



DIGITAL Enlightenment STUDIES

VOLTAIRE FOUNDATION
UNIVERSITY OF OXFORD

Compte rendu

Baird I. (éd.) 2020. *Data Visualization in Enlightenment Literature*. Cham : Springer International Publishing.

Christophe Schuwey, Université Bretagne Sud, christophe.schuwey@univ-ubs.fr

Schuwey C. « Compte rendu : Baird I. (éd.) 2020. *Data Visualization in Enlightenment Literature* »

In : *Digital Enlightenment Studies* 2, 85–88.

DOI : [10.61147/des.25](https://doi.org/10.61147/des.25)

Data Visualization in Enlightenment Literature s'inscrit dans le contexte foisonnant des recherches croisant l'infographie et les sciences humaines.¹ Sous la direction d'Ileana Baird, cette collection d'articles sur le long XVIII^e siècle anglais et français s'attache spécifiquement à souligner les « réévaluations » que produisent l'infographie et les approches *data*-centrées dans l'étude des Lumières. Plusieurs publications récentes en font la démonstration, mais un volume entier qui se positionne sur ce sujet est à saluer. Implicitement, il répond en effet aux réserves régulièrement formulées quant aux apports réels du numérique et de l'infographie à l'étude de l'histoire, de la littérature ou encore de la musique.

L'introduction solide d'Ileana Baird rappelle que le choix de cette période en particulier n'est pas anodin. D'une part, le XVIII^e siècle est un siècle d'infographie (c'est le siècle de William Playfair, pour ne citer que lui) ; de l'autre, les études de la première modernité européenne offrent plusieurs sources et bases de données riches et propices à l'analyse quantitative et qualitative. Celles-ci permettent d'examiner aussi bien des ensembles vastes, dans la première partie de l'ouvrage (« Representing Big Data ») que des objets plus ciblés, dans la seconde partie (« Case Studies »). Baird dresse également un état des lieux centré sur les travaux anglais et américains dans lequel elle inscrit clairement le projet d'ensemble, et fait la part belle aux enjeux critiques et historiques de la *datavisualisation*, en se référant à plusieurs reprises aux réflexions de Johanna Drucker.

Les deux premières contributions du volume s'attachent à l'histoire du livre. Simon Burrows démontre comment les interfaces de la base de données *French Book Trade in Enlightenment Europe, 1769–1794*

¹ DOI : [10.1007/978-3-030-54913-8](https://doi.org/10.1007/978-3-030-54913-8).



renouvellent la compréhension de la librairie et de ses sources. Plutôt que de chercher à lisser les difficultés conceptuelles – telles que les découpages géographiques historiques, les catégories, les classifications (qu'est-ce qu'un ouvrage illégal, par exemple) – l'interface numérique met en lumière toute la complexité de l'archive, sa granularité, et fait apparaître ces manques et ces difficultés conceptuelles. En cela, la démarche est un véritable modèle du genre, qui rappelle une nouvelle fois que l'interface n'est pas une affaire esthétique, mais qu'elle relève pleinement du processus scientifique.

L'article de Mikko Tolonen, Mark J. Hill, Ali Zeeshan Ijaz, Ville Vaara et Leo Lahti présentent les résultats d'un important projet débuté en 2013 sur l'*English Short Title Catalogue* de la British Library. En s'appuyant sur la taxonomie d'Alastair Fowler, le projet construit un nouveau canon littéraire à partir des données éditoriales des ouvrages et leur nombre de rééditions. L'historien du livre appréciera la clarté méthodologique, mais aussi la conscience des limites, l'ampleur des données traitées et la prise en compte de divers facteurs (rééditions à court et long terme) atténuant largement les inévitables imprécisions. Le nouveau canon qu'établit l'étude s'intéresse aussi bien aux formats, qu'aux auteurs ou encore aux éditeurs. L'étude quantitative offre des conclusions qualitatives fortes, en révélant notamment l'importance d'ouvrages rarement repérés ou cités aujourd'hui, mais dont la diffusion massive à l'époque eut un rôle majeur dans la culture des Lumières. Les résultats forcent implicitement les études littéraires à se repositionner, en admettant une fois encore que le canon ne correspond pas à l'environnement culturel du XVIII^e siècle, mais bien à une construction postérieure.

