

POUR UNE PRATIQUE RÉFLEXIVE DES DOCUMENTS NUMÉRIQUES

Björn-Olav Dozo

(F.R.S.-FNRS – Université de Liège)

Résumé. L'article examine des exemples de pratiques des documents numériques liés à la recherche universitaire. Par l'examen de quatre axes (conservation, diffusion, usages et transmission), il entend illustrer certains enjeux pratiques, méthodologiques et épistémologiques du passage au numérique. Cette contribution souhaiterait ainsi susciter un intérêt collectif pour cette problématique.

Les documents numériques ont envahi nos disciplines. Toutes recourent au support digital à chaque étape de la production du savoir : de l'examen des sources (textes, photos, bases de données, etc.) à la diffusion des résultats, en passant par l'organisation des données grâce aux outils informatiques, le document numérique est de toutes nos pratiques de recherche.

On constate cependant que les livres conservent l'avantage, nombre de chercheurs préférant le confort d'un livre à celui d'un écran. Ce n'est pas cette opposition imprimé *vs* électronique que je voudrais prioritairement explorer ; celle-ci concerne le livre-objet, le livre matériel. S'il n'existe pas encore de substitut au livre-objet, de « liseuse » suffisamment complète et confortable pour satisfaire le plus grand nombre, le « contenu » du livre (à défaut d'un terme plus pertinent) a déjà quitté sa prison de papier. Tout est (sera) disponible sur Internet. Livres, presse, archives, cartes, photographies, musique, films, rien n'échappe au codage binaire.

Ce passage du papier à l'électronique n'est pas neutre. Il s'accompagne d'une reconfiguration profonde des pratiques de lecture et de consultation. En ce domaine, il a suscité l'apparition d'un paradigme de pratiques différent : la recherche (entendue comme mode d'accès à l'information). John Battelle (2005) l'a bien montré dans son essai *The Search. How Google and Its Rivals*

Rewrote the Rules of Business and Transformed Our Culture. La diversité du web n'est souvent accessible, pour le non-spécialiste, que par la recherche par mots clés. Pour accéder au web, pas d'ouvrages de référence permettant d'aborder un sujet inconnu, pas de bibliographie disciplinaire à dépouiller systématiquement, pas d'index des notions ou des noms propres. Juste un petit curseur clignotant dans une case sur une page blanche. Quelques mots à écrire, une liste de résultats, un coup d'œil rapide sur l'extrait pour juger de la pertinence et l'accès à un site totalement ignoré jusque-là. On s'interrogera légitimement sur les conséquences dans nos pratiques de recherche (au sens disciplinaire cette fois).

Les initiatives comme le projet Google Recherche de Livres (*Google Books*), qui visent à étendre cette forme d'accès aux livres, donc à la matière première de nos disciplines, font mesurer à quel point ce bouleversement réclame une théorie de la pratique, une réflexion sur les usages, non pas généralisée et prophétique, comme le sujet en charrie souvent, mais concrète et circonstanciée, à partir des outils existants et des pratiques effectives. Le temps est venu pour les chercheurs en sciences humaines et sociales de s'interroger sur leurs pratiques des documents numériques et d'en mesurer les implications épistémologiques. Clés d'entrée pour comprendre nos sociétés contemporaines, les documents numériques constituent un objet interdisciplinaire extrêmement mobilisateur : histoire, sociologie, économie, droit, informatique, pédagogie, bibliothéconomie, toutes ces disciplines — et d'autres encore — ont leur mot à dire sur cette réalité polymorphe.

Dans cet article, je voudrais baliser quatre domaines à explorer ; chacun d'eux concerne une facette de nos pratiques d'enseignants-chercheurs : les sources, la diffusion du savoir à travers ces dernières, leur usage et la transmission des connaissances afférentes. Ces quatre domaines me permettront, par l'examen de quelques-unes de leurs actualisations, d'aborder différentes pratiques des documents numériques dans la recherche en sciences humaines et sociales.

