



HAL
open science

Les enjeux épistémologiques des humanités numériques

Alexandre Gefen

► **To cite this version:**

Alexandre Gefen. Les enjeux épistémologiques des humanités numériques. Socio - La nouvelle revue des sciences sociales, 2015, 4, pp.61–74. 10.4000/socio.1296 . hal-01624113

HAL Id: hal-01624113

<https://hal.science/hal-01624113>

Submitted on 26 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Socio

La nouvelle revue des sciences sociales

4 | 2015 :

Le tournant numérique... et après ?

Dossier : Le tournant numérique... Et après?

Les enjeux épistémologiques des humanités numériques

The epistemological issues posed by the digital humanities

ALEXANDRE GEFEN

p. 61-74

Résumés

Français English

En sciences humaines comme en science dures, les numérisations massives des textes et des données produisent depuis plus d'une décennie des big data ou long data ouvrant des pistes de recherche novatrices. Mais les méthodes critiques qui les accompagnent ont des enjeux épistémologiques, institutionnels et pédagogiques considérables. Dans ce qu'on appelle désormais les « humanités numériques », la lecture à distance de corpus constitués par des cartes et graphes offre une forme spécifique de savoir et un paradigme méthodologique et épistémologique qu'il importe de saisir dans toute sa puissante heuristique, sans se laisser entraîner par l'idée naïve d'une production transparente de savoirs par moissonnage des corpus, masse de données qui restent des artefacts muets en l'absence d'une herméneutique spécifique.

In the humanities, as in the “hard” sciences, for more than a decade now the massive digitalisation of texts and data has produced big data or long data opening up new opportunities for research. But the accompanying critical methods pose considerable epistemological, institutional and pedagogical issues. In what is now referred to as the “digital humanities”, the remote reading of a corpus constituted by maps and graphs provides a specific form of knowledge and a methodological and epistemological paradigm the heuristics of which must be grasped in all their potential. One must not allow oneself to be swept into the naïve belief that trawling a corpus of texts results in a transparent production of knowledge, a mass of data which remain silent artefacts in the absence of a specific hermeneutics.

Entrées d'index

Mots-clés : humanités numériques, big data, sciences de la culture, lexicographie, épistémologie des sciences sociales, histoire littéraire, théorie de la littérature, critique, sciences collaboratives

Keywords : digital humanities, big data, sciences of culture, lexicography, epistemology of the social sciences, literary history, theory of literature, criticism, collaborative sciences

Texte intégral

- 1 On assiste aujourd'hui à un tournant informationnel et computationnel du travail scientifique. En sciences dures (sciences du vivant, physique des particules, sciences de l'Univers, chimie quantique, études de population, etc.), de nombreuses recherches produisent désormais des volumes considérables de données hétérogènes (*big data*), nécessitant le développement de nouvelles méthodes d'extraction et de traitement de l'information : la production massive de données par des expériences, des capteurs ou des populations s'offre à une approche empirique des faits où l'on peut extraire les phénomènes par des méthodes algorithmiques à partir des masses de données en limitant l'interférence des présupposés. Parallèlement, en sciences de la culture, l'apport des bibliothèques en ligne et des moteurs de recherche conduit à la redécouverte de vastes corpus anciens et encourage de nombreuses initiatives d'analyse et d'exploitation. Je voudrais montrer, en prenant l'exemple de la critique littéraire, que, dans les sciences sociales, l'extraction de connaissances, le repérage et l'agrégation de faits culturels, la visualisation et la navigation à l'intérieur de corpus textuels plus ou moins massifs sont également des instruments nouveaux de connaissance et de perception. Ces *big data* ou *long data* de la culture, produits par un tournant informationnel du savoir, font de l'historien ou du théoricien de la littérature un *data scientist* comme un autre. Les méthodes critiques qui les accompagnent ont des enjeux épistémologiques, institutionnels et pédagogiques considérables. Dans ce qu'on appelle désormais les « humanités numériques¹ », la lecture à distance (*distant reading*) de corpus constitués par des cartes et graphes offre « une forme spécifique de savoir » (Moretti, 2005 : 1) permettant de découvrir des régularités autant que des nouveautés irrégulières (Goodwin et Holbo, 2011 ; Jockers, 2013). De même, la lecture numérique de corpus plus étroits ou d'œuvres uniques considérées comme des bases de données de faits linguistiques et critiques (*close reading*) propose une « herméneutique intégrative », caractérisée par une désessentialisation du texte, désormais dénudé dans sa texture et placé au centre d'un empire hypertextuel et de systèmes variationnels qui le décentrent (Viprey, 2005 ; Mayaffre, 2007). D'où l'apparition d'un paradigme méthodologique et épistémologique qu'il importe de saisir dans toute sa puissante heuristique, sans se laisser entraîner par l'idée naïve d'une production transparente de savoirs par moissonnage du *big data* des corpus, masse de données qui restent des artefacts muets en l'absence d'une herméneutique spécifique.

