Ed.), *Didactique des mathématiques* (pp. 197-242). Lausanne : Delachaux & Niestlé.

[37] Vergnaud, G. (1994). Multiplicative Conceptual Field. What and Why"". in G. Harel, J. Confrey (Eds). *The Development of Multiplicative Reasoning in the learning of Mathematics*. Albany State, University of New York Press

[38] Vergnaud, G. (1988). Question de représentation et de formulation dans la résolution de problèmes mathématiques. *Annales de Didactique et de Sciences Cognitives* (1), 33-35.

[39] Vergnaud, G. (1985). Concepts et schèmes dans une théorie opératoire de la représentation. *Psychologie française*, Tome 30, 3.4, 245-252.

INTERACTIONS ENTRE ÉLÈVES DANS LA CLASSE NON MAITRISÉES PAR L'ENSEIGNANT. Étude exploratoire par une enquête par questionnaire auprès d'enseignants de l'école primaire

Océane Touron[1], Jean-Claude Régnier[2],

PEER INTERACTIONS NOT MASTERED BY THE TEACHER IN PRIMARY SCHOOL CLASSES EXPLORATORY STUDY WITH INQUIRY SURVEY OF PRIMARY TEACHERS

RÉSUMÉ

Cet article a pour objectif d'interroger le rapport des enseignants aux interactions entre pairs qu'ils ne maitrisent pas dans leurs classes, notamment quant à la place qu'ils font à ces interactions dans leurs propositions pédagogiques. Après un rapide retour sur la question des interactions au sein de la classe, nous présenterons le traitement des premiers résultats d'une enquête en ligne menée auprès d'enseignants de l'école primaire grâce à l'analyse statistique implicative. Nous verrons comment l'interprétation de ces résultats nous permet d'identifier deux principaux profils d'enseignants, qui se distinguent dans leur rapport très favorable ou au contraire plutôt défavorable à ces interactions.

Mots-clés : *Interaction, apprentissage, cadre, maitrise, pédagogie, choix, placement*

ABSTRACT

This article is intended to question the relation of primary teachers to peer interactions that they do not master during learning situations, especially regarding the space they make for these interactions in their educational proposals. We will first be promptly interested in interactions in the schoolroom. Secondly, we will present the data processing of an online survey conducted with teachers, thanks to statistical implicative analysis. We will finally show how these results interpretation enable us to identify two main teacher profiles, which are distinguished by a strongly favorable relation to those interactions in one hand and a rather unfavorable relation to those interactions on the other hand.

Keywords*: interaction, learning, environment, control, pedagogy, choice, positioning*

[1] Doctorante en sciences de l'éducation et de la formation. ED 485 EPIC Université de Lyon. Laboratoire UMR 5191 ICAR. oceane.touron@univ-lyon2.fr

[2] Professeur des universités émérite. UMR 5191 ICAR http://www.icar.cnrs.fr/membre/jcregnier/, Université Lumière Lyon 2. Professeur invité à l'Université d'État de Tomsk (Russie) – National Research Tomsk State University – jean-claude.regnier@univ-lyon2.fr

1 Introduction

De nombreux éducateurs et parmi eux, nombre d'enseignants s'interrogent sur leur propre posture en classe, en relation à la place de l'élève en tant qu'apprenant autour duquel s'organisent les situations d'enseignement et d'apprentissage et la vie même de la classe. Ils s'interrogent sur les questions de liberté de mouvement et de choix dont cet élève peut disposer au sein de l'espace scolaire. Des choix pédagogiques faits par les enseignants découle la construction d'une infinité de situations d'enseignement et d'apprentissage, donnant lieu à des modalités de travail variées et variables ainsi qu'à l'utilisation de supports très différents les uns des autres, mais aussi à des interactions plus ou moins nombreuses et riches entre l'adulte et l'enfant, entre l'enseignant et l'apprenant ou entre les élèves, c'est-à-dire entre pairs. Dans l'imaginaire collectif, les interactions entre pairs sont traditionnellement plutôt réprimées à l'école afin de favoriser les modalités de travail individuel ; on pense généralement l'enseignement scolaire comme un travail silencieux et statique de l'élève. Néanmoins, les conceptions éducatives prônant des pratiques pédagogiques favorisant des interactions entre pairs ont progressivement été davantage perçues dans la communauté des enseignants. Ainsi ce qui fut appelé les pédagogies dites alternatives ou actives, datant majoritairement de la fin du XIXème et du début du XXème siècle, semblent canaliser un plus grand intérêt à notre époque. Force est de constater que les pratiques favorisant les échanges et le mouvement sont aujourd'hui davantage encouragées par les programmes et directives de l'Éducation Nationale, qui font une plus grande place aux interactions et aux variations de modalités de travail dans la classe dès l'entrée à l'école. Les dernières modifications des programmes d'enseignement de l'école maternelle précisent que « [L'enseignant] favorise les interactions entre enfants et crée les conditions d'une attention partagée, la prise en compte du point de vue de l'autre en visant l'insertion dans une communauté d'apprentissage. » (Ministère de l'Éducation Nationale, de la Jeunesse et des Sports, 2020)

Dès lors que l'enseignant est lui-même encouragé à faire varier les modalités qu'il propose au cours des situations d'enseignement-apprentissage dans sa classe, nous pouvons imaginer qu'il sera plus enclin à susciter des interactions entre les élèves. Le travail en groupes, en binômes ou trinômes, et même la très classique disposition des tables d'élèves en rangées, donnent lieu à des échanges. Mais tous ces échanges ne peuvent pas être totalement maîtrisés par l'enseignant. En effet, il paraît cohérent d'admettre que l'adulte, souvent seul face à un groupe de vingt à trente élèves en moyenne, ne peut ni tout voir ni tout entendre. Il est donc intéressant de s'interroger sur ce qui se passe vraiment lors de ces échanges qui échappent à l'attention de l'enseignant et qui peuvent être tant encouragés que redoutés, voire réprimés, par le professionnel dans sa pratique d'enseignement. Lorsque deux élèves échangent pendant une séance de travail individuel, quel intérêt trouvent-ils, qui participe du processus d'apprentissage à l'œuvre ? Echangent-ils uniquement pour se faciliter la tâche dont la résolution est demandée par l'enseignant ou encore par manque d'intérêt pour le travail qui leur est proposé ? Qu'en est-il des propositions pédagogiques où les échanges libres entre élèves font partie intégrante de la vie de la classe ? Quelles conséquences ces échanges ont-ils réellement sur les apprentissages en tant que résultats du processus d'apprentissage ? Ce sont ces interrogations qui ont orienté la construction de notre objet de recherche, et nous ont amené à nous intéresser à la place et au rôle des interactions entre pairs au cours des

situations d'enseignement et d'apprentissage à l'école primaire organisées par l'enseignant dans la classe dont il a la responsabilité.

2 Retour sur la question des interactions au sein de la classe

Les recherches sur les interactions humaines dans la classe sont nombreuses, notamment au sujet de la relation maitre-élève. Les relations et interactions entre élèves sont plutôt étudiées en lien avec des problématiques sociales vives, telles que les inégalités socioéconomiques, l'immigration ou le handicap, ou encore dans le cadre des pédagogies basées sur la collaboration ou la coopération. On peut remarquer des travaux sur cette problématique, notamment autour des pédagogies coopératives, de la notion de tutorat (Berzin, 2005) et dans le domaine des mathématiques (Schubauer-Leoni et al., 1985), souvent appliqués au travail en groupes (Pléty, Cosnier, 1985). On trouve également de nombreux travaux de recherche sur l'impact des interactions entre pairs dans l'acquisition des compétences langagières. De manière générale, il ressort des recherches sur cet objet que ces interactions sont largement favorables aux apprentissages tout au long de la scolarité (Cuisinier et al., 2007).

L'objet de notre étude s'inscrit donc dans la lignée de ces recherches sur les interactions et la coopération au sein de l'espace scolaire. Il se dégage des problématiques sociales, mais insiste sur un type d'interaction particulier, interactions désignées comme « non-maitrisées par l'enseignant ». Ces interactions horizontales ont un statut qui peut être qualifié d'« illégalisme scolaire » (Sirota, 1988), en cela qu'elles sont souvent réprimées et qu'elles sont peu ou pas prises en compte dans la planification des apprentissages, en particulier dans des conceptions pédagogiques dites traditionnelles. Ce statut bancal s'explique en partie par l'imprévisibilité de ces interactions du point de vue pédagogique et didactique. En situation de recherche, il en résulte la nécessité d'adapter les grilles d'observation à cette dimension imprévisible (Vasquez, Martinez, 1990).

Le lien entre apprentissage et interaction est établi, notamment dans le cadre de l'apprentissage par imitation, dont les mécanismes sont particulièrement à l'œuvre chez le petit enfant imprégné de son milieu. De nombreuses études sur le rôle des neurones miroirs dans le développement cérébral et émotionnel de l'enfant (Gueguen, Ansembourg, 2014) montrent que ce dernier a besoin de l'intervention de l'autre pour apprendre et se structurer. Ces recherches tendent à montrer l'intérêt de développer un enseignement faisant une plus grande part aux interactions entre élèves et à la posture d'observateur chez l'enseignant, tel que le proposent de nombreux pédagogues depuis plus d'un siècle selon des orientations différentes qu'il conviendrait de préciser (Freinet, Decroly, Montessori, entre autres). Plusieurs études s'intéressent au lien entre les représentations des enseignants et les pratiques pédagogiques qu'ils mettent en œuvre et tendent à attester de ce lien de causalité (Cuisinier et al., 2007). Afin de tenter de prendre la mesure de l'impact des interactions entre pairs non-maîtrisées par l'enseignant, sur les apprentissages des élèves, il parait donc indispensable de s'intéresser en premier lieu à la manière dont ces interactions sont perçues, identifiées et traitées par l'enseignant en charge de l'organisation pédagogique de la classe.

Dans le cadre de cette recherche, nous avons donc mené auprès d'enseignants du premier degré une enquête visant à établir un panorama de leurs pratiques et de leurs conceptions quant aux interactions entre pairs qui ont lieu dans leurs classes et dont ils

n'ont pas directement la maîtrise. Quels sont les choix pédagogiques de l'enseignant ? Au cours des situations de classe, quelle posture adopte l'adulte, quelle présence auprès des élèves et quelles interactions avec eux ? A quel degré les échanges entre élèves sont-ils permis, dans quel cadre et sous quelle forme ? Ces interactions sont-elles encouragées, réprimées ? L'enseignant tente-t-il de toujours les maîtriser ou donne-t-il de l'autonomie aux élèves dans ces interactions ? Nous tenterons donc de mettre en lien les choix pédagogiques des enseignants et la manière dont ils considèrent les interactions entre élèves non maîtrisées par l'enseignant.

Nous chercherons ici plus spécifiquement à savoir si les interactions entre pairs non-maitrisées par les enseignants sont considérées ou non comme constructives pour les apprentissages par ceux-ci, et quelle place ils laissent à ces interactions dans leurs choix pédagogiques. Notre hypothèse de travail est la suivante : plus il y a d'interactions entre pairs non maitrisées par l'enseignant dans une classe, plus la réaction de l'enseignant tendra à les permettre et à les inclure dans le processus d'apprentissage. Nous supposons que ce sont les choix pédagogiques faits par l'enseignant qui créent de l'espace dans la classe pour que ces interactions aient lieu. Nous nous attendons cependant également à voir apparaitre un autre profil d'enseignant, qui considèrerait les interactions entre pairs qu'il ne maitrise pas comme plutôt dommageables aux apprentissages dans la classe, ce qui pourrait alors être interprété comme une problématique plutôt de l'ordre de la discipline que de l'acquisition de compétences par les élèves.

Notre objectif est dans un second temps de mettre en lien les résultats de cette analyse avec d'autres variables de l'enquête afin de tenter de dégager un ou plusieurs profil(s) majoritaires et d'interpréter de façon plus fine les croisements de variables effectué en premier lieu. Nous chercherons ainsi à déterminer si des styles pédagogiques marqués se distinguent chez les enseignants sondés selon qu'ils sont plutôt en début ou en fin de carrière, jeunes ou moins jeunes, avec un seul niveau de classe, deux niveaux ou plus. Nous nous intéresserons particulièrement à la distinction entre âge de l'enseignant et nombre d'années d'ancienneté. Il est clair qu'un enseignant relativement jeune peut difficilement avoir plusieurs années d'expérience professionnelle à son actif. En ce qui concerne les enseignants les plus âgés, s'ils ont en général plusieurs années d'expérience professionnelle, on ne peut exclure une entrée tardive dans le métier qui en fait un novice. Nous cherchons à identifier le rôle de ces deux caractéristiques dans les représentations de la place accordée aux interactions entre pairs non maitrisées par l'enseignant en classe.

3 Approche méthodologique de la construction des données

Les données dont nous disposons ici ont été construites au moyen d'une enquête par questionnaire mis en ligne par l'intermédiaire de l'application LimeSurvey (version 3.17.5) disponible sur l'environnement numérique de travail du site de l'Université Lumière Lyon2. Le lien du questionnaire a été transmis par courriel sur des listes de diffusion destinées à des enseignants et auxquelles nous avons eu accès directement ou même indirectement. L'échantillon est alors constitué par les individus ayant complété partiellement ou complétement le questionnaire. Nous ne chercherons pas à évaluer le degré de représentativité de cet échantillon qui demeure un échantillon de convenance non assimilable à un échantillon aléatoire. Dans ce domaine d'étude et dans les circonstances au sein desquelles nous nous sommes retrouvés, il ne nous a pas été possible d'aller plus loin dans le respect des exigences méthodologiques requises par la définition

stochastique de la représentativité d'un échantillon vis-à-vis de la population dont il est extrait. Toutefois, nous pouvons en donner quelques caractéristiques qu'il est possible de comparer à des données issues des statistiques officielles du Ministère de l'Éducation nationale ou de l'INSEE (voir tableau DEPP 2021 plus bas).

Le questionnaire comporte globalement 25 questions. Nous retenons 14 de ces questions, identifiées comme étant les plus pertinentes au vu de notre problématique de recherche et exploitables dans le cadre de cette analyse. Celles-ci couvrent les thématiques et visent les objets suivants :

Code	Thématique/Objet des questions
V001	Sexe
V002	Age (calculé au jour du dépôt du questionnaire à partir de la date de naissance)
V003	Ancienneté professionnelle
V004/5/6	Organisation pédagogique de la classe
V007/8/9	Place des élèves dans la classe
V010/…/16	Disposition spatiale de la classe
V018/…/21	Répartition des élèves lors des travaux en groupe
V022/…/24	Constitution de la classe par niveau scolaire
V025	Statut pédagogique des interactions entre élèves dans la classe
On parle d'échanges entre élèves. Ces échanges peuvent être verbaux, gestuels, documentaires, ou de l'ordre de l'observation (un élève montre à un autre). On entend par « maitriser les échanges » le fait d'en connaître au moins partiellement le contenu ou l'objet.	
V026 V027	Fréquence des interactions entre élèves sans la maîtrise du contenu par l'enseignant
V028/…/V033	Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre de …
V034	Réaction de l'enseignant lorsqu'il constate les échanges non maîtrisés entre les élèves
V035/…/42	Nature, place et rôle des échanges entre élèves pendant les temps d'enseignement

Tableau 13 : Thématiques et objets des questions et codage

Ces 14 questions ont été ainsi formatées et modélisées selon les variables ou vecteurs-variables statistiques suivants.

Code de la question	Format de la question	Modélisation statistique
V001	fermée	variable qualitative nominale
V002, V003	ouverte	variable quantitative continue
V004/5/6 V007/8/9 V010/…/16 V018/…/21 V022/…/24 V028/…/V033 V34 V035/…/42	fermée	variable qualitative nominale vecteur-variable à composantes binaires
V025	fermée	variable qualitative nominale
V026 V027	fermée	variable qualitative ordinale

Tableau 2: format des questions et modélisation statistique

À ce jour notre base de données est constituée de 60 individus enseignant à l'école primaire.