Les deux études suivantes adoptent des approches textuelles et lexicométriques. À partir du corpus de l'*Eighteenth Century Collections Online*, John Regan propose d'analyser les représentations européennes de l'ailleurs à partir des termes associés à l'Asie, l'Afrique et l'Europe au cours du XVIII^e siècle. Les listes de fréquences et les réseaux lexicaux qu'il dégage font apparaître la similitude des vocabulaires utilisés pour l'Amérique et l'Afrique, et la différence avec l'Europe. L'étude fait apparaître une évolution claire du vocabulaire employé tout au long du siècle, et souligne les frontières géographiques complexes associées à l'Europe. On aurait pu souhaiter que l'auteur utilise des listes de mots vides pour exclure les termes géographiques (noms de pays et de régions) associés aux continents. Leur présence importante dans les listes amène ultimement à se demander si le système de valeurs associé à l'Afrique se construit prioritairement autour du terme « Afrique ».

Billy Hall se sert enfin de différentes méthodes quantitatives, dont le *topic modelling*, pour isoler des tendances dans la recherche sur la poésie au XVIII^e siècle. En se basant sur les corpus de deux revues clés (*Eighteenth-Century Studies* et *The Eighteenth Century: Theory and Interpretation*), il examine non seulement les principaux sujets de cette critique, mais aussi les questions de genre et l'évolution de la place des poétesses dans le canon. L'article offre un lot de résultats passionnants, notamment sur le canon poétique critique, et démontre que le quantitatif complexifie des narratifs établis au sujet des évolutions de la critique, et notamment la place que celle-ci accorderait aux poétesses (p.186). Sur le mode du *trial and error*, Hall détaille aussi bien les échecs de son enquête que les méthodes qui produisent des résultats convaincants,



offrant ainsi une précieuse base de connaissance sur différents algorithmes utile à d'autres projets, même sur des sujets éloignés.

La deuxième partie de l'ouvrage, consacrée aux « études de cas », débute par une puissante réflexion historique et théorique sur la fonction des grilles dans l'infographie, réflexion qui résonne directement avec celles de l'introduction et nourrit la réflexion fondamentale sur les modes de représentations et leurs effets sur la compréhension des données. Pour ce faire, Jakub Zdebik prend pour objet la *Machine chronologique* de Jacques Barbeau-Dubourg (1753), le *Tableau économique* de François Quesnay (1758), et les illustrations de Linné dans le *Systema Naturae* (1768), et prolonge ces réflexions jusqu'aux avatars modernes de la forme.

Courtney A. Hoffman propose ensuite une étude pédagogique sur les outils de datavisualisation, appliqués au roman épistolaire *The History of Emily Montague* (1769). À partir d'une question de recherche – visualiser les lettres du roman par date et lieu d'expédition plutôt que par moment de réception, telles que présenté dans le texte – elle démontre successivement comment Voyant, Palladio et Tableau permettent de modéliser sa recherche et ouvre des pistes inattendues.

Ileana Baird renouvelle ensuite l'analyse sociolittéraire de la *Dunciad* d'Alexander Pope (1743). Baird constate d'abord que les commentaires des différentes éditions de l'œuvre, outre leur amoncellement sur l'espace de la page, reflètent souvent les jugements des commentateurs plutôt que d'éclairer les enjeux de ces citations. À cela, le numérique offre des solutions nouvelles. En se fondant sur une analyse de personnages et de colocations, elle reconstruit alors la structure sociale du poème sous forme de réseau. Baird souligne la centralité de John Dennis dans le poème, et met ainsi en lumière la reconnaissance paradoxale que Pope accorde ainsi à son adversaire critique. L'analyse ne manque pas de relever la complexité du dossier éditorial de la *Dunciad*, mais mène néanmoins son enquête sur la dernière édition seulement. Ces multiples éditions, qui, comme dans les satires de Boileau, changent régulièrement de cibles, ne mériteraient-elles de modéliser le réseau en diachronie ? Au vu de la qualité du résultat obtenu sur une édition, c'est une perspective possible qui nous paraît des plus stimulantes, et qui démontrerait d'autant mieux la nécessité d'une approche numérique.