1. Sources : numérisation et conservation des documents

Rétrospectivement, on peut affirmer que c'est l'annonce par Google de son projet *Google Print* (rebaptisé depuis *Google Books*) qui a servi de catalyseur au débat sur la numérisation massive des livres et lui a donné une large dimension

publique. Fidèle à son projet original — « organiser toute l'information du monde » —, Google s'est lancé dans la numérisation et la mise à disposition des livres (quitte à créer un « marché » du livre sur support informatique) ; initialement, les ouvrages provenaient de grandes bibliothèques universitaires américaines et anglaises (Universités de Harvard, de Stanford, du Michigan et d'Oxford, auxquelles il faut ajouter la *New York Public Library*). Ce n'est pas le lieu ici de retracer en détail l'historique du projet, même si cette démarche pourrait s'avérer intéressante afin de rendre compte de ce tournant de l'histoire culturelle mondiale¹. Je me contenterai d'évoquer les grandes lignes des débats suscités par cette initiative.

Un article de Jean-Noël Jeanneney² dans *Le Monde* fut à l'origine du principal débat en langue française. Le même auteur publia en 2005 un pamphlet, plaidoyer pour un sursaut de l'Europe face à l'ogre privé américain. Cet essai, caricatural par certains aspects, fut à l'origine du projet européen d'une grande bibliothèque publique en ligne, visant à défendre la « culture européenne » : *Europeana*³. Le *New York Times*, lors du lancement de celle-ci, pointait surtout la politique culturelle de la France, soucieuse de sa puissance :

France has never been shy about promoting its culture, so few were surprised when it took a close interest in a new digital library intended to showcase Europe's history, literature, arts and science. But when the new site, called Europeana, begins life on Thursday, more than half of its two million items will come from just one of the 27 countries in the European Union: France. (St. Castle 2008)

J'ai pu défendre ailleurs, avec François Provenzano, que les arguments de J.-N. Jeanneney renvoyaient à une vision essentialiste de la notion de culture, prolongeant la politique française en la matière⁴ (B.-O. Dozo & Fr. Provenzano 2007).

Dans sa charge polémique, J.-N. Jeanneney confondait deux dimensions différentes du projet de Google : la numérisation des fonds de bibliothèques,

1 Pour en situer les grandes étapes, voir l'historique du projet dans la présentation qu'en fait Google : <http://books.google.com/googlebooks/history.html>.

2 Président de la Bibliothèque nationale de France à l'époque des faits (2004-2005).

3 Le site est accessible à l'adresse <http://www.europeana.eu>.

4 L'argument de J.-N. Jeanneney a perdu de sa puissance quand Google a signé des partenariats avec des bibliothèques européennes aux vastes fonds francophones. Citons notamment la bibliothèque cantonale et universitaire de Lausanne, la bibliothèque municipale de Lyon et la bibliothèque universitaire de Gand.

bien souvent constitués d'ouvrages indisponibles et/ou libres de droit (mais pas uniquement, j'y reviendrai), et le programme « Partenaires », destinés aux éditeurs. Ce programme doit permettre à chaque éditeur de contrôler sa présence sur *Google Books*, en choisissant le contenu disponible à la visualisation par l'utilisateur du service de Google. Certains petits éditeurs y ont vu un moyen de bénéficier d'une plus grande visibilité. Une enquête visant à connaître les véritables retombées du projet serait évidemment nécessaire pour évaluer la pertinence des *a priori* de chacun. L'histoire du livre et la sociologie de la lecture ont avec ce projet un objet fascinant.