Caractéristiques	Age moyen (écart-type)	Enseignante	Enseignant
Données échantillonnales	39 ans (12 ans)	48/60 (80%)	12/60 (20%)
Données officielles	Secteur public 1er degré: 43 ans 2nd degré : 45 ans Secteur privé sous contrat 1er degré: 44 ans 2nd degré : 46 ans	Secteur public 1er degré: 85.2% 2nd degré : 58.7% Secteur privé sous contrat 1er degré: 92.0% 2nd degré : 65.8%	Secteur public 1er degré: 14.8% 2nd degré : 41.3% Secteur privé sous contrat 1er degré: 8% 2nd degré : 34.2%

Tableau 3: d'après *Repères et Références statistiques : enseignement, formation, recherche*. DEPP 2021

4 Construction de la base de données en vue du traitement dans le cadre de l'Analyse Statistique Implicative – ASI

L'analyse statistique implicative (ASI) est, initialement, une méthode d'analyse de données non symétrique, conçue à l'origine à la fin des années 70 par Régis Gras (1979, 2018). Nous ne reviendrons pas ici sur les développements de ce cadre théorique de l'ASI. (Gras, Régnier, Guillet, 2009), (Gras, Régnier, Lahanier-Reuter, Marinica, Guillet, 2017) (Gras, Régnier, 2017). Les traitements sont assistés par le recours au logiciel CHIC dans sa version 7. Le nom CHIC est l'acronyme de Classification Hiérarchique Implicative Cohésitive dont l'historique est exposé par Raphaël Couturier et Saddo Ag Almouloud (Gras et *al.*, 2017 p.219-223).

Code de la question	Nature de la variable	Modélisation statistique dans le cadre de l'ASI par des vecteurs-variables à n composantes binaires	Variables binaires sur {0 ; 1}
V001	Qualitative nominale	(V001Femme, V001Homme)	2
V002	Quantitative transformée 4 en segments par l'algorithme intégré au logiciel CHIC	(Age1, Age2, Age3, Age4)	4
V003	idem	(Ancien1, Ancien2, Ancien3, Ancien4)	4
V004/5/6 V007/8/9 V010/…/16 V018/…/21 V022/…/24 V028/…/33 V34 V035/…/42	Qualitative nominale	(V004, V005, V006) (V007, V008, V009) (V010, V011, …, V016) (V018, V019, V020, V021) (V022, V023, V024) (V028, …, V033) (V034_1, …, V034_4, V034Autre, V034NR) (V035, …, V042)	3 3 7 4 3 6 6 8
V025	Qualitative nominale	(V025NR, V025Autre, V025Ouiavec, V025Ouisans, V025Ouimais)	5
V026 V027	Qualitative ordinale avec prise en compte de la non-réponse	(V026NR, V026jamais, V026rarement, V026parfois, V026souvent, V026Trèssouvent) (V027NR, V027jamais, V027rarement, V027parfois, V027souvent, V027Trèssouvent)	6 6

Tableau 4 : questions associées aux variables statistiques transformées en variables binaires constituant la base de données initiale soumise aux traitements du logiciel CHIC 7.

Pour construire la base de données, nous sommes partis des réponses fournies à 14 des 25 questions du questionnaire (annexe 1). La base initiale de données est d'abord constituée de 59 variables binaires à valeurs dans {0 ; 1} et de 2 variables numériques : âge et ancienneté professionnelle. Cette base est enregistrée sous la forme d'un tableau au format *.csv.

Nous avons procédé au traitement avec le logiciel CHIC 7 avec le paramétrage suivant : modèle de l'implication selon la théorie classique avec la loi binomiale.

À l'issue du premier traitement, l'ensemble des valeurs de l'âge et de l'ancienneté professionnelle a été soumis à l'algorithme de partitionnement en 4 segments. Le résultat suivant a été obtenu :

Variable V002_âge	Variable V003_Ancienneté professionnelle
Partitions optimales : V002_Age1 de 22.193 à 29.4428 V002_Age2 de 31.7591 à 39.9945 V002_Age3 de 41.7577 à 50.4531 V002_Age4 de 52.909 à 63.2745	Partitions optimales : V0031 de 0 à 5 V0032 de 6 à 13 V0033 de 15 à 26 V0034 de 30 à 42

Tableau 5 : Partitions des ensembles des valeurs de l'âge et de l'ancienneté professionnelle

La nouvelle base de données est alors constituée de 67 variables binaires.

Nous avons procédé à un tri des variables selon leur rôle :

Variables principales $N_p = 51$	Variables supplémentaires $N_s = 16$
(V007, V008, V009) (V010, V011, V012, V013, V014, V015, V016) (V018, V019, V020, V021) (V025NR, V025Autre, V025Ouiavec, V025Ouisans, V025Ouimais) (V026NR, V026jamais, V026rarement, V026parfois, V026souvent, V026Trèssouvent) (V027NR, V027jamais, V027rarement, V027parfois, V027souvent, V027Trèssouvent) (V028, V029, V030, V031, V032, V033) (V034_1, V034_2, V034_3, V034_4, V034Autre, V034NR) (V035, V036, V037, V038, V039, V040, V041, V042)	(V001Femme, V001Homme) (Age1, Age2, Age3, Age4) (Ancien1, Ancien2, Ancien3, Ancien4) (V004, V005, V006) (V022, V023, V024)

Tableau 6: variables principales et variables supplémentaires

Nous avons alors soumis cette base de données à la procédure de recherche des relations de quasi-équivalence afin d'en réduire le nombre de colonnes (Gras et al., 2017, p119-124). Nous avons un seuil d'indice de quasi-équivalence supérieur à 0.90. Nous rapportons en annexe 3 le tableau des quasi-équivalence. Nous faisons alors le choix des variables représentant les classes de quasi-équivalences Il ressort les résultats suivants :

Classes de quasi-équivalence	Représentant choisi	Variables retirées de la base de données finale
V027parfois, V026parfois	V026parfois	V008 V020 V025NR V025Ouiavec V025Ouimais V027NR V027parfois V027souvent V027Tsouvent V033 V035 V040 V34_NR
V025Ouiavec, V039, V033	V039	
V025NR, V34_NR, V027NR, V026NR	V026NR	
V042,V028	V042	
V014, V008	V014	
V027souvent, V026souvent	V026souvent	
V041, V025Ouimais, V020	V041	
V035, V032	V032	
V026Tsouvent, V040, V027Tsouvent	V026Tsouvent	

Tableau 7 : Réduction de la base de données par la quasi-équivalence de variables principales

Il en résulte une réduction de 13 colonnes dans la base de données à partir laquelle nous allons conduire nos analyses, c'est-à-dire à partir de 38 variables binaires considérées en tant que variables principales et 16 variables binaires considérées comme illustratives.

5 Traitements et analyses des données dans le cadre de l'ASI

Nous avons alors conduit les traitements des données à partir de la base réduite enregistrée au format *.csv. Nous avons exploré l'ensemble des relations non symétriques entre variables binaires en nous limitant aux indices d'implication dont la valeur est supérieure ou égale à 0.90. Nous avons réalisé une distinction des niveaux 0.90 (vert), 0.95 (bleu) et 0.99 (rouge).

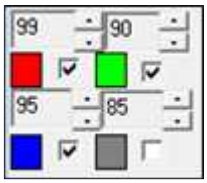


Figure 17 : distinction des relations non-symétriques entre variables binaires selon le niveau de l'intensité d'implication

Nous obtenons alors le graphe implicatif suivant dans lequel les flèches en pointillé indiquent les fermetures transitives. Nous retenons ainsi 23 relations de quasi-implication.

Niveau de l'intensité d'implication $\varphi(a \rightarrow b)$	0.99	0.95	0.90	Fermeture transitive
Nombre	2	7	9	5

Tableau 8 : fréquence des relations de quasi-implication selon les niveaux des indices d'implication

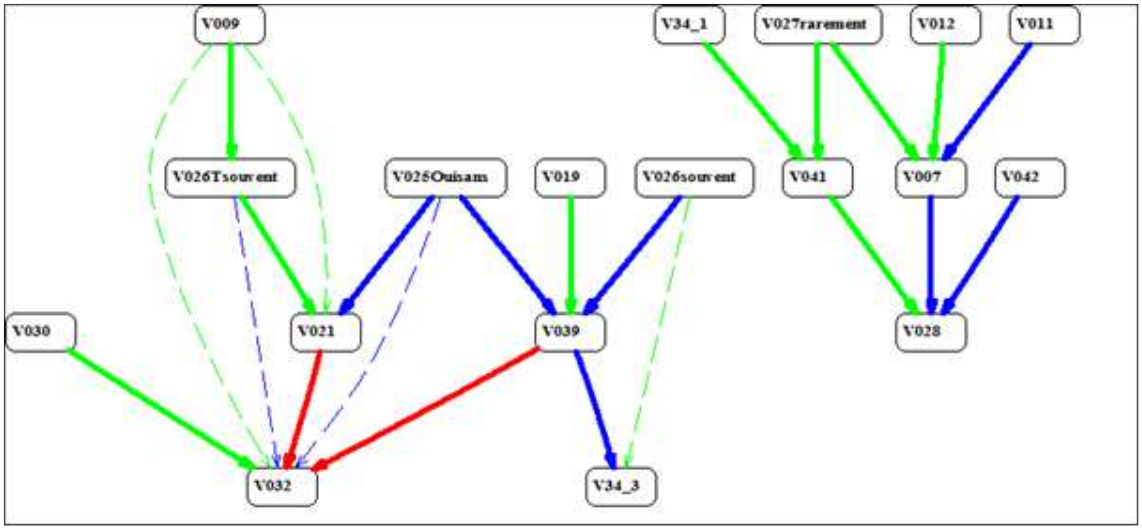


Figure 2: graphe implicatif aux niveaux d'intensité d'implication 0.99, 0.95 et 0.90

Nous observons que ce graphe implicatif est constitué de deux sous-réseaux que nous allons analyser et interpréter successivement. Nous allons nous intéresser particulièrement aux caractéristiques de l'identité professionnelle et pédagogique des enseignants et des liens qu'elles peuvent entretenir. Dans cette perspective nous nous proposons d'analyser les relations entre les variables portant sur la fréquence des interactions entre pairs non-maîtrisées par l'enseignant dans sa classe et celles portant sur les réactions suscitées chez l'enseignant par ce type d'interactions.

5.1 Analyse et interprétation des résultats fournis par le sous-réseau n°1

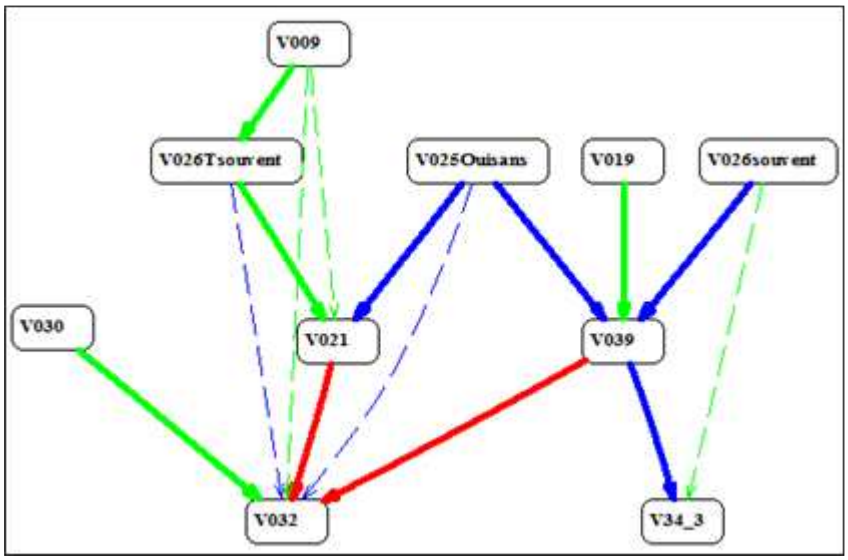


Figure 3 : graphe implicatif du sous-réseau n°1

5.1.1 Analyse du chemin V025Ouisans⇒V021⇒V032

Au sein du sous-réseau numéro 1, nous remarquons le chemin n°1

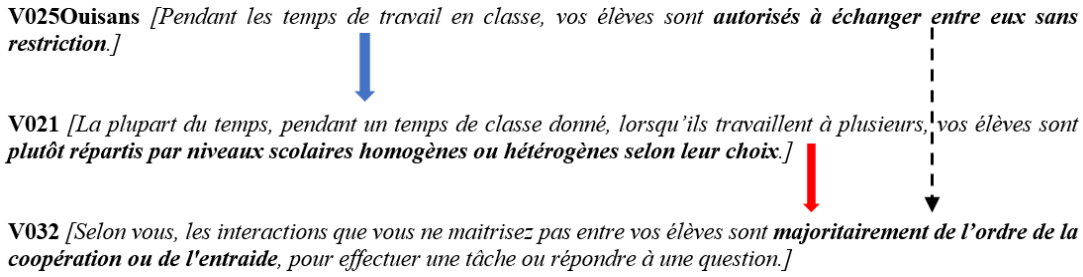


Figure 4 : graphe du chemin n°1

Ce chemin établit une relation quasi-implicative entre la liberté d'échange sans restriction accordée aux élèves par l'enseignant et le choix donné aux élèves de constituer eux-mêmes leurs groupes de travail. Les enseignants disant faire ces choix pédagogiques considèrent les interactions entre élèves non maîtrisées pendant le temps de classe comme un outil coopératif. On peut supposer que cet outil est mobilisé spontanément par les élèves, du fait de leur liberté de choix concernant les échanges entre eux et la liberté de constitution des groupes de travail.

Nous recourons aux variables illustratives (supplémentaires) pour mieux comprendre la construction de ce chemin. Nous rapportons en annexe 4 les niveaux de contribution des catégories déterminées par les variables supplémentaires à construction du chemin n°1. La notion de contribution est définie dans (Gras et al., 2017 p.111-112)

Le groupe d'âge qui contribue le plus à ce chemin est le groupe 4, soit la tranche d'âge de 52,9 à 63,3 ans. Le groupe d'ancienneté qui contribue le plus est le groupe 3, avec une ancienneté de 15 à 26 ans. Ce sont plutôt des hommes que des femmes. Cette caractéristique n'apparaît plus comme notable dans l'ensemble des contributions que nous avons calculées dans cet article. On peut remarquer que ces personnes sont donc les plus âgées mais pas celles qui ont le plus d'ancienneté dans le métier, et donc supposer qu'il s'agit majoritairement de personnes ayant connu une réorientation professionnelle au cours de leur vie. Nous relevons également que les variables V005 et V024 contribuent à ce chemin avec un risque très faible, respectivement de 0.000311 et 0.00833. La variable V005 est choisie par les enseignants qui disent utiliser en classe une pédagogie non traditionnelle identifiée. La variable V024 concerne quant à elle les enseignants dont la ou les classes comportent trois niveaux ou plus. Les enquêtés concernés sont donc majoritairement des enseignants qui travaillent avec une pédagogie non traditionnelle spécifique et dans des classes multiniveaux.

L'étude de ce chemin nous amène à porter notre attention sur la variable V021.