La construction des corpus

- 2 « Le diable est dans les détails », dit un vieil adage : pour mieux comprendre les enjeux épistémologiques de ces nouvelles formes critiques, il importe de préciser les opérations et les méthodes concrètes mises en œuvre au cœur des humanités

numériques. L'épistémologie disciplinaire est indissociable des pratiques et des usages concrets, tant techniques que sociologiques : structures des équipes académiques, souvent bouleversées par l'émergence de nouveaux acteurs et l'interdisciplinarité, place des informaticiens, rôle des contributeurs anonymes ou du moins externes lorsque les projets intègrent du *crowdsourcing*², etc. Coproduits par des ingénieurs et des chercheurs dans une logique souvent *bottom-up* (c'est-à-dire partant des besoins concrets et des possibilités techniques), gouvernés par les invitations et les interdictions propres aux outils techniques autant que par des programmes de recherche disciplinaire intellectuellement maîtrisés, profondément affectés par les campagnes patrimoniales de numérisation et par les modalités juridiques d'exploitation des données, les méthodes et les savoirs des humanités numériques sont indissociables des normes, des algorithmes et des corpus qu'ils déploient³.

3 Quelles sont, justement, les opérations que présuppose la « philologie numérique⁴ » lorsqu'elle constitue des corpus ? Il faut rappeler que, pour l'analyse numérique du *long data* de l'histoire culturelle, la première étape consiste en la numérisation ou l'agrégation de textes en corpus, puis en leur saisie, leur conversion et leur équipement en métadonnées (métadonnées bibliographiques ou, plus finement, utiles à leur classement). Ces opérations sont loin d'être transparentes, pour de multiples raisons. En premier lieu, les acteurs des humanités numériques dépendent dans une large mesure des politiques de numérisation de masse, celles des bibliothèques en particulier ; or ces dernières sont elles-mêmes contraintes par l'état physique des documents et par la possibilité d'équiper un nombre considérable de textes en métadonnées pertinentes : le problème est particulièrement sensible pour les périodiques, dont l'exploitation productive suppose un découpage par articles et par auteurs. Deux autres facteurs s'opposent à l'idée d'une analyse strictement empirique : d'une part, il est difficile d'identifier de manière automatique ou manuelle, au sein des textes, des informations d'indexation signifiantes (en particulier ce qu'on appelle les « entités nommées », noms, dates et lieux), alors même que ce balisage constitue un préalable nécessaire à certains types d'études (de réception, par exemple) ; d'autre part, pour les textes anciens, les informations de tirage et de diffusion sont rares, ce qui rend assez peu fiable la mesure de leur représentativité. En somme, dans l'anarchie des corpus numériques des bibliothèques, le rêve d'exhaustivité d'une bibliothèque de Babel est aussi éloigné de sa réalisation que l'avènement espéré par Tim Berners-Lee d'un web sémantique (Berners-Lee, Hendler et Lassila, 2001) à l'intérieur d'Internet. Ainsi, souvent, le travail sur le *big data* doit se faire par bricolage, à partir de corpus partiellement représentatifs et partiellement organisés, imposant des biais considérables à toute analyse quantitative et requérant des chercheurs une conscience méthodologique aguerrie. L'interrogation, très impressionnante, du Google Ngram Viewer⁵ ne saurait dispenser « de réfléchir aux questions d'échantillonnage, de représentativité, d'homogénéité du corpus, des conditions dans lesquelles les données sont produites, agrégées ou combinées », comme le rappelle Michel Wieviorka (2013 : 26), en particulier lorsqu'il s'agit de fouiller des données ouvertes, mais dont les standards de constitution ne sont pas clairement connus – ce qui est précisément le cas de Google Books sur lequel s'appuient les expériences de « culturonomique » (quantification des tendances culturelles) les plus simples et les plus spectaculaires.