On trouve cette proposition dans l'article de Simon D. Fleming, qui examine les ventes des concertos de Charles Avison, en analysant les professions, genres, localisations géographiques et autres caractéristiques des souscripteurs. En transformant les listes en données traitables en graphiques et tableaux, il ouvre ainsi une fenêtre sur la réalité sociale de la pratique musicale en Angleterre, aussi bien professionnelle qu'amateur. En analysant les évolutions, il offre également une contribution importante à l'étude du marché musical des Lumières, élément indispensable pour comprendre la culture musicale du XVIII^e siècle.

L'ouvrage se clôt sur une conclusion forte d'Emily C. Friedman pour les données ouvertes, qui plaide notamment pour un accès massif, centralisé et standard aux corpus et aux données. Les problèmes qu'elle identifie ne sont ni nouveaux ni spécifiques à l'Angleterre, mais la manière dont elle les expose en rappelle l'urgence de cette question : la dispersion des corpus numérisés, leur difficulté d'accès, sans même parler de ceux qui ne sont pas disponibles librement est un frein majeur aux méthodes quantitatives. Friedman



promeut également l'idée, développée ailleurs, que chaque publication scientifique évaluée par des pairs et basée sur des ensembles de données numériques ou visualisées devrait inclure, en annexe, les corpus et outils utilisés (« Ideally, every peer-reviewed publication that relied on enumerative or visualized datasets would require [...] a publicly available attachment of the corpora and tools used » p.366).

De fait, *Data Visualization in Enlightenment Literature* illustre cette problématique. Le livre est une somme d'expériences, de savoirs et de résultats d'une très grande richesse. Si certaines bases dont il est question sont disponibles (en libre accès ou payant), la mise à disposition des corpus préparés pour les différentes études profiterait assurément à toute la communauté scientifique, afin de tester d'autres hypothèses ou de l'intégrer à d'autres corpus. Le modèle du livre, que les politiques d'évaluation des universités imposent, se révèle passablement inadapté au propos et au projet.

Au rang des remarques, on regrette que les auteurs problématissent peu les formes visuelles qu'ils utilisent. Seules les contributions de Burrows, Zdebek et Hall interrogent véritablement la pertinence des représentations employées. Comme souvent dans les humanités numériques, la question, au fond, demeure celle des données. Les graphiques et réseaux sont utilisés, mobilisés, mais sont peu discutés, comme s'ils émanaient naturellement des données et que leur forme allait de soi. Un ouvrage portant « visualisation » dans son titre n'aurait-il pas mérité une réflexion plus approfondie sur ces modes classiques de représentation ? Dans une telle somme méthodologique, on aurait aussi pu souhaiter un point sur la question des réseaux de neurones et les grands modèles de langage. S'ils ne sont pas pertinents pour tous les sujets, leurs capacités de traitement multiparamétriques font qu'ils excellent dans les analyses complexes de texte, notamment pour le *topic modelling*. Quelles sont les implications de ces technologies pour les différents projets présentés ?²

On peut enfin se demander si le ton de l'ouvrage est le plus adapté au projet qu'il se donne. Conformément aux pratiques du champ, plusieurs articles tiennent à la fois du propos de sciences humaines et du *data paper* : ils sont structurés en fonction des outils et méthodes utilisés, en présentant la disponibilité et les caractéristiques d'un corpus, les instruments d'analyse et la ou les questions exploratoires. Si l'objectif est toutefois de souligner la « réévaluation » des Lumières auquel l'ouvrage conduit, et que l'un de ses publics cibles est celui qu'il faudrait convaincre, alors certaines contributions auraient pu mettre leurs conclusions plus en valeur, et un peu moins l'accent sur le processus. On reconnaît évidemment que cette remarque relève de ces injonctions contradictoires propres à l'interdisciplinarité, qui supposent des normes de rédaction spécifiques à chacun des champs.

Quoi qu'il en soit, les citations en témoignent, la communauté des humanités numériques et de l'infographie a fait son miel de l'ouvrage depuis sa parution, et c'est peut-être l'essentiel. Car ces points de discussions, qui dépassent en partie cet ouvrage en propre, ne doivent en aucun cas occulter l'extrême richesse et la qualité de cet ensemble d'articles qui servira assurément à la communauté scientifique. De l'introduction à la conclusion, il est difficile, après sa lecture, de tenir encore l'infographie et les approches fondées sur les données à distance des sciences humaines.

² On rappellera toutefois que, l'ouvrage étant paru en 2021, la question n'était pas aussi urgente qu'aujourd'hui.