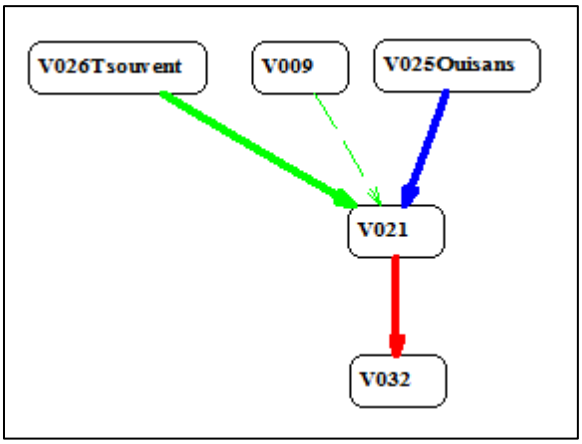


Figure 5: Mode cône centré sur la variable V021

V009 *[Dans votre classe, les élèves* ***choisissent leur place et en changent librement****]*

V026Tsouvent *[Vos élèves* ***interagissent très souvent*** *entre eux sans que vous maitrisiez le contenu de ces échanges]*

V025Ouisans *[Pendant les temps de travail en classe, vos élèves sont* ***autorisés à échanger entre eux sans restriction****.]*

V021 *[La plupart du temps, pendant un temps de classe donné, lorsqu'ils travaillent à plusieurs, vos élèves sont* ***plutôt répartis par niveaux scolaires homogènes ou hétérogènes selon leur choix****]*

V032 *[Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont* ***majoritairement de l'ordre de la coopération ou de l'entraide****, pour effectuer une tâche ou répondre à une question]*

Figure 6 : graphe du sous-réseau centré sur V021 par le mode cône

Nous pouvons relever une chaîne de relations quasi-implicatives principalement axée sur la liberté des élèves dans la classe, selon les enseignants interrogés. Les élèves ont en effet la possibilité de choisir leur place et d'en changer librement, les enseignants déclarent que des interactions entre pairs qu'ils ne maitrisent pas ont lieu très souvent, et les élèves choisissent également la composition de leurs groupes de travail le cas échéant. La conséquence finale de ce chemin est la considération accordée aux interactions entre pairs qu'ils ne maitrisent pas par les enseignants sondés. Pour ceux-ci, ces interactions jouent majoritairement un rôle coopératif lié aux apprentissages en cours, de la même manière que pour le chemin étudié précédemment.

Si nous examinons les contributions (annexe 5) des groupes d'individus déterminés par les variables supplémentaires à la construction du chemin V26Tsouvent⇒V021⇒V032, il ressort que les catégories qui concourent le plus et de manière significative à un niveau de risque α nettement inférieur à 0.05, valeur seuil communément admise, sont Age3 (α=0.0124), Ancien3 (α=0.0237), V005 (α=0.000805) et V024 (α=0.00181).

Plus explicitement, il s'agit plutôt des enseignants de la tranche d'âge plus de 41 ans et moins de 51 ans, ayant une ancienneté professionnelle entre 15 et 26 années, pratiquant une pédagogie non traditionnelle identifiée parmi celles développées par Freinet, Montessori, Steiner, …, dans des classes multiniveaux.

En comparant ces résultats à ceux relevés pour le chemin V025Ouisans⇒V021⇒V032, nous notons une différence au niveau de la tranche d'âge

qui contribue majoritairement à ce chemin, puisque c'est ici la tranche Age3 qui contribue significativement et non Age4. Les enseignants concernés ont une ancienneté en corrélation avec leur âge, nous pouvons donc en déduire qu'il s'agit de professionnels entrés dans le métier en fin d'études supérieures et qui ne se sont pas réorientés depuis.

Une variable comparable à la V021 est la variable V019 : « la plupart du temps, pendant un temps de classe donné, lorsqu'ils travaillent à plusieurs, vos élèves sont plutôt répartis par niveaux scolaires hétérogènes ». Si la variable V021 mettait en avant l'intention des enseignants qui la choisissent pour déterminer les groupes de travail selon la volonté des élèves, la variable V019 suggère plutôt le choix des enseignants de constituer des groupes hétérogènes pour les travaux de groupes.

Nous pouvons ainsi nous intéresser à deux chemins qui ont pour origine la variable V019.

5.1.2 Analyse des chemins V019⇒V039⇒V032 et V019⇒V039⇒V034_3

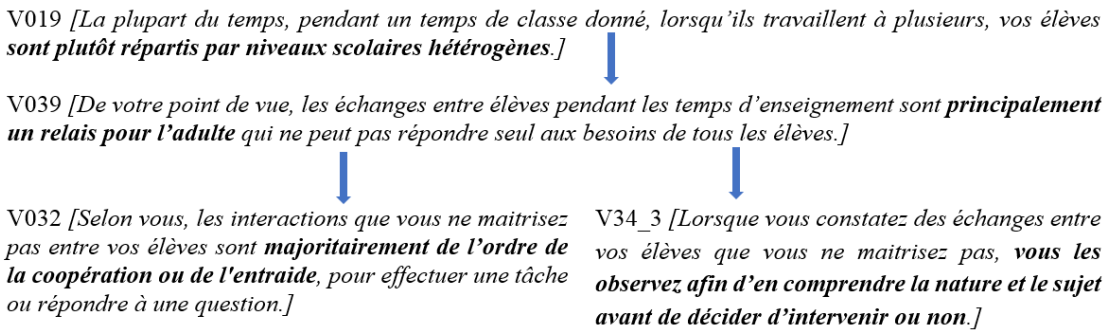


Figure 7: sous-réseau d'origine V019

Bien que l'option retenue par les enseignants qui choisissent V019 pour constituer les groupes de travail dans leurs classes soit de les constituer eux-mêmes plutôt que d'en laisser le choix aux élèves, l'idée forte reste l'utilité pour l'enseignant de recourir aux interactions entre pairs qu'il ne maitrise pas pendant ses temps d'enseignement comme un outil pédagogique. On retrouve également la notion de coopération apportée par la variable V032. Il est intéressant de noter que le chemin V019⇒V039⇒V034_3 nous apporte une information sur l'attitude des enseignants lorsque ces interactions se produisent et qu'ils les remarquent : ils se posent plutôt en observateurs pour évaluer la pertinence de l'interaction dans le contexte et intervenir ou pas. On peut rapprocher ce choix de celui de la variable V019 car ces deux items comportent l'idée de rendre possible un certain brassage entre les élèves, tout en conservant un certain contrôle de l'enseignant sur ce qui se passe.

5.1.3 Analyse du chemin V025Ouisans⇒V039⇒V032

V025Ouisans *[Pendant les temps de travail en classe, vos élèves sont **autorisés à échanger entre eux sans restriction**]*

V039 *[De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont **principalement un relais pour l'adulte** qui ne peut pas répondre seul aux besoins de tous les élèves]*

V032 *[Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont **majoritairement de l'ordre de la coopération ou de l'entraide**, pour effectuer une tâche ou répondre à une question]*

Figure 8 : chemin d'origine V025Ouisans

Il apparait une chaîne de relations quasi-implicatives entre la liberté d'échange entre pairs laissée aux élèves en classe, la considération des interactions entre pairs non-maitrisées par l'enseignant comme un adjuvant pour celui-ci pendant les situations d'apprentissage, et l'idée que ces interactions sont principalement d'ordre coopératif. Nous pouvons donc en déduire que parmi les enseignants sondés, ceux qui laissent libre cours aux interactions entre pairs en classe sont également ceux qui considèrent ces interactions comme un atout et un outil du point de vue pédagogique et relationnel.

Comme pour le chemin V025Ouisans⇒V021⇒V032, c'est le groupe 4 d'âge, les plus âgés, et le groupe 3 d'ancienneté professionnelle qui contribuent le plus au chemin V025Ouisans⇒V039⇒V032. De même, les groupes déterminés par les variables V005 et V024 sont ceux qui contribuent le plus à ce chemin avec un risque très faible, respectivement de 0.000805 et 0.0144. Nous pouvons donc tirer des conclusions similaires pour ces deux chaînes de relations quasi-implicatives, à savoir que les enseignants qui travaillent avec une pédagogie dite alternative dans des classes multiniveaux sont ceux qui ont le plus tendance à laisser de la place aux interactions entre élèves sans qu'elles soient maitrisées pendant les temps d'enseignement, et à les considérer comme un outil pédagogique valable.

Plusieurs des relations étudiées précédemment nous amènent à nous intéresser plus particulièrement à la variable V039 : « de votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves. »

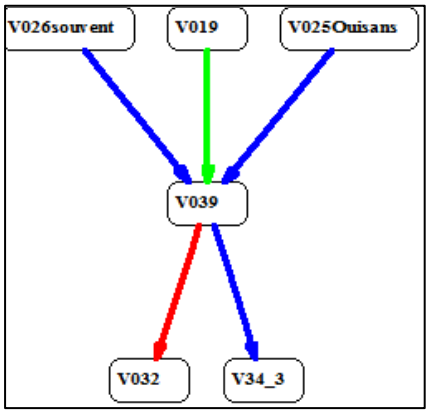


Figure 9 : Mode cône centré sur V039

Pour les enseignants qui sont d'accord avec ce point, on relève préalablement que des interactions entre pairs non maitrisées par l'adulte se produisent souvent dans la classe (V026souvent), que les groupes de travail sont plutôt hétérogènes (V019) et que les élèves peuvent échanger sans restriction entre eux (V025Ouisans). Le choix de la variable V039, soit la considération des interactions entre pairs que l'enseignant ne maitrise pas comme un relais pour celui-ci, apparait comme étant majoritairement à l'origine d'une attitude particulière des enseignants face à ces interactions : elles sont considérées comme étant plutôt de l'ordre de la coopération ou de l'entraide (V032), et l'enseignant adopte une attitude observatrice lorsqu'elles se produisent pour juger de leur pertinence avant de décider d'intervenir ou non sur ces échanges (V34_3).

5.1.4 Analyse de l'arc V030⇒V032 et du cône centré sur V032

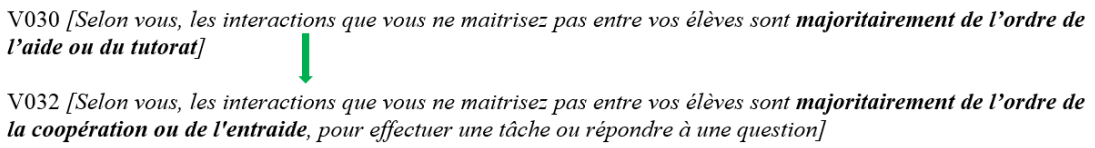


Figure10 : arc V030⇒V032

En s'intéressant ensuite à l'arc V030⇒V032, nous portons notre attention sur la relation quasi-implicative établie entre ces deux variables V030 et V032, qui sont des items relevant d'une même question comportant plusieurs réponses possibles dans l'enquête menée auprès des professionnels. Un lien peut donc être établi entre la tendance à considérer les interactions entre pairs non-maitrisées par l'enseignant comme relevant de l'aide ou du tutorat (relation d'aide dans un seul sens, d'un individu vers l'autre) et la tendance à les considérer comme relevant de la coopération ou de l'entraide (relation d'aide à double-sens entre deux individus). Les enquêtés qui choisissent la variable V030 ont tendance à choisir également la variable V032 ; nous pouvons donc supposer que ces enseignants considèrent que les interactions entre élèves non maîtrisées par l'enseignant sont plutôt un outil pédagogique pertinent quelle que soit la nature de la relation d'aide qui s'établit entre deux individus pendant le temps d'apprentissage.

On observe que plusieurs variables ont une relation quasi-implicative directe ou transitive avec la variable V032. Les variables concernées ont en commun d'être centrées autour de la liberté de l'élève dans ses choix pendant le temps de classe : les élèves choisissent leur place et en changent librement (V009), choisissent leur répartition lors de travaux de groupes (V021), sont autorisés à échanger entre eux sans restriction (V025Ouisans) et interagissent très souvent sans que l'enseignant maitrise le contenu de ces échanges (V026Tsouvent). Elles se regroupent également autour de la considération que les interactions entre pairs non maitrisées par l'enseignant sont un outil pédagogique pertinent : elles sont un relai pour l'adulte en charge de la classe (V039), relèvent de l'aide ou du tutorat (V030). Nous pouvons donc en déduire un lien notable entre les choix pédagogiques de l'enseignant et la manière dont il considère les interactions entre pairs qu'il ne maitrise pas dans la classe.

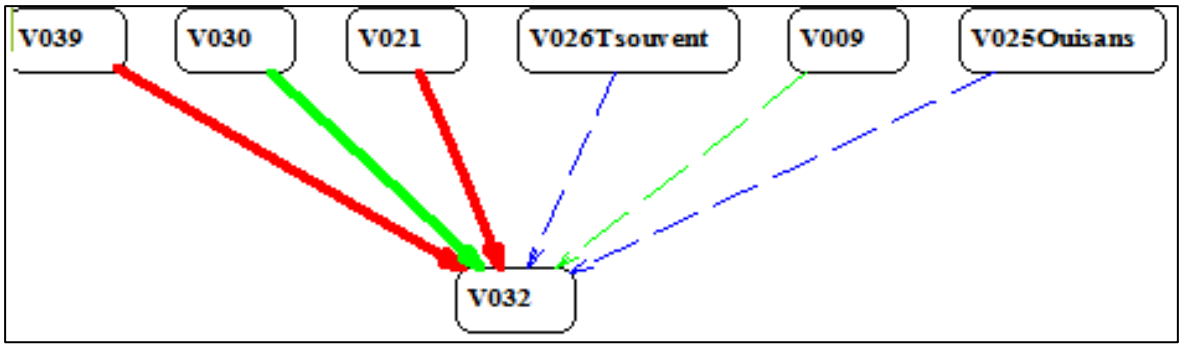


Figure 11 : Mode cône centré sur la variable V032

Enfin, il nous semble pertinent de nous recentrer aussi sur les variables V026Tsouvent et V026souvent, qui sont proches en termes de sens puisque les enseignants qui les choisissent déclarent qu'il se produit souvent ou très souvent des interactions entre pairs qu'ils ne maitrisent pas dans leurs classes.

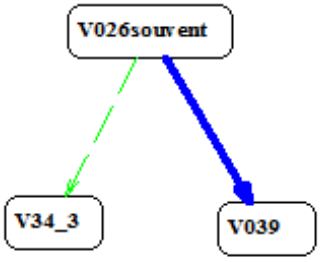


Figure 12 : Mode cône centré sur la variable V026souvent

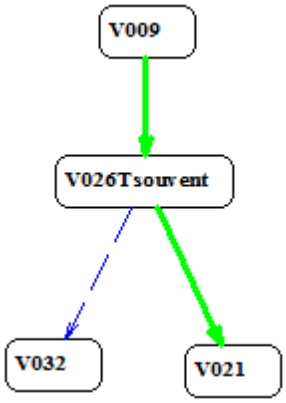


Figure 13 : Mode cône centré sur la variable V026Tsouvent

Il est intéressant de noter que, bien qu'elles soient proches en termes de sens, ces deux variables apparaissent comme n'ayant pas les mêmes causes ni les mêmes conséquences dans notre traitement statistique. Leur seul point de rencontre est qu'elles appartiennent toutes deux à des chemins qui ont pour conséquence finale la variable V032 mais elles ne semblent pas avoir de lien direct probant. Cette distinction nous permet d'introduire une nuance dans notre analyse. Le chemin V009⇒V026Tsouvent⇒V021 met en avant le fait que les enseignants qui déclarent que des interactions entre pairs qu'ils ne maitrisen t pas se produisent très souvent dans leurs classes sont principalement ceux qui disent laisser la liberté de choix de leur place à leurs élèves ; la conséquence qui apparait est que les groupes de travail dans les classes de ces enseignants sont constitués au choix des élèves. La relation quasi-implicative entre V026souvent et V039 montre quant à elle que les enseignants qui déclarent que des interactions entre pairs qu'ils ne maitrisent pas se produisent souvent dans leurs classes considèrent majoritairement que ces interactions sont un relais pour l'enseignant. Cette nuance nous permet de confirmer qu'il existe une distinction entre deux sous-ensembles dans notre Groupe numéro 1, entre les enseignants qui disent accorder une liberté de choix presque totale à leurs élèves, et ceux qui, tout en laissant une liberté relativement grande aux élèves, semblent garder un peu plus de

contrôle sur les placements et les échanges. Le chemin V026souvent > V039 > V34_3 met d'ailleurs en évidence, par l'intervention de la variable V34_3 (« lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas, vous les observez afin d'en comprendre la nature et le sujet avant de décider d'intervenir ou non ») comme conséquence finale, cette volonté d'exercer une certaine maitrise sur ces interactions qui, au départ, échappent au contrôle de l'enseignant, sans pour autant choisir l'intervention systématiquement.