4 Par ailleurs, ni la conversion numérique des textes, ni leur modélisation structurelle par des normes⁶ visant à les abstraire selon des règles générales en séparant contenu textuel originel, présentation originelle (découpage textuel, typographie, mise en page, etc.) et présentation finale sur ordinateur ou tablette, ne sont transparentes : convertir des récits en des bases de données, passer de la narration ou de la diction à de l'information, c'est identifier des noms et des concepts signifiants, transposer des

typographies et des mises en pages, annoter un texte « à l'ancienne » ou avec des outils collaboratifs innovants⁷, démembrer une structure en repérant l'énonciation éditoriale et les enjeux esthétiques des supports. Il s'agit de déterminer ce qui relève du texte ou du paratexte, d'identifier des structures, de choisir les normes orthographiques, en faisant à chaque fois le distinguo entre information signifiante et information négligeable : on choisira, par exemple, pour un texte classique de conserver ou non les « privilèges d'impression », les formes de ponctuation et de typographie d'époque, les informations spatiales de mise en forme des pages, les données concernant la matérialité du texte ; on considèrera ou non les « tomes », « livres » ou « parties » comme relevant d'un même fichier, etc. S'il existe des outils de détection automatique des structures et de rapprochement des termes sémantiques par association à des bases de données externes, et s'il est envisageable de s'appuyer par exemple sur des cartes de mots-clés signifiants, générés par des algorithmes, toutes ces opérations relèvent d'une lourde ingénierie des connaissances. Elles présupposent des modélisations plus ou moins normatives. Les processus de sélection et de structuration des entités des textes impliquent donc des sémantiques – si ce n'est des ontologies –, qu'elles soient explicites et gérées par outils des arbres *ad hoc*, ou implicites et apportées par la tradition.

- 5 Ainsi, la standardisation et la modélisation égalisatrice par les normes ouvertes, « démocratiques » et générales du web et la délégation de certaines opérations de sémantisation à des algorithmes de classement empiriques et transversaux ne sauraient faire oublier à quel point les sémantiques restent indissociables de savoirs disciplinaires et d'une tradition interprétative sans laquelle ni l'établissement ni l'interprétation des œuvres ne font sens. Il n'y a donc pas une œuvre numérisée mais des standards et des degrés d'encodage et de sémantisation : entre un texte numérisé en « texte pur », comme on en trouve encore sur le site du projet Gutenberg (la première des bibliothèques numériques proposant en libre accès des ouvrages libres de droits), et un texte sur lequel le moindre détail aura été encodé, de la couleur de l'encre à la signification d'une référence permettant de lier un texte à un autre, c'est par un facteur cinquante qu'auront été multipliés le temps de travail, la taille du fichier et la densité informationnelle exploitable (rendant au passage le fichier final XML quasi illisible avant sa transformation pour affichage en HTML). Le degré de détail pouvant être atteint par une transcription numérique est extrêmement variable et la transcription dans la norme TEI génère sa propre « TEIlogie⁸ », ses règles et ses débats, ses normes et ses bons usages, extension numérique de la philologie classique et des problématiques des éditions papier, en circonvenant des impossibilités propres à ces dernières (pensons, par exemple, à la possibilité d'encoder deux transcriptions ou deux interprétations d'un même passage) grâce à la maîtrise d'un langage d'encodage, à un travail philologique et herméneutique lourd, et au déploiement d'outils logiciels de transcription et d'affichage complexes. Car les modalités d'affichage d'un texte sont aussi riches que les possibilités logicielles des interfaces et évoluent avec elles, permettant des formes de googlation et de reconstruction du texte modulable selon le projet et les besoins du lecteur : dans une édition du *Grand Cyrus* de Madeleine de Scudéry⁹, nous avons ainsi proposé un texte « à la carte », que le lecteur peut afficher soit dans sa typographie et sa mise en page ancienne, soit dans une version modernisée, soit encore dans une version « philologique » mettant en regard la reproduction de la page originelle et sa transcription ; une œuvre dont la lecture peut passer par une structure de résumé dépliant et une aide sous la forme de bulles documentaires, aussi bien que par une présentation continue sans paragraphes, respectant la densité et la compacité des in-folio d'époque : s'il existe des heuristiques plus ou moins fécondes de lecture numérique du texte, et si la dénudation des textes dans un format standardisé

permet des opérations de comparaison et d'exploitation infinies et non prédéterminées, aucune ne possède un caractère universel et obvie.

- 6 L'« herméneutique intégrative » permise par le numérique ne considère donc le texte littéraire comme un objet informationnel ordinaire et empirique analysable par une machine qu'au prix d'une sélection *a priori* délicate (possibilité de l'agrégation dans un corpus d'une version particulière) et d'une décomposition structurelle et sémantique très complexe. Les méthodes permettant de faire de l'œuvre littéraire un ensemble de faits textuels ordinaires, examinables empiriquement dans leur connexion à d'autres faits textuels ou références, et susceptibles de comparaisons, supposent donc le recours à un savoir philologique et historique ancien avant toute expérimentation, et exposent donc tout savoir empirique aux aléas des conditions de sa constitution.