Le lancement d'*Europeana* a montré combien les gestionnaires du projet avait une conception d'Internet fort différente de celle de Google : ce dernier importe dans le domaine du livre un modèle issu de la culture Internet (en particulier du logiciel libre, avec pour devise « *release early, release often* »), privatisé pour les intérêts d'une société. Ce modèle a vu se multiplier au fil du temps les ajustements successifs du service *Google Books* (évolution de l'interface, évolution des partenariats, apparition de la possibilité de télécharger des livres libres de droit en PDF, généralisation de l'*opt-out*⁵, etc.). *Europeana* est, en revanche, un projet public, financé par l'Europe et fruit d'une collaboration entre une multitude de partenaires institutionnels. En raison des hautes ambitions qu'il voulait rencontrer dès son lancement, le projet fut maintenu en longue gestation avant que les utilisateurs ne puissent profiter pleinement de ses services : si l'interface fut rapidement présentée et testable, elle n'était que la vitrine d'un savoir-faire technique sans contenu. Cette conception « traditionnelle » pour laquelle a opté le projet européen eut pour conséquence une inadaptation première du produit aux besoins du public : lors de son lancement, le 20 novembre 2008, l'affluence fut telle que les serveurs du projet n'y résistèrent pas et que le site fut directement indisponible. Il fallut ce lancement raté pour que la logique Internet s'impose : le projet est maintenant présenté en version beta et une section « laboratoire d'idées » propose des modules en phase de test.

À cette différence de conception de l'Internet correspond aussi une différence de financement : celui de Google est privé (fondé sur la publicité

5 L'*opt-out* est la pratique consistant à tout inclure lors de la numérisation, quitte à ce que les éditeurs ou les ayant-droits refusant de faire partie du projet se manifestent eux-mêmes et soient retirés par la suite.

ciblée et les partenariats commerciaux), ce qui rend son projet fragile sur le long terme ; celui d'*Europeana*, quoique public, n'en offre pas pour autant de garanties éternelles (que deviendra Europeana si l'Europe ne juge plus le projet prioritaire ?). On peut lire sur le site que « la dernière tâche du projet sera de recommander un modèle économique capable d'assurer la pérennité du portail »⁶. Par ailleurs, une éventuelle faillite de Google a poussé de nombreuses universités américaines⁷ à former un consortium pour créer un entrepôt numérique partagé, le Hathi Trust. L'objectif est la mise en commun de tous leurs fonds numériques et l'intégration de toutes les formes d'archives numériques. Un portail d'accès aux livres du monde, en quelque sorte.

Google vise lui-même l'intégration généralisée de ses différents services, qui indexent toutes sortes de contenus et réfèrent l'un à l'autre. *Google Scholar*, qui répertorie toute la littérature scientifique, renvoie ainsi naturellement à *Google Books*. Ce dernier a pour vocation de devenir un portail d'accès aux livres sous toutes leurs formes, fût-ce sous la seule forme d'une référence. Dans ce but, Google a conclu un partenariat avec OCLC⁸, responsable du catalogue *WorldCat*, regroupant les références de toutes les bibliothèques du monde. Google permet donc de repérer une œuvre, puis de la localiser dans les bibliothèques les plus proches. Il produit ensuite une carte (grâce à *Google Maps*, intégré lui aussi) qui indique le chemin le plus court entre son domicile et cette bibliothèque...

Je l'ai dit, tous ces services sont gratuits. Le modèle économique de Google, essentiellement fondé sur la publicité, a déjà été analysé à de nombreuses reprises. Il faudrait cependant mentionner une ressource complémentaire dont Google a fait grand bruit : l'accord entre Google et les éditeurs américains⁹.

6 Voir <http://www.europeana.eu/portal/aboutus.html>.

7 On y remarque notamment la présence de l'Université du Michigan, l'une des premières partenaires de *Google Books* (voir la composition du Hathi Trust à cette adresse : <http://www.hathitrust.org/partnership>, consulté le 15 mars 2009).

8 <http://www.oclc.org/news/releases/200811.htm>.

9 Voir l'accord à l'adresse <http://books.google.com/booksrightsholders/agreement-contents.html>. Voir aussi le document explicatif de J. Band (2008). Olivier Ertzschedl fournit sur son blog une analyse intéressante de cet accord (O. Ertzscheid 2008a & 2008b). Il faut aussi noter le point de vue assez corporatiste d'A. Gallimard (2009) sur cet accord entre Google et les éditeurs américains, point de vue que d'autres éditeurs français — François Bon de publie.net notamment — ne partagent pas complètement. Voir Fr. Bon (2009).