Au terme de l'analyse de ce premier groupe de relations quasi-implicatives, on voit apparaitre un profil assez net chez une partie des enquêtés. Il s'agit d'enseignants qui utilisent généralement une pédagogie non traditionnelle identifiée et qui travaillent plutôt en classes multi-niveaux. Ils optent pour le placement libre des élèves, travaillent avec des groupes hétérogènes ou constitués librement. Ils considèrent les interactions entre pairs comme un outils dans le processus d'apprentissage, ils déclarent que celles-ci se produisent souvent ou très souvent en classe. Lorsque ces interactions se produisent, l'adulte n'intervient que s'il le juge nécessaire après observation de la situation.

5.2 Analyse et interprétation des résultats fournis par le sous-réseau n°2

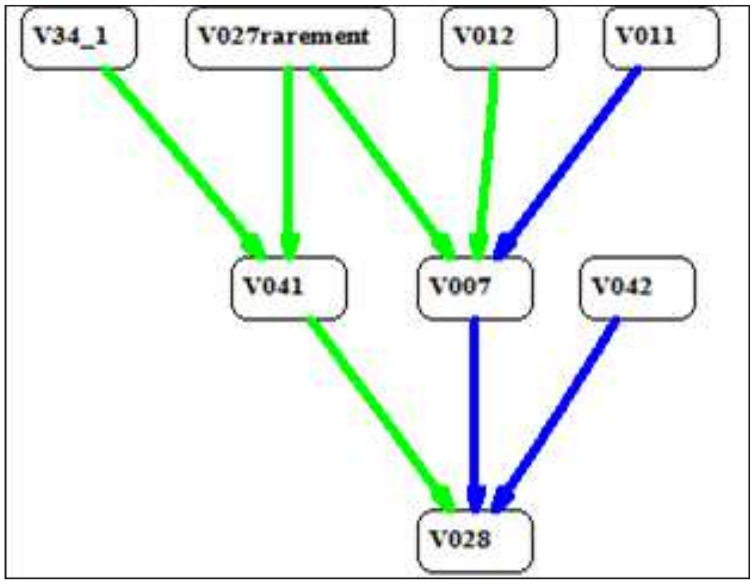


Figure 14 graphe implicatif du sous-réseau n°2

5.2.1 Analyse du chemin V011⇒V007⇒V028

Parmi les chaînes de relations qui apparaissent dans le sous-réseau n°2, nous relevons tout d'abord celles dont l'intensité d'implication est le plus élevé, soit supérieur ou égal à 0.95, c'est-à-dire le chemin en lequel nous attribuons le plus confiance en sa validité. Il s'agit du chemin V011⇒V007⇒V028 qui met en avant une disposition de la classe plutôt traditionnelle, en rangs de deux bureaux, où les élèves ont une place attitrée où ils demeurent la majorité du temps. Les enseignants qui déclarent faire ces choix pédagogiques considèrent plutôt que les interactions entre élèves non maîtrisées par l'enseignant au sein de leur classe relèvent du bavardage sans lien avec les apprentissages.

V011 *[Votre salle de classe est disposée en **rangs de deux bureaux**.]*

V007 *[Dans votre classe, **les élèves ont une place attitrée** à laquelle ils travaillent tout le temps ou la plupart du temps.]*

V028 *[Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement **de l'ordre du bavardage** ou de l'échange récréatif, elles ne concernent pas le travail.]*

Figure 15 : chemin d'origine V011

A la suite de l'étude de ce chemin, il parait intéressant de se centrer plus précisément sur la variable V007

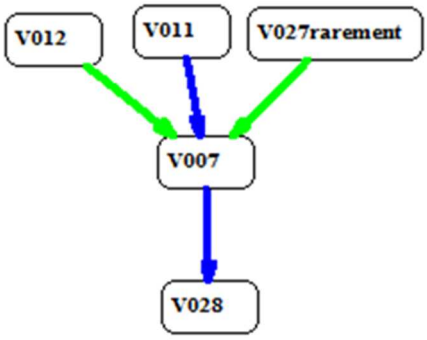


Figure 16 : Mode cône centré sur V007

Nous relevons trois antécédents de la variable V007 : V012 (« votre/vos salle.s de classe est/sont disposée.s en rangées de trois bureaux ou plus. »), V011 (« votre/vos salle.s de classe est/sont disposée.s en rangs de deux bureaux. ») et V027rarement (« des échanges que vous ne maitrisez pas en classe se produisent rarement »). La mise en relation de ces quatre variables forme un ensemble cohérent : les enseignants qui attribuent une place attirée à leurs élèves sont plutôt ceux dont les classes sont disposées en rangs ou en rangées, et qui déclarent que des interactions entre pairs qu'ils ne maitrisent pas se produisent rarement dans leurs classes.

La conséquence unique de la V007 qui apparait lors du traitement de nos données est la variable V028 (« selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre du bavardage ou de l'échange récréatif, elles ne concernent pas le travail. »). Un lien manifeste est mis en avant entre le choix pédagogique d'attribuer une place attitrée aux élèves en classe et l'idée que les interactions entre pairs que l'enseignant ne maitrisent pas, ne concernent pas les apprentissages.

Pour l'arc V007⇒V028, (voir annexe 6) les tranches qui contribuent le plus sont la tranche d'âge 3, soit les sondés âgés d'environ 41 ans à 50 ans, et la tranche d'ancienneté 3, soit les enseignants ayant 15 à 26 ans d'ancienneté dans le métier. Deux variables contribuent à la construction de cette relation de quasi-implication avec un risque très faible : la variable V004 avec un risque de 0.00461 et la variable V022 avec un risque de 0.000663. Il s'agit donc d'enseignants dont l'organisation pédagogique relève plutôt de caractéristiques d'une pédagogie traditionnelle (V004), et qui ont majoritairement des élèves d'un seul niveau dans leur classe (V022).

Si la variable V028 est la seule conséquence de la variable V007, elle comporte en revanche plusieurs causes ainsi que nous pouvons le remarquer en s'intéressant plus particulièrement à cette variable.

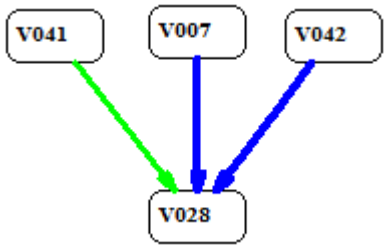


Figure 17 : Mode cône centré sur V028

Comme l'arc V007⇒V028, l'arc V042⇒V028 est retenu avec un niveau de confiance élevé avec un intensité d'implication supérieur ou égal à 0.95. Celui-ci met en avant le fait que les enseignants qui considèrent que les interactions entre élèves non maîtrisées par l'enseignant en classe sont majoritairement de l'ordre du bavardage (V028), sont également ceux qui pensent que ces échanges peuvent être bons pour certains élèves mais pas pour d'autres (V042). On peut déduire de cette relation quasi-implicative une conception nuancée de ces interactions par les enseignants qui choisissent ces variables. Ceux-ci semblent penser que les interactions entre élèves non maîtrisées par l'enseignant sont plutôt récréatives et sans lien avec le travail et peuvent être dommageables pour les apprentissages pour certains élèves, sans toutefois les condamner tout à fait.

En revanche, les enseignants qui choisissent plutôt la variable V041 déclarent penser que les interactions entre élèves non maîtrisées par l'enseignant sont inévitables mais qu'elles doivent être contenues pour poursuivre les apprentissages, et adoptent donc un point de vue plus fataliste vis-à-vis de ces interactions. L'arc V041⇒V028 a pour contributeurs principaux des enseignants issus des tranches d'âge et d'ancienneté 4 (risques de 0,125 et 0.0958), ce sont donc les enseignants plutôt en fin de carrière et proches de la retraite. Les variables qui contribuent le plus à établir ce lien sont les mêmes que pour l'arc V007⇒V028. Il s'agit donc là aussi d'enseignants dont l'organisation pédagogique relève plutôt de caractéristiques d'une pédagogie traditionnelle (V004, risque de 0.00461), et qui ont majoritairement des élèves d'un seul niveau dans leur classe (V022, risque de 0.00347).

5.2.2 Analyse du chemin V34_1⇒V041⇒V028

Intéressons-nous enfin à la chaîne de relations quasi-implicatives V34_1⇒V041⇒V028, dont la construction est obtenue à partir d'intensité d'implication supérieur ou égal à 0.90.

V34_1 *[Lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas,* ***vous intervenez afin de recadrer l'échange (sans forcément l'interdire)****.]*

V041 *[De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont* ***principalement inévitables, mais il faut réussir à les contenir*** *pour avancer.]*

V028 *[Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement* ***de l'ordre du bavardage ou de l'échange récréatif****, elles ne concernent pas le travail.]*

Figure 18 : le chemin V34_1>V041>V028

Ce chemin nous apporte une précision intéressante pour compléter l'analyse de la relation quasi-implicative V041⇒V028. En effet, il met en avant comme cause possible la variable V34_1 qui nous informe sur l'attitude des enseignants lorsque des interactions entre pairs qu'ils ne maitrisent pas se produisent : ils interviennent « afin de recadrer l'échange » ; autrement dit, ils admettent des interactions entre pairs, mais n'admettent pas de ne pas les maitriser.

Après avoir analysé ce second sous-réseau de relations quasi-implicatives, nous voyons apparaitre un deuxième profil d'enseignants parmi les personnes sondées. Ces enseignants travaillent plutôt avec une pédagogie traditionnelle. Le placement des élèves est imposé par l'enseignant et fixe dans le temps et l'espace. L'organisation des tables de travail est plutôt en rangs de deux ou en rangées. Les interactions entre pairs non maitrisées par l'enseignant sont déclarées plutôt rares. Ces interactions sont perçues comme récréatives et sans lien avec le travail, voire de nature à le perturber et pouvant être néfastes. Il est nécessaire de les cadrer fermement bien qu'il ne soit pas possible de les éviter complètement. Ces enseignants ont plutôt tendance intervenir systématiquement lorsque ces échanges sont constatés.

Nous pouvons noter le fait que la variable V34_1, relative à l'attitude adoptée par les enseignants face aux interactions entre pairs qu'ils ne maitrisent pas, apparait dans le second groupe implicatif comme une cause. Elle est située au début de la chaîne de relations quasi-implicatives à laquelle elle appartient. Dans le premier groupe, la variable V34_3, relative à l'attitude adoptée par les enseignants, apparait toujours comme la conséquence finale des chaînes de relations quasi-implicatives auxquelles elle appartient. On peut donc supposer que la manière de réagir des enseignants du premier groupe est un choix qui découle d'autres choix pédagogiques, alors que pour les enseignants du second groupe, il s'agit plutôt d'un positionnement de départ qui n'est pas orienté par les options pédagogiques choisies.

6 Exploration du graphe implicatif intégrant la conjonction de variables binaires

Pour compléter l'analyse réalisée précédemment, nous avons prolongé celle-ci en introduisant les variables issues de la conjonction *et* de deux variables. Afin de limiter le développement de cette analyse, nous ne retenons que les valeurs des intensités d'implication supérieure ou égale à 0.98.

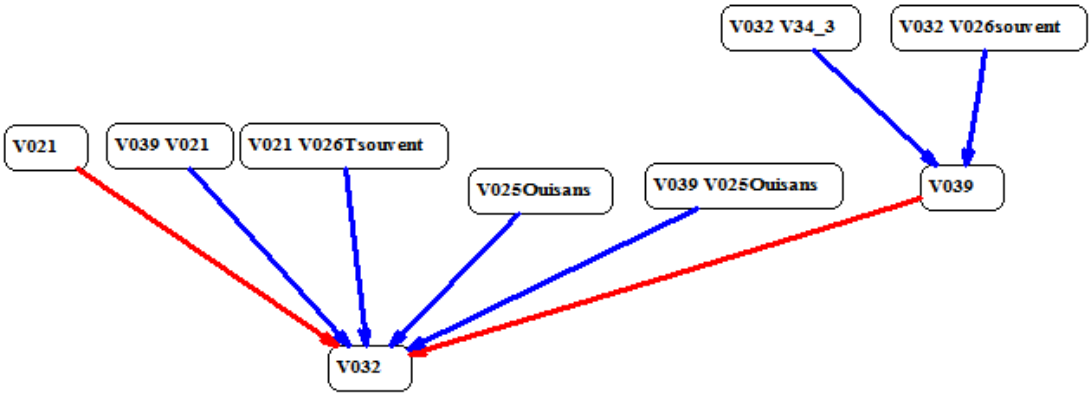


Figure 19 : graphe implicatif prenant en compte les variables issues de la conjonction *et* entre deux variables au seuil d'intensité d'implication supérieur ou égal à 0.98

Nous constatons que la variable V032 selon laquelle les interactions entre élèves relèvent majoritairement de l'ordre de la coopération ou de l'entraide, pour effectuer une tâche ou répondre à une question apparaît comme conséquence de l'ensemble des autres variables.

La variable V032 en conjonction, d'une part, avec la variable V34_3 qui se réfère à une réaction de l'enseignant centrée sur l'observation des échanges entre élèves afin d'en comprendre la nature et le sujet avant de décider d'intervenir ou non, d'autre part, avec le variable V026souvent qui exprime la reconnaissance par l'enseignant des occurrences (souvent) des échanges dans la classe, ont pour conséquent conduit à considérer que les échanges sont un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves.

Nous pouvons donc supposer que le constat fait par les enseignants de la présence fréquente des interactions entre pairs qu'ils ne maitrisent pas, et leur attitude d'observation face à ces interactions, les a amenés à remarquer l'intérêt de laisser ces interactions se dérouler, puisqu'elles aident l'enseignant dans son travail. L'utilité de ces interactions amène les enseignants concernés à attester de la nature coopérative et relative au travail donné par l'enseignant de ces interactions.

7 Exploration de l'arbre cohésitif représentation d'une classification hiérarchique orientée

Pour compléter notre exploration des données telle que réalisée dans les sections précédentes où nous avons porté notre attention sur les relations entre variables, nous allons cette fois exploiter les données de notre base en nous appuyant sur une classification hiérarchique orientée de l'ensemble des variables. (Gras et al., 2017, p.89-103). En utilisant la procédure suivante du logiciel CHIC 7, nous obtenons alors :

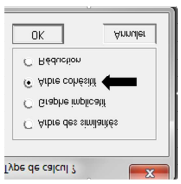


Figure 20

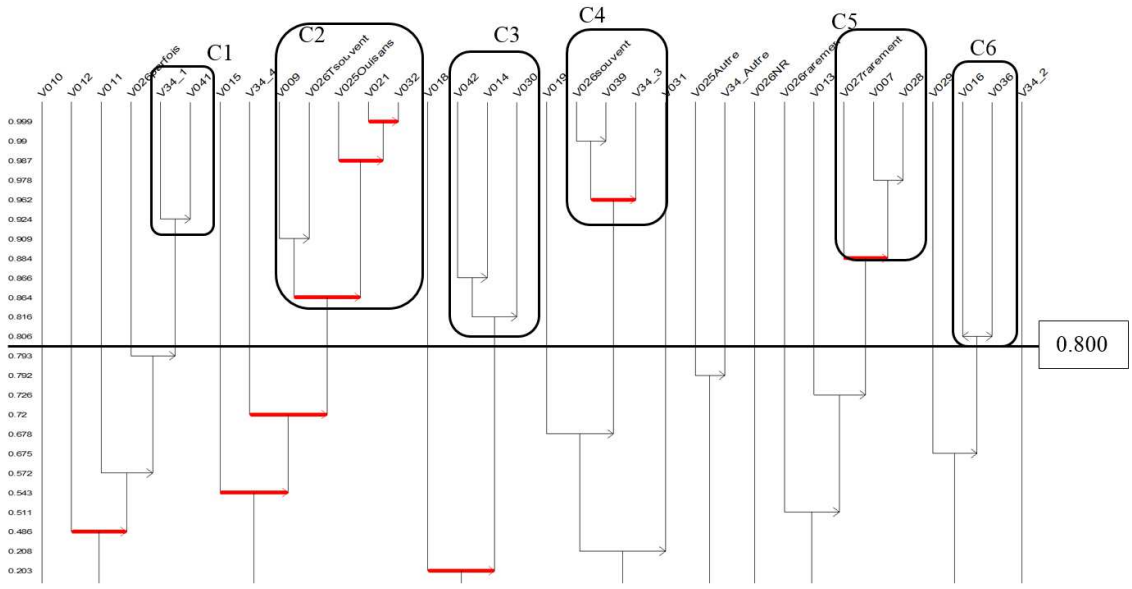


Figure 21 : Représentation graphique de l'arbre cohésitif

Si nous ne retenons que les classes formées à un niveau de la cohésion (Gras et al, 2017, p.91 Définition 7)) supérieur à 0.800, nous obtenons alors 6 classes. Nous les décrivons en lisant l'arbre cohésitif horizontalement de gauche à droite.