Les enjeux de l'analyse quantitative

- 7 Loin de se satisfaire de concordanciers, ou tables lexicales de concordances, les opérations de fouilles textuelles approfondies, qui permettent par exemple de dégager des thématiques, supposent ce qu'on appelle la « lemmatisation », c'est-à-dire le regroupement des différentes formes d'un mot, et ne font qu'ouvrir le travail interprétatif : comment par exemple interpréter des statistiques sur la « colocation » fréquente ou au contraire très rare de deux mots, concepts ou noms propres, et développer une intelligence de linguistique de corpus, sans recourir à des expériences configurées par une intuition et un savoir disciplinaire ? Certes, un texte individuel peut ainsi être visualisé comme un nuage de mots, ou, après traitement par lemmatisation et neutralisation des mots grammaticaux, comme un nuage de concepts ; des comparaisons peuvent être faites par des recours à des formes stylistiques ou grammaticales, et un certain nombre de « découvertes » inattendues peuvent être espérées de telles visualisations. Pourtant, la lexicométrie et l'analyse sémantique et thématique ne peuvent répondre qu'à des questions posées selon des problématiques critiques nécessairement exogènes, car, avant même toute interprétation, les modalités de visualisation permettant de rendre interprétables des résultats supposent des choix mathématiques pour produire des graphes simplement affichables et lisibles.
- 8 Le passage au *big data* permet-il de faire un vrai saut épistémologique dans lequel la part des présuppositions et des *a priori* laisserait vraiment place à des résultats empiriques ? Les expériences faites par Franco Moretti proposent de construire « une forme spécifique de savoir », où, par exemple, « le genre devient un spectre à bande large, dont aucun texte singulier ne pourra jamais représenter la multiplicité interne » dans le cadre d'une « conception matérialiste de la forme » (Moretti, 2005 : 91) proposant des objectivations scientifiques par l'établissement de « cartes », de « graphes », d'« arbres » opérant à distance des textes. Mais les résultats mathématiques ne sont rien sans des représentations et la production de graphes et de formes. D'où un premier saut, consistant à passer à des résultats visualisables : « la quantification pose le problème, la forme apporte la solution », écrit Moretti (2005 : 58) pour expliquer le passage à des représentations unidimensionnelles déployant, en regard de la chronologie, les possibles formels des œuvres littéraires du XIX^e siècle qui constituent son corpus et son objet. Par ailleurs, loin d'être neutres, les résultats de cette tentative de retrouver une sorte d'universel du roman sont nourris de sous-entendus interprétatifs liés à une conception héritée de Lukács – historiciste, si ce n'est évolutionniste – de la « vie » des formes et des productions culturelles (Escola, 2008). Ce passage du descriptif au narratif est donc un retour à un grand récit interprétatif,

dans lequel les écarts constatés (apparition de tel ou tel sous-genre) et les fluctuations sont moins considérés comme des possibles non réalisés que comme des indices de surdétermination historique inéluctable, dans laquelle les chefs-d'œuvre reconnus restent des points d'inflexion et de référence fondamentale.

9 Le geste épistémologique le plus fort des travaux de Moretti est d'avoir tenté d'objectiver, pour les abstraire, les différentes variables définissant les œuvres littéraires (genres, formes, etc.) afin de permettre des changements d'échelle d'analyse. L'accroissement numérique des corpus et surtout cet « aplatissement » des textes littéraires dans l'espace numérique selon des normes propres aux besoins généraux de cohérence et d'indexation du web et selon des sémantiques globales offrent des perspectives particulièrement intéressantes : ils permettent d'interroger des corpus qui ne hiérarchisent pas *a priori* les entités informationnelles, en sorte que peuvent émerger par un rapport concret au « terrain » des entités « authentiquement empiriques, à jamais éloignées de la simple illustration de concepts pré-donnés » (Monnin, 2012¹⁰). Ainsi les modes de sémantisation généraux déployés pour les textes littéraires, comme les corpus élargis par rapport aux définitions restrictives du champ littéraire et critique et qui contiennent autant des grandes œuvres canoniques peuvent proposer des rapprochements originaux et conduire à des constats à mille lieues des pré-supposés de départ.