Cet accord, dans les grandes lignes, donne la possibilité à Google de vendre un accès numérique aux livres toujours sous *copyright*, tout en étant épuisé. Pour Olivier Ertzsched, Google est devenu, par cet accord, éditeur et libraire de « littérature grise ». À mon sens, au-delà de la métaphore, Google est plutôt devenu un fournisseur d'accès payant à cette littérature, dont les premiers « consommateurs » sont les universitaires, voire les universités, qui ont fourni une partie des volumes numérisés. Cela induit un paradoxe : les universités paient pour l'accès à un service — certes plus vaste que leur propre fonds — qu'elles ont contribué à créer... Les contrats entre Google et les différentes bibliothèques universitaires ont évolué, mais une constante est observée : Google est tenu de remettre à l'université une copie numérique (en format image dans les premiers contrats) de tous les livres numérisés¹⁰.

Si la forte domination de Google doit pousser les utilisateurs à rester vigilants et critiques — il ne faudrait pas confondre les activités et les services gratuits avec de la philanthropie —, le potentiel de ces nouveaux outils ne doit pas être négligé. En effet, l'accord prévoit une clause intéressante pour la recherche en littérature. On y lit ainsi :

*In addition to the institutional subscriptions and the free public access terminals, the agreement also creates opportunities for researchers to study the millions of volumes in the Book Search index. Academics will be able to apply through an institution to run computational queries through the index without actually reading individual books.*¹¹

Cette possibilité d'accès à des millions d'ouvrages évoque de manière surprenante les ambitions de Franco Moretti, professeur de littérature anglaise à Stanford, berceau historique de Google et université partenaire de *Google Books*. Fr. Moretti s'exprime sur sa démarche de recherche dans une interview accordée au *Guardian* :

Macro-criticism — it's an expression I personally never use, but it describes the direction in which I would like to move. In fact, the person with whom I've set up a laboratory on electronic data and literary theory speaks precisely of macro literary history. That's what it is. (J. Sutherland 2006)

Le dernier livre de Fr. Moretti, traduit en français sous le titre *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature* (2008), suggère quelques directions à explorer pour pratiquer cette macro-histoire de

¹⁰ Ce que l'université est en droit de faire de ces copies est extrêmement limité.

¹¹ <http://books.google.com/intl/en/googlebooks/agreement/#5>.

la littérature. Pour ne prendre qu'un exemple, son travail sur l'explosion de la production de romans dans différentes sociétés à un moment de leur histoire, met en évidence l'adéquation particulière entre ce genre littéraire et un état de société, que cela soit en Grande-Bretagne, en Italie, en France, aux États-Unis ou au Japon. Ce constat n'est réalisable qu'à partir de l'examen de très vastes corpus, dont *Google Books* constitue un modèle.

Néanmoins, les corpus en eux-mêmes ne suffisent pas :

[...] *larger and larger banks of data are becoming available, and we have absolutely no idea of how to deal with them. In just a few years, all the texts in existence will be online, and searchable. We really do not know how to pose useful questions to that mass of information.*

Merely having all these resources available doesn't mean, of itself, that they produce interesting answers. They won't. You have to pose interesting questions. That is what people do in labs and that we, in English studies, are completely unprepared to do. The fact is, in the very near future we shall have a vastly different field of materials but, really, no different frame of mind to do something with those materials. (J. Sutherland 2006)

L'objectif de la recherche a toujours été de poser de bonnes questions, quel que soit le corpus. Cependant, l'avènement des corpus numériques massifs demande que l'on réévalue le type de questions posées et la portée de celles-ci. Si le phantasme d'une exhaustivité à la Pic de la Mirandole ressurgit, il faut insister sur le processus d'élaboration de ces corpus et ne pas les utiliser comme des acquis immuables. Les modalités de constitution de la base de données de *Google Books* mériteront toute l'attention des chercheurs quand ceux-ci pourront l'exploiter.

La politique de Google en matière de numérisation fait actuellement le bonheur des juristes : les questions juridiques restent l'une des problématiques concernant les documents numériques les plus riches et les plus étudiées, étant donné les répercussions matérielles immédiates qu'elles peuvent avoir. La justification par le « *fair use* » des choix opérés par *Google Books* a beaucoup occupé les avocats des éditeurs et de Google avant l'accord dont il a été question plus haut. Pourtant, cet accord ne règle pas la question une fois pour toutes : il permet juste à Google de trouver un terrain d'entente avec les éditeurs américains¹².