7.1 La classe C1={V34_1⇒V041}

Cette classe de degré 1 est formée au nœud n°6 avec une valeur de la cohésion coh(C1)=0.924. Cette classe regroupe 2 variables reliées par une règle simple. L'ordre dans la classe est conforme à la relation de quasi-implication :

Variable binaire	V34_1	V041
Occurrences	9	14

Tableau 9 : ordonnancement des variables selon les occurrences

a) Quelle est votre réaction lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas ? V034_1[Vous intervenez afin de recadrer l'échange (sans forcément l'interdire)]
b) De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement : V041[Inévitables, mais il faut réussir à les contenir pour avancer]

Tableau 10: contenu sémantique des variables binaires de la classe C1

Une règle se forme à ce niveau de la hiérarchie selon laquelle la réaction de l'enseignant est plutôt d'intervenir pour recadrer l'échange entre élèves sans forcément l'interdire dans la mesure où ces échanges sont inévitables et qu'il faut réussir à les contenir pour avancer.

Caractéristiques définies par les variables supplémentaires	Contribution à la construction de la classe avec un niveau de risque	Nombre d'individus appartenant au groupe optimal d'effectif = 6
Age1= De 22 ans à 29.5 ans	0.152	3
Ancien1= Moins de 5 années	0.0621	4
Homme	0.111	2
V006 = pédagogie mixte qui s'inspire de différents courants pédagogiques	0.0522	4
V023= Classe à deux niveaux	0.0684	3

Tableau 11 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C1

Ce sont plutôt les enseignants les plus jeunes et qui ont la plus faible ancienneté professionnelle, qui déclarent pratiquer une pédagogie mixte qui s'inspire de différents courants pédagogiques tant ceux traditionnels que ceux des pédagogies comme celles proposées par Montessori ou Freinet dans des classes à deux niveaux.

L'analyse de la classe C1 vient s'inscrire en miroir du sous-réseau 1 étudié précédemment. En effet, dans le sous-réseau 1 apparaissaient des enseignants d'âge et d'ancienneté plutôt élevés, qui considéraient les interactions entre élèves non maîtrisées par l'enseignant comme un outil et qui laissaient plutôt cours à ces interactions. Nous voyons ici apparaitre un profil d'enseignants d'âge et d'ancienneté plutôt peu élevés, qui auront en revanche une tendance à la méfiance et à l'intervention plus prononcée face à ces interactions.

7.2 La classe C2={(V009⇒V026Tsouvent)⇒(V025Ouisans⇒(V021⇒V032)}

Cette classe de degré 5 est formée au nœud n°10 avec une valeur de la cohésion coh(C2)=0.864. Elle regroupe 5 variables orientées par la relation de quasi-implication. Nous pouvons remarquer ici qu'il y a 2 inversions par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Variables binaires ordonnées dans la classe	V009	V026Tsouvent	V025Ouisans	V021	V032
Occurrences	10	12	8	13	27
Ordonnancement issu du traitement de l'implication statistique	2	3	1	4	5

Tableau 12 : ordonnancement des variables selon les occurrences

Si nous souhaitons tenir compte du critère de cohérence o(C2) de la classe C2 défini (Gras et al, 2017 p.98-103) par la probabilité d'obtenir un nombre i d'inversions strictement supérieur à celui observé réalisation de la variable aléatoire I, nous obtenons alors :

$$o(C2) = Prob\{I > 2\} = \frac{100}{120} \approx 0.833$$

Cette valeur suggère que nous pouvons considérer avec un niveau de cohérence très satisfaisant. Cette valeur est obtenue par tabulation (annexe 7, m=5)

a) Dans votre (vos) classe(s), les élèves... V009[Choisissent leur place et en changent librement]
b) V026_Diriez-vous qu'il arrive que vos élèves interagissent entre eux sans que vous maitrisiez le contenu de ces échanges ? V026Trèssouvent
c) V025_Statut pédagogique des interactions entre élèves dans la classe. Pendant les temps de travail en classe, vos élèves sont-ils autorisés à échanger entre eux ? V025Ouisans

d) La plupart du temps, pendant un temps de classe donné, lorsqu'ils travaillent à plusieurs, vos élèves sont plutôt répartis V021[L'un ou l'autre selon leur choix]
e) Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre... V032[De la coopération ou de l'entraide, pour effectuer une tâche ou répondre à une question]

Tableau 13 :contenu sémantique des variables binaires de la classe C2

Il s'agit d'une règle complexe organisée autour de 5 variables. Partant du fait que les enseignants qui laissent aux élèves la possibilité de choisir leur place et d'en changer librement reconnaissent qu'il arrive très souvent que leurs élèves interagissent entre eux sans qu'ils maitrisent le contenu de ces échanges. Nous pouvons considérer que ce sont eux qui autorisant sans restriction les échanges pendant les temps de travail en classe ont alors tendance à introduire de la souplesse dans la répartition en groupe des élèves dans la mesure où ils considèrent que les interactions qu'ils ne maitrisent pas entre élèves, sont principalement de l'ordre de la coopération ou de l'entraide, pour effectuer une tâche ou répondre à une question.

Caractéristiques définies par les variables supplémentaires	**Contribution à la construction de la classe avec un niveau de risque**	**Nombre d'individus appartenant au groupe optimal d'effectif =14**
Age4= De 52,9 ans à 63,3 ans	0.125	4
Ancien4= de 30 à 42 années	0.0562	3
Homme	0.125	4
V005 = pédagogie non traditionnelle identifiée	5.73e-06	11
V024= Classe à trois niveaux ou plus	0.0371	4

Tableau 14 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C2

Les enseignants qui contribuent le plus à la construction de cette classe sont plutôt des hommes parmi les plus âgés ayant le plus d'ancienneté professionnelles pratiquant une pédagogie non traditionnelle identifiée dans des classes d'au moins trois niveaux. L'analyse de la classe C2 vient donc confirmer les résultats obtenus en analysant différents chemins du sous-réseau 1.

7.3 La classe C3={(V042⇒V014)⇒V030}

Cette classe de degré 3 est formée au nœud n°11 avec une valeur de la cohésion coh(C3)=0.816. Elle regroupe 3 variables orientées par la relation de quasi-implication. Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Variables binaires ordonnées dans la classe	V042	V014	V030
Occurrences	13	14	18
Ordonnancement issu du traitement de l'implication statistique	1	2	3

Tableau 15 : ordonnancement des variables selon les occurrences

La cohérence o(C3) de la classe C3 est alors r la probabilité d'obtenir un nombre i d'inversions strictement supérieur à celui observé réalisation de la variable aléatoire I, nous obtenons alors :

$$o(C3) = Prob\{I > 0\} = \frac{5}{6} \approx 0.833$$

Là encore, cette valeur suggère que nous pouvons considérer avec un niveau de cohérence très satisfaisant. Cette valeur est obtenue par tabulation (annexe 6, m=3)

a) De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement :V042[Bons pour certains élèves et néfastes pour d'autres]
b) Votre/vos salles(s) de classe est/sont disposée(s)... V014[En îlots]
c) Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre... V030[De l'aide ou du tutorat]

Tableau 16 : contenu sémantique des variables binaires de la classe C3

Partant du point de que les échanges entre élèves bien que non maîtrisés par l'enseignant pendant le temps d'enseignement sont pour bons certains mais néfastes pour d'autres, requiert une disposition de la salle de classe en ilots, alors ils ont tendance à interpréter ces interactions dans l'ordre de l'aide ou du tutorat.

Caractéristiques définies par les variables supplémentaires	**Contribution à la construction de la classe avec un niveau de risque**	**Nombre d'individus appartenant au groupe optimal d'effectif = 8**
Age4= De 52,9 ans à 63,3 ans	0.208	2
Ancien4= de 30 à 42 années	0.0547	4
Homme	0.0148	4
V005 = pédagogie non traditionnelle identifiée	0.0398	4
V023= Classe à deux niveaux	0.0136	5

Tableau 17 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C3

Cette classe C3 est principalement déterminée par les enseignants les plus âgés et les plus anciens professionnellement parmi les individus ayant répondu et pratiquant une pédagogie non traditionnelle identifiée dans des classes à deux niveaux.

L'analyse de la classe C3 fait apparaitre une potentielle troisième catégorie d'enseignants, moins prononcée que les deux profils mis en lumière lors de l'analyse de nos sous-réseaux de relations quasi-implicatives. Ce troisième profil serait alors plutôt représentatif d'individus qui, comme ceux du sous-réseau 1, sont relativement âgés et anciens dans leur métier, et utilisent plutôt des pédagogies non traditionnelles identifiées, mais qui, comme ceux du sous-réseau 2, portent un regard mitigé voire méfiant sur les interactions entre élèves non maîtrisées par l'enseignant. Ces enseignants étant ceux qui ont plutôt des classes à deux niveaux (contre trois niveaux ou plus pour le profil 1 et un niveau unique pour le profil 2), nous pouvons nous demander si ce facteur n'aurait pas un rôle important à jouer dans la place et la considération accordée aux interactions entre pairs non-maitrisées par l'enseignants.

7.4 La classe C4={(V026souvent⇒V039)⇒V34_3}

Cette classe de degré 3 est formée au nœud n°5 avec une valeur de la cohésion coh(C4)=0.962. Elle regroupe 3 variables orientées par la relation de quasi-implication. Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Variables binaires ordonnées dans la classe	V026Souvent	V039	V034_3
Occurrences	12	20	23
Ordonnancement issu du traitement de l'implication statistique	1	2	3

Tableau 18 : ordonnancement des variables selon les occurrences

La cohérence o(C4) de la classe C4 est alors la probabilité d'obtenir un nombre i d'inversions strictement supérieur à celui observé réalisation de la variable aléatoire I, nous obtenons alors :

$$o(C4) = Prob\{I > 0\} = \frac{5}{6} \approx 0.833$$

Là encore, cette valeur suggère que nous pouvons considérer avec un niveau de cohérence très satisfaisant. Cette valeur est obtenue par tabulation (annexe 6, m=3)

a) V026_Diriez-vous qu'il arrive que vos élèves interagissent entre eux sans que vous maitrisiez le contenu de ces échanges ? V026souvent
b) De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement : V039[Un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves]
c) V034_Quelle est votre réaction lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas ? V034_3[Vous les observez afin d'en comprendre la nature et le sujet avant de décider d'intervenir ou non]

Tableau 19 :contenu sémantique des variables binaires de la classe C4

Puisque partant du fait que souvent les élèves interagissent pendant le temps d'enseignement sans que l'enseignant n'en maîtrise le contenu alors ce dernier considère ces échanges comme un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves. Tout cela conduit à la réaction de l'enseignant consistant à observer afin de comprendre la nature et le sujet de ces échanges avant de décider d'intervenir ou non.

Caractéristiques définies par les variables supplémentaires	**Contribution à la construction de la classe avec un niveau de risque**	**Nombre d'individus appartenant au groupe optimal d'effectif =**
Age1= De 22 ans à 29.5 ans	0.0941	7
Ancien1= Moins de 5 années	0.119	7
V004 = pédagogie traditionnelle	0.00502	6
V023= Classe à deux niveaux	0.0575	6

Tableau 20 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C4

Cette classe C4 est principalement déterminée par les enseignants les plus jeunes et les plus nouveaux professionnellement parmi les individus ayant répondu, et pratiquant une pédagogie traditionnelle dans des classes à deux niveaux.

Il est intéressant d'interpréter ces résultats en les comparant à ceux de la classe C1. Les sondés concernés par les classe C1 et C4 ont en commun d'être plutôt jeunes et avec peu d'ancienneté dans l'enseignement, et d'avoir plutôt des classes à deux niveaux. Si les enseignants de la classe C1 pratiquent plutôt une pédagogie mixte et se montrent méfiants envers les interactions entre élèves non maîtrisées par l'enseignant, ceux de la classe C4 pratiquent plutôt une pédagogie traditionnelle et considèrent plutôt ces interactions comme un outil pédagogique pertinent. Cette analyse comparative est surprenante car nous pourrions nous attendre à ce que les interactions soient vues plus positivement par des enseignants qui utilisent une pédagogie mixte que par ceux qui utilisent une pédagogie

traditionnelle. Nous pouvons ainsi faire l'hypothèse, à la lumière de la faible ancienneté des enseignants contribuant aux classes C1 et C4, que le choix pédagogique de l'enseignant qui opte pour une pédagogie mixte en début de carrière peut le mettre en difficulté face à la gestion de son groupe-classe.

7.5 La classe C5={V027rarement⇒(V007⇒V028)}

Cette classe de degré 3 est formée au nœud n°8 avec une valeur de la cohésion coh(C5)=0.884. Elle regroupe 3 variables orientées par la relation de quasi-implication. Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Variables binaires ordonnées dans la classe	V027rarement	V007	V028
Occurrences	3	13	23
Ordonnancement issu du traitement de l'implication statistique	1	2	3

Tableau 20 : ordonnancement des variables selon les occurrences

La cohérence o(C5) de la classe C5 est alors la probabilité d'obtenir un nombre i d'inversions strictement supérieur à celui observé réalisation de la variable aléatoire I, nous obtenons alors :

$$o(C5) = Prob\{I > 0\} = \frac{5}{6} \approx 0.833$$

Là encore, cette valeur suggère que nous pouvons considérer avec un niveau de cohérence très satisfaisant. Cette valeur est obtenue par tabulation (annexe 6, m=3)

a) V027_A quelle fréquence se produisent des échanges que vous ne maitrisez pas en classe ? V027rarement
b) Dans votre (vos) classe(s), les élèves… V007[Ont une place attitrée à laquelle ils travaillent tout le temps ou la plupart du temps]
c) Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre... V028[Du bavardage ou de l'échange récréatif, elles ne concernent pas le travail]

Tableau 21 : contenu sémantique des variables binaires de la classe C5

Partant de l'idée que les échanges non maîtrisés par l'enseignant ne se produisent que rarement, alors l'organisation de la classe est basée sur une structure dans laquelle les élèves ont une place attitrée

Caractéristiques définies par les variables supplémentaires	**Contribution à la construction de la classe avec un niveau de risque**	**Nombre d'individus appartenant au groupe optimal d'effectif = 10**
Age3= De 41,75 ans à 50,5 ans	0.0512	4
Ancien3= 15 à 26 années	0.0274	5
V004 = pédagogie traditionnelle	0.00461	5
V022= Classe à un niveau	0.000663	8

Tableau 22 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C5

Cette classe C5 est principalement déterminée par les enseignants quadragénaires ayant entre 15 et 26 années d'ancienneté professionnelle, c'est-à-dire cela les situe environs entre un tiers (1/3) et deux tiers (2/3) d'une carrière de 42 années, et pratiquant

une pédagogie traditionnelle dans des classes à un niveau. Les enseignants qui contribuent à la classe C5 semblent donc se raccrocher au profil établi après l'analyse du sous-réseau 2 de relations quasi implicatives, bien qu'ils aient moins d'ancienneté.