10 La recherche plein texte permet déjà des recontextualisations lexicales ou conceptuelles parfois inattendues, et les opérations de rapprochement textuel automatisé font émerger des accointances lexicales ou des réécritures (voire des reprises non assumées) entre des passages parfois de prime abord éloignés, sans connaissance au départ des sources et des modes de transmission intertextuelle, nous exposant à bien des surprises. Par-delà le « formalisme quantitatif » développé à Stanford par Franco Moretti (analyse de mots grammaticaux et de traits formels/énonciatifs), d'autres méthodes de *data mining* (fouille de données) offrent des perspectives intéressantes : sans que le point de départ soit un texte particulier, il est ainsi possible de cartographier les affinités entre différents textes selon une requête particulière ; il est également envisageable d'effectuer des classifications automatiques de documents selon des méthodes de classification bayésienne ou selon celle des « k plus proches voisins », algorithmes permettant de déterminer empiriquement des communautés de vocabulaire. Au contraire, il est possible de proposer des analyses contrastives entre des corpus prédéfinis. Des expériences de classification thématique ou stylistique et surtout d'interrogation des classifications admises extrêmement variées peuvent alors être tentées, et des modes de représentation originaux proposés. Qu'il s'agisse de suivre sur la longue durée l'évolution d'un concept (en comparant le lexique des « voisins distributionnels » autour des termes et noms propres pivots), d'une forme ou d'un champ à l'aide de diagrammes¹¹, d'entraîner une machine à découvrir des œuvres cachées dans une production éditoriale qui n'a pas été étiquetée comme littéraire, de s'interroger sur les conditions de réception d'une œuvre en comparant des catalogues et des canons, ou de se laisser surprendre en essayant de regarder les corrélations possibles entre métadonnées et formes textuelles, on peut produire toute une série d'expériences d'imagerie littéraire. Les possibilités expérimentales permises par l'agrégation de données hétérogènes sont infiniment variées et vont des cartes géolocalisant des métadonnées géographiques, des cartes des réseaux relationnels (entre personnages, auteurs, etc.) à la fouille des réseaux sociaux permettant d'interroger les modalités de construction de l'auctorialité (c'est-à-dire l'autorité juridique et symbolique de l'écrivain) et de la valeur, la sociologie (vérifier par exemple les hypothèses de la sociologie bourdieusienne des champs) aussi bien que l'histoire des poétiques (en produisant une histoire objectivée de l'évolution des styles)

ou la théorie littéraire (en la confrontant aux formulations variables de ses concepts et de l'idée même de littérature).

Conclusion

- 11 L'émergence des humanités numériques va de pair avec une perte d'insularité disciplinaire et géographique, et l'accès permis par les bases de données numériques au contexte culturel et aux détails d'une aventure littéraire offre le projet d'une lecture active de l'œuvre et d'une forme novatrice d'appropriation de la critique, foncièrement interdisciplinaire et désormais ouverte au débat et à l'intelligence collective. Les humanités numériques sont favorisées et favorisent en retour l'émergence d'une épistémologie empirique et décentrée, indissociable d'une approche pragmatique, interdisciplinaire, décloisonnée des faits culturels. Celle-ci ambitionne de se couper de catégories préconstruites et des schémas préfabriqués au profit d'une ontologie centrée sur leurs objets et d'une heuristique computationnelle et comme autonomisée (Ramsay, 2011). Ce tournant épistémologique conduit à la production de logiques moins essentialistes, ou critiques, que descriptives, transactionnelles et expérimentales. Cette épistémologie consonne avec d'autres courants majeurs contemporains : la sociologie contemporaine de Latour ou de Goffmann centrée sur les usages, qui reconnaît l'autonomie des acteurs et des objets dans une « ontologie plate » ; les philosophies wittgensteiniennes des pratiques ordinaires comme la « philosophie expérimentale » et les modèles matérialistes produits par les sciences cognitives (Harman, 2009 ; Smith, 2011, Halpin et Monnin, 2012). Ce qu'Eileen Joy nomme les « object-oriented reading modes », propres au tour informationnel et computationnel des humanités numériques (Joy, 2011), s'accompagne ainsi d'une interdisciplinarité nouvelle, notamment avec les *data sciences* (sciences des données), comme d'un tournant descriptif, désessentialisant et interrogeant les frontières mêmes de la littérature, dont les concepts sont censés être décrochés des hiérarchies héritées et des pièges des méthodes hypothético-déductives. Plus encore que la sociologie ou l'anthropologie culturelle, les humanités numériques font du texte littéraire, pour le meilleur ou pour le pire, un discours parmi d'autres, connecté à d'autres, et dont les nœuds relèvent non d'un système de valeurs mais d'une « ontologie plate ».
- 12 Loin d'être mécanique et purement quantitative, la deuxième génération des travaux des humanités numériques est « qualitative, interprétative, émotive, générative », comme l'écrivent les auteurs du *Digital Humanities Manifesto 2.0* (Schnapp et Presner, 2009 ; voir Berry, 2011). L'analyse des discours et de leurs situations d'énonciation concrètes permettrait, par exemple, de circonvenir les téléologies du « littéraire », si périlleuses lorsque l'on fait, en l'anthropomorphisant, une histoire de la littérature et de son autonomisation supposée au XIX^e siècle : d'autres canons, d'autres regroupements, d'autres logiques seraient discernables si l'on disposait de corpus suffisamment larges et bien édités pour voir se transformer les genres et les valeurs littéraires dans l'espace large des savoirs et des représentations. Émergeraient alors, par les expériences de fouille des textes, des cohérences ou des signaux étranges permettant de brasser à nouveaux frais des structures consacrées, de faire apparaître des pans oubliés de la littérature, de signaler des cas atypiques posant à l'interprétation de nouvelles questions... Nous voilà loin du risque d'un nouveau scientisme où la littérature – ses noms, ses champs, ses valeurs –, à l'état de données, serait réduite à une curieuse activité humaine statistique qui se mesure sans se lire, se dessine sans s'interpréter, se dévoile sans avoir besoin de se laisser raconter. Ce travail a bénéficié

