

« Raw Data » is an Oxymoron Lisa Gitelman (dir.), Cambridge, MIT Press, 2013.

Jean-Christophe Plantin

DANS **COMMUNICATION & LANGAGES** 2013/3 N° 177 , PAGES 155 À 156
ÉDITIONS **NECPLUS**

ISSN 0336-1500

DOI 10.4074/S0336150013013100

Date de mise en ligne : 01/11/2017

Article disponible en ligne à l'adresse

<https://shs.cairn.info/revue-communication-et-langages1-2013-3-page-155?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour NecPlus.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

« RAW DATA » IS AN OXYMORON

GITELMAN Lisa (dir.), Cambridge, MIT Press, 2013.

Le livre édité par Lisa Gitelman vise à tordre le coup à un oxymore trop présent à l'heure des *big data* et *open data* : il existerait des données « brutes » qui n'attendraient que l'action du chercheur, de l'informaticien ou du designer graphique pour livrer tout leur sens. Les auteurs de cet ouvrage s'appuient sur des recherches issues de l'histoire des sciences ou des STS (sciences, technologie et société) pour démontrer que la donnée n'est jamais – malgré son étymologie – « donnée » : elle résulte au contraire d'une construction qui doit être mise en lumière au travers des contextes d'émission, des acteurs intervenant dans ce processus et des usages prescrits aux données. L'intérêt principal de cet ouvrage provient du décentrement qu'il offre par rapport aux enjeux actuels concernant les données numériques, à travers l'histoire de la production, de l'utilisation et de la circulation des données au sein de plusieurs secteurs allant de la science à la citoyenneté : mathématiques, économie, astronomie, sciences citoyennes ou encore *tactical media*.

Daniel Rosenberg nous apprend ainsi que l'usage du terme « donnée » ne date pas des révolutions statistiques et bureaucratiques du XIX^e siècle, ni de l'informatique du XX^e siècle : il est présent dès le XVIII^e siècle, notamment à travers les travaux de Joseph Priestley. Le codage des occurrences du terme dans les corpus de livres de Google Books et d'ECCO (Eighteenth-Century Collections Online) met également en avant le tournant qu'a pris le terme durant ce siècle : alors qu'il désignait auparavant un paramètre axiomatique de l'expérience scientifique, il va peu à peu désigner le résultat d'une expérience ou d'une collection. Rosenberg nous affirme ainsi que le terme « donnée » est intrinsèquement rhétorique : il sert de base à l'argumentation et doit être relié au contexte et aux différentes stratégies argumentatives auxquels il sert de base.

Quel meilleur moyen de combattre le mythe de la donnée brute que de mettre en avant les multiples étapes de sa création ? Si les

informaticiens actuels se plaignent souvent du douloureux travail de « nettoyage » des données au cours duquel celles-ci sont structurées et rendues promptes aux réutilisations, celui-ci n'a rien de nouveau : Kevin R. Brine et Mary Poovey nous décrivent l'étude d'Irving Fisher à la fin du XIX^e siècle sur l'usage du terme « capital » dans les livres d'économie de son époque, qui lui a demandé d'« enlever les éléments incorrects ou inopportuns des données disponibles, de combler les données manquantes, et de les formater de manière à ce qu'elles s'ajustent avec les autres données » (p. 70). De même, Ellen Gruber Garvey nous présente une autre aventure de *data mining* en version *Steampunk* des années 1830 au cours de laquelle les militants abolitionnistes américains Angelina Grimké Weld, Theodore Weld et Sarah Grimké ont trié, catégorisé et annoté des publicités d'esclavagistes provenant de vingt mille journaux, afin d'en faire un outil de combat contre l'esclavage.

Les différentes contributions au livre de Lisa Gitelman soulignent également les difficultés relatives au traitement temporel des données, entre passé et futur. Dans le cadre des recherches sur l'« accélération séculaire », Matthew Stanley décrit les difficultés pour reconstituer le positionnement de la lune lors d'éclipses passées à partir de la consultation d'archives : les astronomes n'utilisaient guère le terme « éclipse » pour rendre compte de cet événement, mais davantage des descriptions vagues telles que « le jour s'est transformé en nuit » (p. 85). Ces multiples approximations appellent alors tout un travail de reconstruction *a posteriori* des données. À l'inverse, David Ribes et Steven J. Jackson se demandent comment garantir la comparabilité à long terme des données. À partir de l'analyse ethnographique d'une initiative de sciences citoyennes qui dure depuis seize ans, les auteurs décrivent les différents « trucs » utilisés pour pallier les constants changements dans les conditions de collecte des données, aussi bien au niveau des personnes les réalisant, de l'environnement mesuré que des infrastructures pour stocker ces données.