¹² La question du *fair use* sur Internet, notamment dans les contenus créés par les utilisateurs (« *user-generated contents* »), devrait passionner les juristes encore longtemps.

Internet pose en outre des questions spécifiques liées à la technique œuvrant derrière le médium : on peut citer par exemple la problématique de l'« *embedding* », dont le lecteur exportable de Youtube est le meilleur exemple¹³. Quel statut juridique donner à une œuvre conservée sur un serveur et intégrée (par un « appel » direct à l'œuvre stockée, et donc sans duplication de l'original¹⁴) dans d'autres pages Internet ? Est-ce une copie ? La suppression de l'original entraîne automatiquement la suppression de tous ses artefacts, mais la possibilité existe d'en jouir à plusieurs en même temps et dans les mêmes conditions. André Gunthert (2008) y voit la mort du *copyright*... Certains commentaires envisagent ces usages comme une vieille pratique de « piratage » de bande passante : on pointe vers un contenu qu'on n'héberge pas, ce qui diminue les frais liés au trafic en les déplaçant vers d'autres. Ces pratiques, avant d'être répandues par les sites de partage de vidéos et de photos, existaient déjà aux premiers temps du web, lorsqu'une image était volumineuse par rapport aux taux de transferts alors relativement limités.

L'image est aussi au fondement de problèmes juridiques plus classiques concernant le droit d'auteur. On connaît bien les problèmes de reproduction d'images, notamment dans les revues scientifiques. Le cas des revues électroniques est souvent encore plus délicat. En France, un cadre légal trop contraignant, la fameuse loi DADVSI, promulguée trop rapidement et soumise aux fortes pressions des lobbys de l'industrie de la culture défendant aveuglément leurs intérêts, a entraîné des conséquences fâcheuses pour certaines revues : le problème de la protection des droits d'auteur et des droits dérivés pour les œuvres d'art et les images a contraint, par exemple, *Études photographiques* à fermer pendant quelque temps¹⁵. D'autres problèmes, plus subtils, liés à la diffusion des images existent également, telle l'indexation interdite pour les

13 Cette intégration dans des pages externes est aussi possible pour les livres numérisés par *Google Books* (voir <http://code.google.com/intl/fr/apis/books/>).

14 En tout cas pas une duplication pérenne (il ne faut jamais oublier que tout transfert de données informatiques se fait par duplication de l'original).

15 Voir <http://www.arhv.lhivic.org/index.php/2007/06/24/427-le-droit-aux-images-et-la-publication-scientifique>, <http://www.arhv.lhivic.org/index.php?q=photographie+num%C3%A9rique> et <http://etudesphotographiques.revues.org/index1213.html>. D'autres lois (européennes dans ce cas, mais protégeant les intérêts des mêmes lobbys) suscitent aussi l'inquiétude des chercheurs, concernant cette fois les archives sonores : <http://phonotheque.hypotheses.org/608> (consulté le 15 mars 2009).

autres moteurs de recherche des photographies de *Life* numérisées et hébergées par Google¹⁶.

2. Édition numérique et diffusion du savoir

La publication électronique en sciences humaines et sociales a déjà suscité d'intenses réflexions. L'émergence de l'édition électronique¹⁷ durant la dernière décennie a donné lieu à des débats passionnants, mêlant arguments scientifiques, politiques, économiques, voire idéologiques¹⁸. Des concepts spécifiques ont vu le jour ; un lexique partagé par les chercheurs s'intéressant au domaine s'est progressivement imposé (archives ouvertes, « voie royale », collecte des méta-données, *pre-print*, *post-print*, portail de revues, rétro-archivage, barrière temporelle flottante, etc.). Une étape a été franchie quand différentes initiatives d'observation de ce passage au numérique sont apparues¹⁹. Dans ce domaine en perpétuelle évolution, où les nouveautés remplacent ce qui existait, la création de lieux de mémoire permettra d'écrire l'histoire de ce passage.