d'une aide d'État gérée par l'Agence nationale de la recherche dans le cadre des investissements d'avenir portant la référence ANR-11-IDEX-0004-02.

Bibliographie

Références bibliographiques

BARTSCHERER, Thomas et COOVER, Roderick (dir.), 2011, *Switching Codes: Thinking Through Digital Technology in the Humanities and the Arts*, Chicago, University of Chicago Press.

BASSEZ, Danielle, 2006, *Écrits dans les marges : de la pratique du gribouillage comme art gourmand de la lecture*, Le Chambon-sur-Lignon, Cheyne.

BERNERS-LEE, Tim, HENDLER, James et LASSILA, Ora, 2001, « The semantic Web », *Scientific American* (en ligne), mis en ligne le 17 mai 2001. Consultable en ligne : <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>>.

BERRY, David M., 2011, « The computational turn: Thinking about the digital humanities », *Culture Machine*, vol. 12 (en ligne). Consultable en ligne : <<http://www.culturemachine.net/index.php/cm/article/viewArticle/440>>.

BERRY, David M. (dir.), 2012, *Understanding Digital Humanities*, Handmills (Basingstoke, Hampshire), New York, Palgrave Macmillan.

CHATELAIN, Jean-Marc (dir.), 1999, *Le Livre annoté : de l'Antiquité à Stendhal*, *Revue de la Bibliothèque nationale de France*, n° 2, Paris, BnF.

DOUEIHI, Milad, 2011, *Pour un humanisme numérique*, Paris, Éditions du Seuil.

ESCOLA, Marc, 2008, « Voir de loin. Extension du domaine de l'histoire littéraire », *Acta fabula*, vol. 9, n° 6, Essais critiques, juin. Consultable en ligne : <<http://www.fabula.org/revue/document4291.php>>.

GARCIA, Tristan, 2011, *Forme et objet : un traité des choses*, Paris, Presses universitaires de France.

GOLD, Matthew K. (dir.), 2012, *Debates in the Digital Humanities*, Minneapolis, University of Minnesota Press.

GOODWIN, Jonathan et HOLBO, John (dir.), 2011, *Reading Graphs, Maps, and Trees: Responses to Franco Moretti*, Anderson, Parlor Press. Consultable en ligne : <<http://www.parlorpress.com/pdf/ReadingMapsGraphsTrees.pdf>>.

GUICHARD, Éric, 2013, « L'internet et les épistémologies des sciences humaines et sociales », Équipe Réseaux, Savoirs et Territoires, ENS (en ligne), février. Consultable en ligne : <<http://barthes.ens.fr/articles/Guichard-RSL.html>>.

HALPIN, Harry et MONNIN, Alexandre (dir.), 2012, « Philoweb: Toward a philosophy of the Web », numéro spécial de *Metaphilosophy*, vol. 43, n° 4, juillet, p. 361-532.

HARMAN, Graham, 2009, *Prince of Networks: Bruno Latour and Metaphysics*, Melbourne, Re.press.

JACKSON, H. J., 2001, *Marginalia: Readers Writing in Books*, New Haven et Londres, Yale University Press.

JAHAH, Marc, 2012a, « Les 4 royaumes de l'annotation (1500-2012) », *Marginalia – Savoir(s) d'annotateurs*, mis en ligne le 13 avril. Consultable en ligne : <<http://marginalia.hypotheses.org/18186>>.