7.6 La classe C6={V016⇒V036}

Cette classe de degré 1 est formée au nœud n°12 avec une valeur de la cohésion coh(C6)=0.806. Cette classe regroupe 2 variables reliées par une règle simple.

Nous pouvons remarquer qu'il n'y a aucune inversion par rapport à l'ordonnancement obtenu par le traitement des graphes implicatifs.

Variables binaires ordonnées dans la classe	V016	V036
Occurrences	10	10
Ordonnancement issu du traitement de l'implication statistique	1 ex-aequo	1 ex-aequo

Tableau 23 : ordonnancement des variables selon les occurrences

Comme le montre la représentation graphique ci-dessous extraite de l'arbre cohésitif, nous avons là une classe de quasi-équivalence.

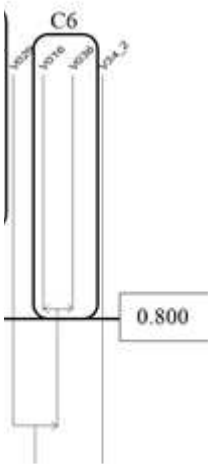


Figure 22 : Classe C6 extraite de la représentation graphique de l'arbre cohésitif

a) Votre/vos salles(s) de classe est/sont disposée(s)... V016[Avec plusieurs dispositifs qui cohabitent (type « flexible seating »)]
b) De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement : V036[Indispensables pour la cohésion du groupe]

Tableau 24 : contenu sémantique des variables binaires de la classe C6

Une règle se forme à ce niveau de la hiérarchie selon laquelle les enseignants qui organisent leur salle de classe selon plusieurs dispositifs qui cohabitent de type « flexible seating » ont une propension à considérer les échanges entre élèves comme Indispensables pour la cohésion du groupe. Nous pourrons ajouter : et réciproquement.

Caractéristiques définies par les variables supplémentaires	Contribution à la construction de la classe avec un niveau de risque	Nombre d'individus appartenant au groupe optimal d'effectif =
Age3= De 41,75 ans à 50,5 ans	0.088	2
Age4= De 52,9 ans à 63,3 ans	0.072	2
Ancien4= 30 à 42 années	0.11	1
Femme	0.208	5
V005 = pédagogie non traditionnelle identifiée	0.0312	3
V024= Classe à au moins trois niveaux	0.0332	2

Tableau 25 : Contributions des groupes d'individus définis par les variables supplémentaires à la construction de la classe C6

Pour la construction de cette classe, les contributions proviennent plutôt des enseignantes appartenant aux deux tranches d'âges les plus élevés, âgées de plus de 40 ans et ayant la plus forte ancienneté, plus de 30 années, qui auraient une tendance à pratiquer une pédagogie non traditionnelle identifiée dans des classes d'au moins trois niveaux.

La classe C6 nous apporte une information complémentaire sur le profil identifié lors de l'analyse du sous-réseau 1 de relations quasi-implicatives. Dans le sous-réseau 1, nous n'avions en effet pas d'information concernant le placement des élèves. Nous retrouvons ici un profil d'enseignants similaire à celui que nous avons pu construire en analysant le sous-réseau 1, et apprenons que ceux-ci disent majoritairement organiser leur salle de classe en installant plusieurs dispositifs de placement des élèves qui cohabitent. La mise en avant de cette variable est cohérente avec le choix d'une pédagogie non traditionnelle identifiée, puisque ces pédagogies proposent généralement à l'élève de se placer et de se déplacer librement.

8 Conclusion

Le recours au cadre théorique de l'analyse statistique implicative – ASI instrumentée par le logiciel CHIC 7 pour traiter nos données nous a permis de mettre en évidence des liens non symétriques forts que nous pouvons interpréter comme des liens de causalité d'une point de vue probabiliste. Certes ces résultats développés dans le cadre théorique de l'ASI ne sont relatifs qu'à l'échantillon étudié dont la taille et le caractère de convenance, dans la mesure où il n'a pas été construit par des procédures aléatoires, incitent à une grande prudence pour être généralisés à une population plus vaste d'enseignants. Cependant ils sont suffisamment intéressants pour donner lieu à une recherche de plus grande ampleur exploitant, en particulier, les ressources de la didactique des mathématiques (transposition, milieu, théorie anthropologique, …).

À travers cet échantillon, il ressort qu'il peut exister un lien notable entre les choix pédagogiques de l'enseignant et la manière dont il considère les interactions entre élèves qu'il ne maitrise cependant pas dans la classe. Comme nous l'avions supposé, nous voyons apparaitre deux profils d'enseignants très distincts. Le premier profil se rapporte à des enseignants qui travaillent plutôt en classes multi-niveaux et en utilisant une pédagogie non traditionnelle ou encore dite alternative, du type pédagogie Freinet, par exemple, ou autre. Notons que nous n'avons pas questionné les enseignants sur le type précis de pédagogie non traditionnelle à laquelle ils font implicitement référence dans

leur pratique pédagogique. Ces enseignants qui déclarent recourir à des pratiques de pédagogie non traditionnelles, déclarent laisser une grande possibilité de choix à leurs élèves au sein de la salle de classe en leur permettant d'interagir très librement pendant les temps d'apprentissage. Ils considèrent que les interactions entre élèves non maîtrisées par l'enseignant constituent un processus interne au dispositif pédagogique qui soutient la relation d'aide entre les sujets-élèves-apprenants et qui est un appui pour le sujet-enseignant en charge de la classe. Le second profil qui apparait, identifie des enseignants qui travaillent plutôt dans des classes à un seul niveau et avec une pédagogie traditionnelle. Ce profil concerne plutôt des enseignants qui ont une propension à déterminer eux-mêmes quand, où et comment doivent travailler les élèves. Ils semblent plutôt réfractaires aux interactions entre élèves qu'ils ne maitriseraient pas. Ils ne les bannissent pas totalement car ils considèrent toutefois que ces interactions ne peuvent pas être tout le temps évitées et qu'elles peuvent parfois même s'avérer bénéfiques, mais ces enseignants auront plutôt tendance à reprendre le contrôle sur ces interactions lorsqu'elles se produisent et à considérer qu'elles ne concernent pas les apprentissages, voire qu'elles les freinent.

L'identification de ces deux profils semble corroborer notre hypothèse de départ, qui supposait que plus il y a d'interactions entre élèves non maitrisées par l'enseignant dans une classe, plus la réaction de l'enseignant tendra à les permettre et à les inclure dans les processus d'enseignement et d'apprentissage. En effet, il est clair qu'il y aura beaucoup plus d'interactions entre élèves dans une classe multi-niveaux utilisant, par exemple, la pédagogie Freinet ou celle inspirée des conceptions de Montessori, qui permettent une plus grande liberté de choix et de déplacement dans la classe et telle que l'on peut en rencontrer chez les enseignants du premier profil, que dans une classe à niveau simple organisée selon les principes d'une pédagogie traditionnelle, où les élèves travaillent majoritairement à une place attribuées et avec des voisins fixes.

Nous pouvons néanmoins soulever une nouvelle interrogation : comment expliquer que les enseignants qui ont des classes multi-niveaux, traditionnellement réputées plus difficiles à mener, soient ceux qui ont une propension à choisir de recourir aux interactions entre élèves qu'ils reconnaissent ne pas maitriser comme un processus à prendre en compte dans l'organisation pédagogique plutôt que de vouloir les contenir davantage qu'en simple niveau ? Deux suppositions peuvent être faites : peut-être les enseignants qui sont favorables *a priori* à laisser se produire ces interactions entre élèves sans pour autant les maitriser, sont-ils plus enclins à choisir des classes à niveaux multiples pour encourager la coopération entre des élèves de plusieurs classes d'âge. Il est également possible d'imaginer que peut-être ces enseignants, dont on a remarqué qu'ils relèvent plutôt de la catégorie des plus âgés et des plus anciens professionnellement, contraints d'accepter une classe multi-niveaux, ont constaté par eux-mêmes un bénéfice à laisser intervenir des interactions entre pairs qu'ils ne maitrisaient pas. Toutes ces questions constituent des sources pour développer des pistes de recherche dans le champ des sciences de l'éducation et de la formation mais aussi dans ceux des didactiques disciplinaires et de la didactique professionnelle des enseignants.

Références

[40] Ministère de l'Éducation nationale, de la Jeunesse et des Sports. (2020). *Bulletin officiel n°31 du 30 juillet 2020 - Programme d'enseignement de l'école maternelle*. https://cache.media.education.gouv.fr/file/31/89/4/ensel712_annexe_1312894.pdf

[41] Berzin, C. (2005). Interactions entre pairs et apprentissages à l'école maternelle. Le cas du tutorat : Intérêt et limites. *Spirale - Revue de recherches en éducation*, *36*(1), 7‑15. https://doi.org/10.3406/spira.2005.1320

[42] Cuisinier, F., Gauducheau, N., Clavel, C., Costet, J.-P., & Pelloux, P. (2007). Psychologie des interactions sociocognitives et apprentissages scolaires : Regards croisés de pédagogues et de chercheurs. *Les Sciences de l'education - Pour l'Ere nouvelle*, *Vol. 40*(4), 47‑79.

[43] 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.

[44] Gras, R., Régnier, JC., Lahanier-Reuter, D. Marinica, C., Guillet, F. (Eds) (2017). *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales*. Toulouse : Cépaduès

[45] Gras, R., Régnier, J.C., (2017). Fondements théoriques de l'A.S.I. In R. Gras, J.C. Régnier, D. Lahanier-Reuter, C. Marinica, F. Guillet, (Eds) (2017) *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales.* Toulouse : Cépaduès p.43-61

[46] Gras, R., (2018) *La théorie de l'analyse statistique implicative ou l'invraisemblance du faux*. Toulouse : Cépaduès

[47] Gueguen, C., & Ansembourg, T. d'. (2014). *Pour une enfance heureuse : Repenser l'éducation à la lumière des dernières découvertes sur le cerveau*. Robert Laffont.

[48] Pléty, R., & Cosnier, J. (1985). *Ethologie de l'interaction chez des enfants du 1er cycle de l'enseignement secondaire au cours d'un apprentissage des mathématiques en groupes dans la résolution de problèmes*. s.n.

[49] Schubauer-Leoni, M.-L., Perret-Clermont, A.-N., & Mugny, G. (Ed.). (1985). Interactions sociales dans l'apprentissage de connaissances mathématiques chez l'enfant. In *Psychologie sociale du développement cognitif* (p. 225‑250). Peter Lang.

[50] Vasquez, A., & Martinez, I. (1990). Interactions élève-élève. Un aspect non perçu de la socialisation. *Enfance*, *43*(3), 285‑301. https://doi.org/10.3406/enfan.1990.1949

Annexe 1

LimeSurvey
Version 3.17.5+190604

Enquête par questionnaire auprès des enseignants de l'école primaire relative aux représentations à l'égard des diverses interactions qui se produisent au sein de la classe

G01 A propos de vous

1. **Votre genre :** *Féminin / Masculin*
2. **Quelle est votre date de naissance ?**
3. **Vous travaillez dans le contexte ... :** *enseignement public / enseignement privé sous contrat / enseignement privé hors contrat / Autre :*
4. **Votre statut professionnel est ... :** *Enseignant titulaire de l'enseignement public / Enseignant titulaire de l'enseignement privé sous contrat / Enseignant contractuel de l'enseignement public / Enseignant suppléant de l'enseignement privé sous contrat / Enseignant stagiaire de l'enseignement public / Enseignant stagiaire de l'enseignement privé sous contrat / Enseignant en CDD de l'enseignement privé hors contrat / Enseignant en CDI de l'enseignement privé hors contrat / Enseignant bénévole de l'enseignement privé hors contrat / Autre:*
5. **Depuis combien d'années enseignez-vous ?**
6. **Pour l'année scolaire en cours, vous avez en charge... :** *Une classe avec direction de l'école / Une classe sans direction de l'école / Plusieurs classes en tant que complément de service / Plusieurs classes car assumant des remplacements / Autre :*
7. **Quel est votre temps de travail ?** *Temps plein / Temps partiel : 80% / Temps partiel : 75% / Mi-temps hebdomadaire : 50% / Mi-temps annualisé : 50% / Temps partiel : 25% / Autre :*

G02 Vous et votre expérience professionnelle

1. **Pour préparer votre classe vous avez principalement recours aux ressources pédagogiques selon l'ordre de fréquence suivant : (Numérotez chaque case dans l'ordre de vos préférences de 1 à 8)**

Les manuels scolaires
Les ouvrages pédagogiques destinés aux enseignants tels que le livre pour le maitre, etc.
Les ouvrages de pédagogie ou de didactique écrits par des pédagogues, des professionnels ou des chercheurs
Les ressources gratuites accessibles sur Internet
Les ressources payantes accessibles sur Internet
Vos propres ressources
Les ressources collectées lors de sessions de formation
Les ressources partagées entre collègues

2. **Pourriez-vous citer les références de quelques ressources que vous utilisez pour vos préparations de classe ? (Au plus 5)**
3. **De quelles méthodes ou démarches pédagogiques vous inspirez-vous le plus ? Vous diriez que l'organisation pédagogique de votre (vos) classe(s) ...**

a) *Relève plutôt de caractéristiques d'une pédagogie traditionnelle /b) Relève plutôt de caractéristiques d'une pédagogie non traditionnelle identifiée (Exemple : classe ou école démocratique, Freinet, Montessori, Steiner, etc.) /c) Relève plutôt de caractéristiques d'une pédagogie mixte qui s'inspire de différents courants pédagogiques (traditionnel, Montessori, Freinet, etc.) /d) Autre :*

G03 Les modalités de travail en classe

1. **Dans votre (vos) classe(s), les élèves... (ne cochez qu'une case sauf si vous avez plusieurs classes)**

Ont une place attitrée à laquelle ils travaillent tout le temps ou la plupart du temps / Changent régulièrement de place selon le travail qui leur est demandé / Choisissent leur place et en changent librement / Autre :

2. **Votre/vos salles(s) de classe est/sont disposée(s)... (ne cochez qu'une case sauf si vous avez plusieurs classes)**

En bureaux individuels (séparés les uns des autres) / En rangs de deux bureaux / En rangées de trois bureaux ou plus / En U / En îlots / En coins (par matières ou domaines) / Avec plusieurs dispositifs qui cohabitent (type « flexible seating ») / Autre :

3. **Quelles modalités de travail sont les plus utilisées dans votre/vos classe(s) ? (Numérotez chaque case dans l'ordre de vos préférences de 1 à 6)**

Le travail individuel / Le travail en binôme / Le travail en groupes de trois / Le travail en groupes de quatre ou cinq / Le travail en demi-classe / Le travail en classe entière

4. **La plupart du temps, pendant un temps de classe donné :**

Vous déterminez une modalité de travail qui est la même pour tous les élèves / Vous déterminez plusieurs modalités de travail qui cohabitent / Les élèves choisissent leur modalité de travail / Autre :

5. **Lorsqu'ils travaillent à plusieurs, vos élèves sont plutôt répartis : (ne cochez qu'une case sauf si vous avez plusieurs classes)**

Par niveaux scolaires homogènes / Par niveaux scolaires hétérogènes / L'un ou l'autre selon le travail que vous leur demandez / L'un ou l'autre selon leur choix / Autre :