Pierre Mounier (2008) a suggéré qu'avec le passage au numérique, le métier d'éditeur avait connu de profonds bouleversements ces dix dernières années en s'adaptant à la logique de la « longue traîne », décrite par Chris Anderson (2006) au moyen de ces deux slogans principaux : « *make everything available* » et « *help me find it* ». Qu'en est-il, plus spécifiquement, de l'édition en sciences humaines et sociales ? Si la revue constitue à ce jour le type de publication qui a le mieux négocié son passage au numérique, en adoptant massivement ce modèle économique, qu'en est-il du livre ? La mise en avant de l'article comme vecteur de diffusion de la recherche en sciences humaines et sociales a lieu en partie sur le modèle des sciences dures. Il a des conséquences sur la pratique même de la recherche, allant de pair avec la modification de son

16 Voir <http://blogoscoped.com/archive/2009-01-06-n79.html>.

17 Voir notamment G. Boismenu & G. Beaudry (2002) ; P. Mounier (2003) ; M. Pierssens (2005).

18 Il est possible de se faire une bonne idée de ces débats en consultant la *Scholarly Electronic Publishing Bibliography* (<http://www.digital-scholarship.org/sepb/>), qui recense et sélectionne les documents en anglais utiles à la compréhension des efforts de publication savante sur Internet.

19 La France s'est notamment dotée d'un observatoire de ces ressources et pratiques, le TGE-Adonis (<http://www.tge-adonis.fr/>).

mode de financement : publication collective, multiplication des publications et accélération du rythme de publication. Le format de l'article de revue correspond à cette évolution : d'un format de plus en plus standardisé (le résumé et la structure du développement général vont dans ce sens), il permet une lecture rapide, possible (pour certains) à l'écran.

Le livre de sciences humaines et sociales est-il définitivement inadapté à l'ère numérique ou sa dématérialisation est-elle, au contraire, sa bouée de sauvetage ? Le développement des « liseuses » (Sony Reader, Amazon Kindle, etc.) et des livres au format électronique pourrait ainsi offrir une seconde jeunesse au livre de sciences humaines. L'évolution de ces « lecteurs » et de leurs interfaces, qui permettront, à terme, de prendre facilement des notes numériques, de faire des recherches en texte plein, d'utiliser des copier/coller, etc., le tout en étant connecté à Internet, donnera au lecteur la même liberté qu'il a aujourd'hui avec les revues devant son ordinateur. Certains acteurs français majeurs du secteur de la publication en sciences humaines et sociales l'ont bien compris, tel *Revues.org*, qui étudie très concrètement le lancement d'une section « livres électroniques ». Le portail Erudit propose déjà des volumes numériques. Plus généralement, le projet PKP entend développer un système de gestion du processus de publication tirant pleinement parti du numérique²⁰. Enfin, de manière plus radicale, The University of Michigan Press a décidé de publier toutes ses monographies sous format numérique, avec possibilité d'une impression à la demande²¹. René Audet profite de cette annonce pour s'interroger :

Les études littéraires, comme certaines autres disciplines des sciences humaines, restent particulièrement attachées à la brique (à la monographie volumineuse et parfois verbeuse), alors que plusieurs autres disciplines (pas nécessairement toutes du domaine des sciences inhumaines) ont considérablement fait bouger les paramètres de ce qui est recevable, disciplinairement parlant, comme contribution scientifique. C'est donc ici un très beau cas de médiologie. Nous convenons aisément que la prose académique impose un format au livre qui la reçoit ; mais acceptons-nous de considérer que le support que nous utilisons conditionne notre propos ? Et de façon encore plus sensible : acceptons-nous même l'idée que la monographie puisse délaissé sa place comme étalon de la recherche littéraire aboutie ? qu'elle délaissé le support du livre pour s'incarner sur de nouveaux supports (et ainsi,

20 Voir la présentation à l'adresse <http://pkp.sfu.ca/omp>.

21 Voir l'annonce sur le blog de l'éditeur : http://umichpress.typepad.com/university_of_michigan_pr/2009/03/um-redefining-scholarly-publications-in-the-digital-age.html.