Au terme de la lecture de cet ouvrage, on peut se demander si le terme de « donnée » mérite encore d'être utilisé ou s'il n'apporte pas davantage de confusion. Dans un autre cadre, Johanna

Drucker a proposé de remplacer le terme *data* par *capta*, justement pour souligner l'effort de captation et la construction inhérente à la donnée¹. On pourra toutefois retenir la remarque de Daniel Rosenberg sur l'utilité du terme : alors que les notions de fait, de vérité et d'évidence ont déjà été amplement déconstruites – la récente traduction en français de l'ouvrage de Lorraine Daston et Peter Galison devrait accentuer les remises en question du terme d'objectivité² – il peut s'avérer toutefois nécessaire de conserver « un ou deux irréductibles » (p. 18) : la « donnée » constitue alors le « terme frontière » minimal pour transcender les disciplines et les secteurs d'utilisation.

Si le terme est à conserver, il reste toutefois nécessaire de faire ressortir ses étapes de construction lors de son utilisation, afin d'éviter toute naturalisation. Le débat se pose actuellement de manière vive à travers les multiples infographies et visualisations d'information qui se multiplient en ligne, par exemple dans les secteurs du journalisme de données ou de la recherche en SHS.³ Accompagner une visualisation des données utilisées est une première étape nécessaire, mais insuffisante ; elle peut en outre accompagner une rhétorique de la transparence. On citera sur ces deux points le *Data blog* du *Guardian*⁴, où les journalistes publient systématiquement les données utilisées, ce qui n'apporte toutefois pas les informations nécessaires pour comprendre toutes les étapes de transformations entre la donnée et sa mise en forme, en plus de soutenir une rhétorique de l'*empowerment* par les données.

Devant ce projet de mise en avant des conditions de création des données, une difficulté demeure. Si les articles regroupés par Lisa Gitelman possèdent une telle finesse dans l'analyse des

données, c'est en partie parce qu'ils bénéficient du temps long de la recherche et de la place allouée pour développer une pensée. Comment faire tenir ensemble l'utilisation des données et la présentation de ses étapes de construction dans des secteurs régis par une temporalité moins généreuse, tels que le journalisme ou le design ? L'exercice peut s'avérer périlleux, notamment du fait que l'exploitation visuelle des données se base sur les propriétés synoptiques des formes choisies, où les informations seraient accessibles immédiatement, ce qui laisse peu de place aux développements paratextuels et autres notes de bas de page. Sans revenir à l'index de soixante-douze pages qui a accompagné un des premiers usages des diagrammes – *A chart of biography* de Joseph Priestley (1756) – il est nécessaire d'imaginer de nouvelles pratiques de publication ne se contentant pas de mettre en ligne les données utilisées, mais explicitant également la manière dont celles-ci sont été construites. S'ils ne portent pas directement sur ces enjeux et questions, les différents cas d'étude réunis par Lisa Gitelman permettent toutefois de fournir des outils pour y réfléchir.

JEAN-CHRISTOPHE PLANTIN

DISCOURS D'EXPERTS ET D'EXPERTISE

LÉGLISE Isabelle et GARRIC Nathalie, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien, Peter Lang, 2012.

Le savoir d'expert peut se situer entre savoir scientifique et décision politique. Telle est l'une des idées centrales de cet ouvrage collectif, qui rassemble des contributions issues de diverses disciplines (linguistique, analyse de discours, analyse conversationnelle, sociologie, sociologie politique, sciences de l'information et de la communication). Les auteurs analysent tant les spécificités linguistiques et discursives du champ de l'expertise au sein de différents corpus que le statut de l'expert et ses différents rôles, en rapport avec les pouvoirs décisionnaires, la citoyenneté et les médias.

L'ouvrage se divise en trois parties. Dans la première partie, qui interroge l'expertise dans le champ des médias, Pierre Lejeune et Aurélie Tavernier s'intéressent à la construction d'un « ethos » d'expert. La deuxième partie (rassemblant les contributions de Caroline Protais et de Laurence Dumoulin) montre que les activités d'expertise sont largement convoquées dans le domaine judiciaire. Les corpus relevant de l'expertise psychiatrique ou les discours

1. Johanna Drucker, "Humanities Approaches to Graphical Display", *Digital Humanities Quarterly*, 5(1), 2011. Revue en ligne : <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>

2. Lorraine J. Daston, Peter Galison, *Objectivité*, Les presses du réel, 2012.

3. Comme l'a montré la journée d'étude PRATIC du 25 février 2013, « De la "data science" à l'infovisualisation » : <http://www.omnsh.org/spip.php?breve117>. Deux comptes rendus sont disponibles sur le site InternetActu : <http://www.internetactu.net/2013/03/06/de-la-data-science-a-linfovisualisation-12-quest-ce-quun-data-scientist/> et <http://www.internetactu.net/2013/03/12/de-la-data-science-a-linfovisualisation-22-rendre-tangible-la-friction-des-donnees/>

4. <http://www.guardian.co.uk/news/datablog>