6. **En classe, vos élèves travaillent sur des tâches, activités ou ateliers... (Numérotez chaque case dans l'ordre de vos préférences de 1 à 7, le rang 1 étant le plus fréquent et le rang 7 le moins fréquent)**

a) *dirigés imposés /b) semi-dirigés imposés /c) autonomes imposés /d) dirigés à inscription/choix libre /e) semi-dirigés à inscription/choix libre /f) autonomes à inscription/choix libre /g) . qu'ils déterminent eux-mêmes*

7. **Dans une même classe vous avez : (ne cochez qu'une seule case sauf si vous avez plusieurs classes)**

a) Un seul niveau (exemple : CP) /b) Deux niveaux (exemple : CE1-CE2 ou 4-5 ans) /c) Trois niveaux ou plus (exemple : PS-MS-GS, 6-10 ans) /d) Autre :

G04 Les interactions en classe

1. **En trois mots, qu'évoque pour vous le terme « interaction » dans le contexte de la classe ?**
2. **Pendant les temps de travail en classe, vos élèves sont-ils autorisés à échanger entre eux ?**

a) *Oui, sans restriction /b) Oui avec parcimonie et à voix basse /c) Oui mais uniquement lorsque la tâche qui leur est demandée le nécessite /d) Oui mais uniquement sous le contrôle de l'adulte /e) Oui mais uniquement lorsqu'ils ont terminé leur travail et sans déranger le reste de la classe /f) Non /g)Autre :*

3. **Diriez-vous qu'il arrive que vos élèves interagissent entre eux sans que vous maitrisiez le contenu de ces échanges ?** *Très souvent / Souvent / Parfois / Rarement / Jamais*
4. **A quelle fréquence se produisent des échanges que vous ne maitrisez pas en classe ?**

Très souvent / Souvent / Parfois / Rarement / Jamais

5. **Selon vous, quelles conséquences ces échanges que vous ne maitrisez pas ont-ils sur les apprentissages ?**
6. **Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre...**

a) Du bavardage ou de l'échange récréatif, elles ne concernent pas le travail /b) De la dispute /c) De l'aide ou du tutorat /d) De la triche /e) De la coopération ou de l'entraide, pour effectuer une tâche ou répondre à une question /f) De la sociabilisation, ce sont des échanges amicaux ou empathiques /g) Autre :

7. **Quelle est votre réaction lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas ?**

a) Vous intervenez afin de recadrer l'échange (sans forcément l'interdire) /b) Vous n'intervenez pas /c) Vous les observez afin d'en comprendre la nature et le sujet avant de décider d'intervenir ou non /d) Vous les interrompez car ces échanges ne sont pas autorisés/e) Autre :

8. **De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement :**

a)*Un outil d'apprentissage lorsqu'ils sont bien régulés /b) Indispensables pour la cohésion du groupe /c) Un problème qui freine la classe /d) Intolérables sauf s'ils ont été prévus et expressément demandés aux élèves /e) Un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves /f) Indispensables pour répondre au besoin de sociabilisation des enfants /g) Inévitables, mais il faut réussir à les contenir pour avancer /h) Bons pour certains élèves et néfastes pour d'autres /i) Autre :*

Annexe 2

Vecteurs-Variables	**Code des variables binaires à valeurs dans {0 ; 1}, composantes des vecteurs**	**Statut P= principale S = supplémentaire**
V001_Sexe [G01-1]	V001Femme V001Homme	S
V002_Age [G01-2]	Age1 = De 22 ans à 29.5 ans Age2 =De 31,75 ans à 40 ans Age3 = De 41,75 ans à 50,5 ans Age4 = De 52,9 ans à 63,3 ans	S
V003_Ancienneté professionnelle [G01-5]	Ancien1 = Ancienneté de Moins de 5 années Ancien2 = Ancienneté de 6 à 13 années Ancien3 = Ancienneté de 15 à 26 années Ancien4 = Ancienneté de 30 à 42 années	S
V004/5/6_Vous diriez que l'organisation pédagogique de votre (vos) classe(s) relève plutôt de caractéristiques d'une… [G02-3]	V004 [pédagogie traditionnelle] V005 [pédagogie non traditionnelle identifiée (Exemple : classe ou école démocratique, Freinet, Montessori, Steiner, etc.)] V006 [pédagogie mixte qui s'inspire de différents courants pédagogiques (traditionnel, Montessori, Freinet, etc.)]	S
V007/8/9_Dans votre (vos) classe(s), les élèves... [G03-1]	V007 [Ont une place attitrée à laquelle ils travaillent tout le temps ou la plupart du temps] V008 [Changent régulièrement de place selon le travail qui leur est demandé] V009 [Choisissent leur place et en changent librement]	P
V010/…/16_Votre/vos salles(s) de classe est/sont disposée(s)... [G03-2]	V010 [En bureaux individuels (séparés les uns des autres)] V011 [En rangs de deux bureaux] V012 [En rangées de trois bureaux ou plus] V013 [En U] V014 [En îlots] V015 [En coins (par matières ou domaines)] V016 [Avec plusieurs dispositifs qui cohabitent (type « flexible seating »)]	P
V018/…/21_La plupart du temps, pendant un temps de classe donné, lorsqu'ils travaillent à plusieurs, vos élèves sont plutôt répartis [G03-5]	V018 [Par niveaux scolaires homogènes] V019 [Par niveaux scolaires hétérogènes] V020 [L'un ou l'autre selon le travail que vous leur demandez] V021 [L'un ou l'autre selon leur choix]	P
V022/…/24_Dans une même classe vous avez … [G03-7]	V022 [Un seul niveau (exemple : CP)] V023 [Deux niveaux (exemple : CE1-CE2 ou 4-5 ans)] V024 [Trois niveaux ou plus (exemple : PS-MS-GS, 6-10 ans)]	S
V025_Statut pédagogique des interactions entre élèves dans la classe. Pendant les temps de travail en classe, vos élèves sont-ils autorisés à échanger entre eux ? [G04-2]	V025NR V025Autre V025Ouiavec V025Ouisans V025Ouimais	P

On parle d'échanges entre élèves. Ces échanges peuvent être verbaux, gestuels, documentaires, ou de l'ordre de l'observation (un élève montre à un autre). On entend par « maitriser les échanges » le fait d'en connaître au moins partiellement le contenu ou le sujet.		
V026_Diriez-vous qu'il arrive que vos élèves interagissent entre eux sans que vous maitrisiez le contenu de ces échanges ? [G04-3]	V026NR V026jamais V026rarement V026parfois V026souvent V026Trèssouvent	P
V027_A quelle fréquence se produisent des échanges que vous ne maitrisez pas en classe ? [G04-4]	V027NR V027jamais V027rarement V027parfois V027souvent V027Trèssouvent	P
Selon vous, les interactions que vous ne maitrisez pas entre vos élèves sont majoritairement de l'ordre... [G04-6]		
V028	V028 [Du bavardage ou de l'échange récréatif, elles ne concernent pas le travail]	P
V029	V029 [De la dispute]	P
V030	V030 [De l'aide ou du tutorat]	P
V031	V031 [De la triche]	P
V032	V032 [De la coopération ou de l'entraide, pour effectuer une tâche ou répondre à une question]	P
V033	V033 [De la sociabilisation, ce sont des échanges amicaux ou empathiques]	P
V034_Quelle est votre réaction lorsque vous constatez des échanges entre vos élèves que vous ne maitrisez pas ? [G04-7]	V034_1 [Vous intervenez afin de recadrer l'échange (sans forcément l'interdire)] V034_2 [Vous les interrompez car ces échanges ne sont pas autorisés] V034_3 [Vous les observez afin d'en comprendre la nature et le sujet avant de décider d'intervenir ou non] V034_4 [Vous n'intervenez pas] V034_Autre [Autre] V034_NR [Non réponse]	P
V035/…/42_De votre point de vue, les échanges entre élèves pendant les temps d'enseignement sont principalement : [G04-8]	V035 [Un outil d'apprentissage lorsqu'ils sont bien régulés] V036 [Indispensables pour la cohésion du groupe] V037 [Un problème qui freine la classe] V038 [Intolérables sauf s'ils ont été prévus et expressément demandés aux élèves] V039 [Un relais pour l'adulte qui ne peut pas répondre seul aux besoins de tous les élèves] V040 [Indispensables pour répondre au besoin de sociabilisation des enfants] V041 [Inévitables, mais il faut réussir à les contenir pour avancer] V042 [Bons pour certains élèves et néfastes pour d'autres]	P

Tableau 26 : Description des vecteurs-variables codés issus du questionnaire

Annexe 3 : Traitement de la relation de quasi-équivalence entre les variables binaires de la base de données.

Variable X	Variable Y quasi-équivalente	Variable X	Variable Y quasi-équivalente
V007	V028	V026NR	V027NR
V007	V025Ouimais	V026NR	V34_NR
V008	V014	V026parfois	V027parfois
V008	V34_3	V026souvent	V027souvent
V009	V027Tsouvent	V026souvent	V039
V020	V025Ouimais	V026Tsouvent	V027Tsouvent
V020	V027parfois	V026Tsouvent	V040
V020	V028	V027NR	V34_NR
V020	V34_3	V027souvent	V039
V020	V041	V027Tsouvent	V040
V021	V026Tsouvent	V028	V042
V021	V032	V032	V033
V021	V027Tsouvent	V032	V035
V025NR	V027NR	V032	V039
V025NR	V026NR	V032	V040
V025NR	V34_NR	V033	V039
V025Ouiavec	V033	V033	V040
V025Ouiavec	V039	V033	V035
V025Ouimais	V028	V34_3	V039
V025Ouimais	V041		

Tableau 27

Variable X	Représentant de la classe de quasi-équivalence
V008	V014
V025NR	V026NR
V027NR	V026NR
V34_NR	V026NR
V027parfois	V026parfois
V027souvent	V026souvent
V027Tsouvent	V026Tsouvent
V040	V026Tsouvent
V035	V032
V025Ouiavec	V039
V033	V039
V020	V041
V025Ouimais	V041
V028	V042

Tableau 28

Annexe 4 : Contribution au chemin : V025Ouisans-V021-V032

Groupe optimal : Ind_065 Ind_033 Ind_029 Ind_079 Ind_073 Ind_070

card GO= 6 p = 0.1 1-p = 0.9

Age	La variable Age1 contribue à ce chemin avec un risque de : 0.635 intersection avec le groupe optimal 1 La variable Age2 contribue à ce chemin avec un risque de : 0.415 intersection avec le groupe optimal 1 La variable Age3 contribue à ce chemin avec un risque de : 0.379 intersection avec le groupe optimal 1 La variable Age4 contribue à ce chemin avec un risque de : 0.0256 intersection avec le groupe optimal 3
Ancienneté professionnelle	La variable Ancien1 contribue à ce chemin avec un risque de : 0.38 intersection avec le groupe optimal 2 La variable Ancien2 contribue à ce chemin avec un risque de : 0.815 intersection avec le groupe optimal 0 La variable Ancien3 contribue à ce chemin avec un risque de : 0.0556 intersection avec le groupe optimal 3 La variable Ancien4 contribue à ce chemin avec un risque de : 0.15 intersection avec le groupe optimal 1
Sexe	La variable Fem contribue à ce chemin avec un risque de : 0.531 intersection avec le groupe optimal 4 La variable Masc contribue à ce chemin avec un risque de : 0.111 intersection avec le groupe optimal 2
Organisation pédagogique de la classe	La variable V004 contribue à ce chemin avec un risque de : 0.686 intersection avec le groupe optimal 0 La variable V005 contribue à ce chemin avec un risque de : 0.000311 intersection avec le groupe optimal 6 La variable V006 contribue à ce chemin avec un risque de : 0.891 intersection avec le groupe optimal 0
Composition de la classe	La variable V022 contribue à ce chemin avec un risque de : 0.518 intersection avec le groupe optimal 1 La variable V023 contribue à ce chemin avec un risque de : 0.485 intersection avec le groupe optimal 1 La variable V024 contribue à ce chemin avec un risque de : 0.00833 intersection avec le groupe optimal 3

La variable qui contribue le plus à ce chemin est V005 avec un risque de : 0.000311

Annexe 5 : Contribution au chemin : V026Tsouvent-V021-V032

Groupe optimal : Ind_085 Ind_087 Ind_031 Ind_033 Ind_070 Ind_079 Ind_080
card GO = 7 p = 0.117 1-p = 0.883

Age	La variable Age1 contribue à ce chemin avec un risque de : 0.721 intersection avec le groupe optimal 1 La variable Age2 contribue à ce chemin avec un risque de : 0.824 intersection avec le groupe optimal 0 La variable Age3 contribue à ce chemin avec un risque de : 0.0124 intersection avec le groupe optimal 4 La variable Age4 contribue à ce chemin avec un risque de : 0.157 intersection avec le groupe optimal 2
Ancienneté professionnelle	La variable Ancien1 contribue à ce chemin avec un risque de : 0.745 intersection avec le groupe optimal 1 La variable Ancien2 contribue à ce chemin avec un risque de : 0.572 intersection avec le groupe optimal 1 La variable Ancien3 contribue à ce chemin avec un risque de : 0.0237 intersection avec le groupe optimal 4 La variable Ancien4 contribue à ce chemin avec un risque de : 0.192 intersection avec le groupe optimal 1
Sexe	La variable Fem contribue à ce chemin avec un risque de : 0.325 intersection avec le groupe optimal 6 La variable Masc contribue à ce chemin avec un risque de : 0.417 intersection avec le groupe optimal 1
Organisation pédagogique de la classe	La variable V004 contribue à ce chemin avec un risque de : 0.745 intersection avec le groupe optimal 0 La variable V005 contribue à ce chemin avec un risque de : 0.000805 intersection avec le groupe optimal 6 La variable V006 contribue à ce chemin avec un risque de : 0.721 intersection avec le groupe optimal 1
Composition de la classe	La variable V022 contribue à ce chemin avec un risque de : 0.879 intersection avec le groupe optimal 0 La variable V023 contribue à ce chemin avec un risque de : 0.285 intersection avec le groupe optimal 2 La variable V024 contribue à ce chemin avec un risque de : 0.00181 intersection avec le groupe optimal 4

La variable qui contribue le plus à ce chemin est V005 avec un risque de : 0.000805

Annexe 6 : Contribution au chemin : V007-V028

Groupe optimal : Ind_041 Ind_042 Ind_085 Ind_039 Ind_006 Ind_012 Ind_007 Ind_043 Ind_015 Ind_072
card GO = 10 p = 0.167 1-p = 0.833

Age	La variable Age1 contribue à ce chemin avec un risque de : 0.473 intersection avec le groupe optimal 3 La variable Age2 contribue à ce chemin avec un risque de : 0.421 intersection avec le groupe optimal 2 La variable Age3 contribue à ce chemin avec un risque de : 0.0512 intersection avec le groupe optimal 4 La variable Age4 contribue à ce chemin avec un risque de : 0.619 intersection avec le groupe optimal 1
Ancienneté professionnelle	La variable Ancien1 contribue à ce chemin avec un risque de : 0.3 intersection avec le groupe optimal 4 La variable Ancien2 contribue à ce chemin avec un risque de : 0.773 intersection avec le groupe optimal 1 La variable Ancien3 contribue à ce chemin avec un risque de : 0.0274 intersection avec le groupe optimal 5 La variable Ancien4 contribue à ce chemin avec un risque de : 0.721 intersection avec le groupe optimal 0
Sexe	La variable Fem contribue à ce chemin avec un risque de : 0.407 intersection avec le groupe optimal 8 La variable Masc contribue à ce chemin avec un risque de : 0.323 intersection avec le groupe optimal 2
Organisation pédagogique de la classe	La variable V004 contribue à ce chemin avec un risque de : 0.00461 intersection avec le groupe optimal 5 La variable V005 contribue à ce chemin avec un risque de : 0.935 intersection avec le groupe optimal 0 La variable V006 contribue à ce chemin avec un risque de : 0.123 intersection avec le groupe optimal 5
Composition de la classe	La variable V022 contribue à ce chemin avec un risque de : 0.000663 intersection avec le groupe optimal 8 La variable V023 contribue à ce chemin avec un risque de : 0.513 intersection avec le groupe optimal 2 La variable V024 contribue à ce chemin avec un risque de : 0.806 intersection avec le groupe optimal 0

La variable qui contribue le plus à ce chemin est V022 avec un risque de : 0.000663

Annexe 7 : Variable I_m nombre d'inversions dans une permutation de m éléments

		Nombre d'inversions															
Nb de permutations m !	m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	2	1	1														
6	3	1	2	2	1												
24	4	1	3	5	6	5	3	1									
120	5	1	4	9	15	20	22	20	15	9	4	1					
720	6	1	5	14	29	49	71	90	101	101	90	71	49	29	14	5	1

Tableau 27 : distribution des effectifs de la variable I, nombre d'inversions, en fonction de la taille m de la classe

LE LOGO ASI 11 COMMENTÉ…

Vladimir Lira Véras Xavier De Andrade

Quelques considérations sur la définition du logo :

1) Pour représenter la pandémie, nous considérons d'abord le deuil que représentent les vies perdues dans le monde et dans la communauté scientifique. Alors j'ai porté le noir qui symbolise le deuil.