JAHAH, Marc, 2012b, « Bienfaits et méfaits des annotations partagées », *Marginalia – Savoir(s) d'annotateurs*, mis en ligne le 10 mai. Consultable en ligne : <<http://marginalia.hypotheses.org/18467>>.

JAHAH, Marc, 2013a, « Annotated books online : découvrir les annotations des premiers siècles de l'imprimerie », *Marginalia – Savoir(s) d'annotateurs*, mis en ligne le 31 janvier. Consultable en ligne : <<http://marginalia.hypotheses.org/page/2>>.

JAHAH, Marc, 2013b, « La lumière de l'annotation : pour une esthétique de la pénombre dans les

espaces numériques », *Marginalia – Savoir(s) d'annotateurs*, mis en ligne le 7 mai. Consultable en ligne : <<http://marginalia.hypotheses.org/23140>>.

JAHAH, Marc, 2013c, « Rhétorique des marginalia imprimées », *Marginalia – Savoir(s) d'annotateurs*, mis en ligne le 2 juillet. Consultable en ligne : <<http://marginalia.hypotheses.org/23646>>.

JOCKERS, Matthew L., 2013, *Macroanalysis: Digital Methods and Literary History*, Urbana, University of Illinois Press.

JOY, Eileen, 2011, « Swedish Twitter university lecture: Eileen (again) on object-oriented reading modes ». Consultable en ligne : <<http://punctumbooks.com/blog/swedish-twitter-university-lecture-eileen-again-on-object-oriented-reading-modes/>>.

LANDOW, George P., 2006, *Hypertext 3.0: Critical Theory and New Media in an Era of Globalization*, Baltimore, Johns Hopkins University Press.

MAYAFFRE, Damon, 2002, « L'Herméneutique numérique », *L'Astrolabe. Recherche littéraire et Informatique*, numéro spécial, 2002, p. 1-11. Consultable en ligne : <<http://halshs.archives-ouvertes.fr/hal-00586512/>>.

MAYAFFRE, Damon, 2007, « Philologie et/ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques ? », in Michel Ballabriga et François Rastier (dir.), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation*, actes du XXVII^e colloque d'Albi, « Langages et signification », publiés par Carine Duteil-Mougel et Baptiste Foulquié, Toulouse, Presses universitaires du Mirail. Consultable en ligne : <<http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Mayaffre.pdf>>.

MAYER-SCHÖNBERGER, Viktor, 2011, *Delete: The Virtue of Forgetting in the Digital Age*, Princeton, Princeton University Press.

MCNAMEE, Kathleen, 1992, *Sigla and Select Marginalia in Greek Literari Papyri*, Bruxelles, Fondation égyptologique Reine Élisabeth.

MONNIN, Alexandre, 2012, « L'ingénierie philosophique comme design ontologique : retour sur l'émergence de la "ressource" », *Réel-virtuel : enjeux du numérique*, n° 3, « Archéologie des nouvelles technologies », avril. Consultable en ligne : <<http://reelvirtuel.univ-paris1.fr/index.php?/revue-en-ligne/3-monnin/>>.

MORETTI, Franco, 2005, *Graphs, Maps, Trees: Abstract Models for a Literary History*, Londres et New York, Verso.

MOROZOV, Evgeny, 2011, *The Net Delusion: The Dark Side of Internet Freedom*, New York, PublicAffairs.

MOUNIER, Pierre (dir.), 2012, *Read/write book 2 : une introduction aux humanités numériques*, Marseille, OpenEdition Press. Consultable en ligne : <<http://books.openedition.org/oep/226>>.

RAMSAY, Stephen, 2011, *Reading Machines: Toward an Algorithmic Criticism*, Urbana, University of Illinois Press.

RASTIER, François, 2001, *Arts et sciences du texte*, Paris, Presses universitaires de France.

ROMARY, Laurent et HUDRISIER, Henri, sans date, « Ingénierie linguistique – les ressources linguistiques : TEI – Text encoding initiative ». Consultable en ligne : <<http://www.culture.gouv.fr/culture/dglf/rifal/tei.htm>>.

SCHNAPP, Jeffrey et PRESNER, Todd, 2009, *The Digital Humanities Manifesto 2.0*. Consultable en ligne : <http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf>.

SHERMAN, William H., 2009, *Used Books: Marking Readers in Renaissance England*, Philadelphia, University of Pennsylvania Press.

SIMONE, Raffaele, 2012, *Pris dans la toile : l'esprit aux temps du web*, traduit de l'italien par Gérard Larché, Paris, Gallimard.

