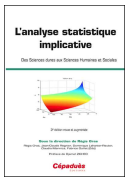

Extrait de la préface de l'ouvrage de **Gras, R., Régnier, JC., Lahanier-Reuter, D. Marinica, C., Guillet, F. (Eds) (2017)** *Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales.* Toulouse : Cépadues rédigée par **Djamel Abdelkhader Zighed**

« Pouvoir alimenter le débat autour du statut des connaissances découvertes par l'ASI ou des méthodes d'inférence ou de prédiction sœurs comme les réseaux bayesiens.

(…) Pour ma part, je vais me contenter de traiter [ce] point mais sans faire une thèse d'exégèse. La principale motivation vient en fait de l'argument avancé par les tenants du *Big Data (voir à ce propos le remarquable livre de Viktor Mayer-Schönberger et Kenneth Cukier paru chez Albert Laffont sous le titre « Big Data : La révolution des données est en marche »*. L'argument des ultras de l'empirisme (les big Data) pèse fortement sur le statut de la science. Il interroge directement celle-ci et particulièrement celle qui cherche à modéliser, à mettre dans une forme mathématique les lois qui gouvernent les mondes physiques et immatériels. L'A.S.I., comme bien d'autres démarches scientifiques, semble plus interpelée au regard des avancées du *Big Data*. Examinons quelques-unes des questions qu'inspire ce nouveau débat épistémologique :

- l'exhaustivité et le rôle de la statistique. Avec les moyens modernes de collecte des données et de calcul comme ceux que possèdent les géants de l'Internet tels que Google, Facebook, etc. avons-nous besoin de l'inférence statistique ?
- ne faut-il pas stocker plus de données et dépenser moins d'énergie à trouver de nouveaux algorithmes ? Deux exemples pour sortir cette question du registre de la provocation. (i.) Les progrès des systèmes de traduction automatique ou de complétion de mots couramment utilisés pour la saisie de textes ne sont pas venus de l'amélioration des algorithmes d'analyse linguistique mais du passage à des milliards de corpus disponibles. (ii.) Les modèles d'étude de la propagation de la grippe, via les réseaux sociaux, se sont avérés plus efficaces en termes de rapidité de détection et de mesure de diffusion que les modèles épidémiologiques utilisés par les réseaux de santé. Dans de nombreux domaines et de plus en plus, on entend dire « donnez-nous plus de données et gardez vos modèles de prédictions d'où qu'ils viennent, de l'A.S.I., de la régression ou de l'économétrie » ;
- l'ordre de grandeur suffit. En effet, quand vous avez 10 clients qui ont acheté un produit A parmi lesquels 9 ont acheté un produit B, les nombres sont importants. Mais quand ce nombre dépasse plusieurs milliers, voire des millions, comme c'est le cas sur certains sites de commerce électronique, ce qui compte ce n'est pas le nombre mais « l'ordre de grandeur », « le rapport », car le nombre est déjà obsolète au moment même où il s'affiche ;
- le *Big Data* suggère que tout est dépendance à des degrés divers, désignée dans l'usage courant par corrélation. À cet égard, l'A.S.I. est au cœur de la question. Mais ses promoteurs sont-ils encouragés à perfectionner les modèles ou invités à collecter de plus en plus de données et à seulement appliquer le plus simple des paradigmes de leur approche. Autrement dit, l'A.S.I., mais également d'autres méthodes d'analyse de données, vont-elles se réduire à de simples calculs de corrélations produites via le paradigme *Map-Reduce* ? »