2) La pandémie a conduit à accélérer de nombreuses recherches dans le monde (non seulement dans le domaine de la santé, mais dans l'éducation, la sociologie, la psychologie, entre autres) en fonction des effets du virus sur la société. Chaque recherche apporte en même temps des variables particulières, il devient évident que le problème d'une nation est un problème de toutes les nations (la solidarité contre l'égoïsme). Tant que le virus ne sera pas contenu dans le monde entier, de nouvelles variantes apparaîtront (le problème de mon voisin est aussi mon problème). Ce processus a conduit à la nécessité de développer des solutions systémiques. En ce sens, j'ai proposé une image simulant une perspective d'un ellipsoïde de révolution (représentant la terre) et les variables qui émergent des recherches dans diverses parties du monde et les relations entre elles dans un graphique implicite.

3) Au milieu de l'obscurité, une lumière au fond montre le rôle des sciences et de l'humanité dans la tentative de résoudre ces questions.

4) Les couleurs rouge, blanche et bleue sont en hommage au pays dans lequel l'ASI est née : la France.

Apresentamos algumas considerações sobre a definição da logomarca:

1) Para representar a pandemia, consideramos em primeiro lugar o luto que representa as vidas perdidas no mundo e na comunidade científica. Por isso, usei o preto que simboliza o luto.

2) A pandemia levou a acelerar inúmeras pesquisas no mundo (não apenas na área da saúde, mas na educação, na sociologia, psicologia, entre outras áreas) em função dos efeitos do vírus para a sociedade. Cada pesquisa traz variáveis particulares, ao mesmo tempo, fica evidente que o problema de uma nação é um problema de todas as nações (a solidariedade contra o egoísmo). Enquanto o vírus não for contido em todo o mundo, novas variantes surgirão (o problema do meu vizinho também é problema meu). Este processo tem levado à necessidade de desenvolvimento de soluções sistêmicas. Nesse sentido, propus uma imagem simulando uma perspectiva de um elipsoide de revolução (representando a terra) e as variáveis que surgem das pesquisas em diversas partes do mundo e as relações entre as mesmas em um gráfico implicativo.

3) Em meio à escuridão, uma luz no fundo mostra o papel das Ciências e da humanidade, na tentativa de resolver essas questões.

4) As cores vermelha, branca e azul são em homenagem ao país no qual surgiu a ASI: a França.

Algunas consideraciones sobre la definición del logo:

1) Usamos el color negro para representar el dolor del luto por las vidas perdidas en el mundo y en la comunidad científica que estamos viviendo hasta este momento por causa de la pandemia.

2) La pandemia ha propiciado una aceleración de numerosas investigaciones en el mundo (y no solo en el área de la salud) sino también en la educación, la sociología, la psicología debido a los efectos de la pandemia. Cada investigación se debate frente variables particulares. Al mismo tiempo, está claro que el problema de una nación es un problema para todas las naciones (solidaridad contra el egoísmo). Mientras el virus no esté contenido en el mundo, surgirán nuevas variantes (el problema de mi vecino es también mi problema). La necesidad de desarrollar soluciones sistemáticas. En ese sentido, tenemos la imagen que simula la perspectiva de un elipsoide de revolución (que representa la tierra) y las variables que surgen de la investigación en diferentes partes del mundo y las relaciones entre ellas en un gráfico implícito. En medio de la oscuridad, una luz de fondo muestra el papel de la ciencia y de la humanidad al tratar de resolver estos problemas.

3) Los colores rojo, blanco y azul están puestos en honor al país en el que surge el ASI: Francia.

Alcune considerazioni sulla definizione del logo:

1) Per rappresentare la pandemia, consideriamo innanzitutto il lutto rappresentato dalle vite perdute nel mondo e nella comunità scientifica. Così ho indossato il nero che simboleggia il lutto.

2) La pandemia ha portato ad accelerare numerose ricerche nel mondo (non solo nel campo della sanità, ma anche nell'istruzione, nella sociologia, nella psicologia, ecc.) in funzione degli effetti del virus sulla società. Ogni ricerca porta, allo stesso tempo, variabili particolari. Diventa evidente che il problema di una nazione è un problema di tutte le nazioni (la solidarietà contro l'egoismo). Finché il virus non sarà contenuto in tutto il mondo, appariranno nuove varianti (il problema del mio vicino è anche il mio problema). Questo processo ha portato alla necessità di sviluppare soluzioni sistemiche. In questo senso, ho proposto un'immagine che simula una prospettiva di un ellissoide di rivoluzione (rappresentante la terra) e le variabili che emergono dalle ricerche in varie parti del mondo e le relazioni tra di loro in un gráfico implicito.

3) In mezzo all'oscurità, una luce in fondo mostra il ruolo delle scienze e dell'umanità nel tentativo di risolvere questi problemi

4) I colori rosso, bianco e blu sono in omaggio al paese in cui è nata l'ASI: la Francia.

会徽寓意：

1）**首先我**们在此向疫情期间全球和学术界逝去的生命致哀，因此我们用黑色表示悼念。

2）**随着新冠病毒**对人类社会的影响，疫情推动了世界各国不仅在医学卫生领域，也在教育学、心理学、社会学以及其他领域的大量的研究。纵使每项研究都有其特定变量，**但**显然一个国家的问题也是世界各国的问题（团结而非利己）。病毒在全世界蔓延，新型变异病毒也随之出现。（邻居的问题亦是自己的问题）这一过程使得我们必须寻求系统性的解决方案，为此我们设计了类似于椭球体的图形来代表地球，变量来代表世界各地的研究，以及隐性图示来代表它们之间的联系。

3）**在黑暗之中，后面的光代表着科学和人**类在试图解决这些问题中的作用。4）红白蓝的色调意在向ASI（蕴涵统计分析）的诞生地法国致意

A few considerations on the definition of the logo:

1) In order to represent the pandemic, we consider first the bereavement that represents the lost lives in the world and in the scientific community. Thus, I chose the black colour, which symbolizes mourning.

2) The pandemic led to the acceleration of abundant research in the world (not only in the field of health, but also of education, sociology, psychology, among others) depending on the effects of the virus on society. Each research bringing about, at the same time, particular variables, it becomes obvious that the problem encountered by one nation is the problem of all nations (solidarity versus selfishness).As long as the virus is not restrained in the whole world, new variants will appear (my neighbour's problem is also mine).This process made it necessary to develop systemic solutions. Thus, I suggested a picture simulating the perspective of an ellipsoid of revolution (representing the Earth) and the variables emerging from research in different parts of the world as well as the relationships between one another in an implicit graph.

3) In the middle of darkness, in the background, a light shows the role of sciences and humankind attempting to solve those questions.

4) The red, blue and white colours pay tribute to the country in which ASI was born: France.

Логотип XI Коллоквиума по вопросам статистического импликативного анализа имеет глубокий смысл и символизирует несколько важных для сегодняшнего мира аспектов.

Во-первых, невозможно обойти стороной кризис, связанный с пандемией, унесшей жизни тысяч людей, в том числе и ученых, что является большой потерей для всего человечества. Черный фон логотипа отражает эту великую утрату.

Во-вторых, пандемия привела к ускорению темпа развития многих исследований во всем мире, не только в области здравоохранения, но и в сфере образования, социологии, психологии, что связано с воздействием вируса на общество. Каждое исследование приносит важные перемены и, таким образом, становится очевидно, что проблема одной нации — это проблема всего человечества в целом. Этот аспект демонстрирует глубокую солидарность мирового сообщества и стран друг к другу. До тех пор, пока в мире не выработается коллективный иммунитет, все новые штаммы вируса будут появляться в разных странах. Это означает, что проблемы, возникающие в соседних странах, будут касаться и нас самих. Данный процесс ведет к необходимости разработки системных решений. Изображение, представленное на логотипе имитирует Землю, погруженную в вихрь кардинальных перемен, которые принес этот год, а также эволюцию идей, возникающих в результате исследований в различных частях света и связь между ними.

В то же время, свет, изображенный на заднем фоне, показывает роль науки и человечества в попытках решения этих проблем.
И, наконец, красный, белый и синий цвета — это дань уважения стране, в которой родилась идея создания XI Коллоквиума по вопросам статистического импликативного анализа – Франции.

Logo ile ilgili bazı açıklamalar:

1) Salgın döneminde dünyada ve bilim camiasında kaybedilen insanların yaslarını temsil etmek için bu yil logomuz siyah renkte olacaktır.

2) Salgın, virüsün toplum üzerindeki etkilerine göre dünyada (sadece sağlık alanında değil, eğitimde, sosyolojide, psikolojide) birçok araştırmanın hızlanmasına yol açmıştır. Her araştırma aynı zamanda belirli değişkenleri de beraberinde getirir. Günümüzde anlıyoruz ki, bir ulusun problemi aynı zamanda diger ulusların problemi de olabilir (komşumun sorunu benim de sorunumdur). Bu süreç, bizlere sistemik çözümler geliştirme ihtiyacını doğurdu. Bu anlamda, dünyanın çeşitli yerlerindeki araştırmalardan ortaya çıkan değişkenleri ve aralarındaki ilişkileri simule eden dünya seklindeki grafikli logo ortaya çıktı.

3) Karanlığın ortasında arkadaki bir ışık, bu soruları çözmeye çalışan bilimin ve insanlığın rolünü göstermektedir.

4) Kırmızı, beyaz ve mavi renkler, ASI'nin doğduğu ülke olan Fransa'yı temsil etmektedir.

🇲🇦 بعض الاعتبارات لتحديد الشعار:

1) لتمثيل الجائحة وحدادا على الأرواح التي فقدها العالم عامة والساحة العلمية خاصة تم استعمال اللون الأسود.

2) أدت الجائحة إلى تسريع عدة بحوث في العالم ليس في المجال الصحي فقط وإنما في مجالات اخرى كالتربية وعلم الاجتماع وعلم لنفس وغيرها تبعا لتأثير الفيروس على المجتمع. كل بحث في نفس الوقت بمتغيرات خاصة، فصار واضحا أن مشكلة أمة ما هي مشكلة لباقي الأمم (التضامن ضد الأنانية). وطالما ان الفيروس لم يتم احتوائه في جميع أنحاء العالم فإنه ستظهر سلالات جديدة (مشكلة جاري هي مشكلتي أنا أيضا). هذه السيرورة تقتضي البحث عن حلول شاملة. لذلك اقترح صورة تحاكي الشكل الإهليليجي للثورة (تمثل الأرض) والمتغيرات التي تفرزها البحوث في مختلف بقاع العالم والعلاقات القائمة بينها عبر رسم بياني استلزامي.

3) وسط الظلام الحالك، يلوح في الأعماق ضوء يبرز دور العلوم والإنسانية في محاولة إيجاد حل هذه الأسئلة.

4) الألوان الأحمر والأبيض والأزرق هي تكريم للبلد الذي ولد فيه " التحليل الإحصائي الاستلزامي" : فرنسا.

INDEX DES CONTACTS DES AUTEURS

Ailin GAMONEDA	observatorioviolencia@outlook.com
Aitzol LASA	aitzol.lasa@unavarra.es
Ana Maria dos Anjos CARNEIRO-LEÃO	ana.acleao@gmail.com
André TOTOHASINA	andre.totohasina@gmail.com
Andrianasy Angelo DJISTERA	angelo.djistera@yahoo.fr
Antoine BODIN	antoinebodin@me.com
Aurino LIMA FERREIRA	aurinolima@gmail.com
Bakys Bruno RALAHADY	ralahadybru@yahoo.fr
Bruce Masonova SOLOZAFY BEMENA	brucebemena@gmail.com
Christian PELLOIS	pellois.christian@wanadoo.fr
Christian RAZANATSOAVINA	cmrazanatsoavina@gmail.com
Cristiana MARINHO DA COSTA	cmcmarinhos@gmail.com
Daniel Rajaonasy FENO	fenodaniel2@yahoo.fr
Eleni DELIYIANNi	deliyianni6@hotmail.com
Erika PEDERSEN	observatorioviolencia@outlook.com
Federico VASCHES	observatoriodepoliticasrn@gmail.com
Flávia PEREIRA RIGHI	flaviarighi87@gmail.com
Germán NIETO	al401111@uji.es
Gilbert SAPORTA	gilbert.saporta@cnam.fr
Greice SCREMIN	greicescremin@gmail.com
Haritz IRIBAS	haritz.iribas@unavarra.es
Iliada ELIA	elia.iliada@ucy.ac.cy
Iván BENZAQUÉN	observatorioviolencia@outlook.com
Jaione ABAURREA	jaione.abaurrea@unavarra.es
Jean-Claude LABERCHE	jeanclaudelaberche@gmail.com
Jean-Claude RÉGNIER	jean-claude.regnier@univ-lyon2.fr
Jeanne GUIET-SILVAIN	jeanne.guiet-silvain@u-paris.fr
Jeanne RAVAOSOLO	ufr.ravaosolo@moov.mg
Josoa Michel TOVOHERY	josoamicheltovohery@gmail.com
Larisa Zamora MATAMOROs	Larisa@uo.edu.cu
Leonardo DALLA PORTA	leo-dallaporta@hotmail.com
María CARDIERI	observatorioviolencia@outlook.com
Miguel WILHELMI	miguelr.wilhelmi@unavarra.es
Nadja Maria ACIOLY-RÉGNIER	nadja.acioly-regnier@univ-lyon1.fr
Nelsa María SAGARÓ DEL CAMPO	nelsa@infomed.sld.cu
Océane TOURON	oceane.touron@univ-lyon2.fr
Olga BELLETICH	olga.belletich@unavarra.es
Pablo CARRANZA	pfcarranza@gmail.com
Pablo GREGORI	gregori@uji.es
Parfait BEMARISIKA	bemarisikap7@yahoo.fr
Paula Virgínia CHAVES CABRAL ANDRADE	paulavchaves@gmail.com
Quércia CARVALHO ELOI	querciacarvalho2012@gmail.com
Rafael SANTOS DE AQUINO	rafael.aquino@ifsertao-pe.edu.br
Raphaël COUTURIER	raphael.couturier@univ-fcomte.fr
Régis GRAS	regispor@hotmail.fr
Saddo AG ALMOULOUD	saddoag@gmail.com
Sofia ANASTASIADOU	sofan@uom.edu.gr
Tolga TEKIN	tolgaatkin@gmail.com
Vladimir LIRA VÉRAS XAVIER DE ANDRADE	vladimir.andrade@ufrpe.br

Actes du 11ème colloque d'Analyse Statistique Implicative

Edition :
Université Bourgogne Franche-Comté - Besançon

Version numérique ISBN : 978-2-9562045-5-8

Version imprimée ISBN : 978-2-9562045-4-1

Année 2021