SLIGHTS, William W. E., 2001, *Managing Readers: Printed Marginalia in English Renaissance Books*, Ann Arbor, University of Michigan Press.

SMITH, Brian Cantwell, 2011, *On the Origin of Objects*, Cambridge et Londres, The MIT Press.

VIPREY, Jean-Marie, 2005, « Philologie numérique et herméneutique intégrative », in Jean-Michel Adam et Ute Heidmann (dir.), *Sciences du texte et analyse de discours : enjeux d'une interdisciplinarité*, Genève, Slatkine érudition, p. 51-68.

WIEVIORKA, Michel, 2013, *L'Impératif numérique, ou La nouvelle ère des sciences humaines et*

sociales ?, Paris, CNRS éditions.

WOLFE, Joanna et NEUWIRTH, Christine M., 2001, « From the margins to the center: The future of annotation », *Journal of Business and Technical Communication*, vol. 15, n° 3, p. 333-371.

Notes

1 Sur la définition du champ, voir Landow (2006), Bartscherer et Coover (2011), Mounier (2012), Gold (2012), Berry (2012). Sur les débats récents, voir Doueïhi (2011), Mayer-Schönberger (2011), Morozov (2011), Simone (2012).

2 On nomme *crowdsourcing* au sens large les formes de participation ouvertes par l'Internet 2.0, centré sur l'interaction et la contribution des internautes (forums de discussion, science citoyenne, démocratie participative, écriture collaborative, financement collectif, journalisme amateur, etc.).

3 Je ne peux ici que souscrire aux affirmations d'Éric Guichard pour qui « le couple {système de signes + support}, que les qualitatistes décrivent comme technique et sans intérêt, conditionne partiellement l'exercice de la pensée, et montre les limites des représentations qui la prétendent spirituelle et immatérielle : l'obligation de compter et de trier, la combinatoire sont parfois vecteurs ou les aiguillons de la conceptualisation » (Guichard, 2013).

4 Sur ce terme, voir Rastier (2001 : 73-98), Mayaffre (2002).

5 <<https://books.google.com/ngrams>>, application qui permet la comparaison sous forme de diagramme du nombre d'occurrences d'un nom, d'une phrase dans un corpus donné et une période définie.

6 Notamment la Text Encoding Initiative (TEI), standard de fait en numérisation patrimoniale littéraire utilisant la norme XML. Pour une présentation, voir Romary et Hudrisier (sans date).

7 Sur la problématique de l'annotation, voir Bassez (2006), Chatelain (1999), ainsi que les travaux classiques et importants signalés par Jajah (2012a, 2012b, 2013a, 2013b, 2013c), McNamee (1992), Jackson (2001), Slights (2001), Wolfe et Neuwirth (2001), Sherman (2009).

8 Voir les débats sur la liste TEI-L : <<http://listserv.brown.edu/archives/cgi-bin/wa?Ao=TEI-L>>.

9 <<http://www.artamene.org>>.

10 Alexandre Monnin s'appuie sur Tristan Garcia (2011) du côté de la métaphysique et sur Bruno Latour du côté de la sociologie.

11 Voir le travail de Mark Algee-Hewitt, « Mapping the sublime: A computer-assisted reading of eighteenth-century aesthetics », engagé à partir de sa thèse (Algee-Hewitt, 2008) : l'usage de diagrammes de Voronoï a permis la découverte des multiples groupes sémantiques qui ont établi des distinctions plus claires entre les sublimes « théologique », « naturel » et les versions « esthétiques » du sublime.

Pour citer cet article

Référence papier

Alexandre Gefen, « Les enjeux épistémologiques des humanités numériques », *Socio*, 4 | 2015, 61-74.

Référence électronique

Alexandre Gefen, « Les enjeux épistémologiques des humanités numériques », *Socio* [En ligne], 4 | 2015, mis en ligne le 28 mai 2015, consulté le 12 juin 2015. URL : <http://socio.revues.org/1296>

Auteur

Alexandre Gefen

Alexandre Gefen est chargé de recherche au Centre d'étude de la langue et des littératures

françaises (Centre national de la recherche scientifique [CNRS] – université Paris 4). Il travaille sur des questions de théorie littéraire appliquées en particulier à la littérature française contemporaine. Fondateur du site Fabula.org et membre du Labex OBVIL (Sorbonne université), il s'intéresse parallèlement aux champs des humanités numériques, par des travaux portant sur le web scientifique, la philologie numérique et ses enjeux épistémologiques, les écritures en réseau et les cultures numériques.

Droits d'auteur

© Éditions de la FMSH