potentiellement, délaissier le filé du discours qui le caractérise [cf. Barthes, « *Littérature et discontinu* »] ?²²

Si le livre au format électronique fait certainement partie de l'avenir de la publication en sciences humaines et sociales, le numérique offre en outre des possibilités inédites. Le concept de *Liquid publication*²³ va dans ce sens : les auteurs évoquent une publication évolutive et collaborative, fondée principalement sur le modèle des logiciels libres. On peut aisément imaginer des dictionnaires ou des encyclopédies, sous la responsabilité scientifique de groupes de recherche²⁴, qui prendraient cette forme. Les capacités de stockage ne posant plus de problème, il est envisageable de donner des niveaux de détail différents en fonction des usages, de conserver l'historique des évolutions des articles, etc. Bref, si la publication électronique en sciences humaines et sociales a déjà beaucoup évolué, elle devrait encore connaître de profonds bouleversements et, en retour, modifier nos pratiques de recherche.

3. Usages des documents numérisés : recherche et sérendipité

La recherche par mots-clés comme mode d'accès à la connaissance a déjà été largement étudiée. J'ai signalé plus haut l'ouvrage de John Battelle (2005). Il faut aussi pointer le quatorzième numéro de la revue *Questions de communication*, consacré aux moteurs de recherche²⁵.

22 Extrait de <http://contemporain.info/audet/archives/465>.

23 Voir le site <http://www.liquidpub.org/>.

24 Cette précision introduit toute la différence avec Wikipedia : si l'on pousse sa logique de fonctionnement jusqu'au bout, on constate que l'encyclopédie en ligne la plus consultée à ce jour n'a pas pour but de fournir un accès à la vérité, mais plutôt de proposer un état de connaissance qui fait consensus. Toute innovation est bannie de Wikipedia si elle n'est pas cautionnée par une source faisant autorité dans le milieu concerné par l'article. Ce modèle fait sens dans le cas d'une encyclopédie collaborative : quel autre critère trouver pour permettre à chacun de prendre part à une entreprise qui n'a pas de « responsable », d'« autorité » — au sens de l'*auctoritas*, de la paternité d'une œuvre, non pas de la hiérarchie, car il existe bel et bien une hiérarchie au sein des contributeurs de Wikipedia, garante de la « forme » et du « fonctionnement », mais pas du « contenu » ? Il limite en revanche fortement toute sortie des sentiers battus. Le modèle de *Knol* (<http://knol.google.com>), encyclopédie concurrente proposée par Google, réhabilite l'auteur et sa pleine responsabilité, mais suscite une multiplication d'articles sur le même sujet, chacun voyant midi à sa porte et pouvant ignorer le travail de l'autre.

25 *Moteurs de recherche. Usages et enjeux. Questions de communication*, n° 14, 2008.

Je souhaiterais revenir sur un concept présenté comme opérant pour rendre compte de la recherche par moteur : la sérendipité, ou la découverte par hasard. Si le terme n'est pas spécifique aux pratiques numériques (je vais y revenir), il n'en désigne pas moins un aspect courant de la recherche lors de l'usage de moteurs de recherche. La sérendipité, c'est la qualification d'une pratique de recherche qui aboutit à un résultat intéressant, quoiqu'inattendu (certains insistent sur la dimension d'échec de la recherche originelle).

Carlo Ginzburg (2006), dans un texte sur l'interface de consultation du catalogue de la bibliothèque de UCLA, *Orion*, décrit comment les hasards de sa consultation et les associations d'idées qu'elle entraîne le font progresser dans sa recherche de manière fortuite. Or le catalogue informatique n'est pas le seul moyen pour faire ce genre de découvertes : certains ont pertinemment insisté sur l'approche complémentaire que constituait l'examen physique des rayonnages d'une grande bibliothèque à toute mise au point bibliographique à partir d'un catalogue en ligne. Les découvertes fortuites sont là aussi fréquentes.

Si le hasard a toujours fait partie des découvertes scientifiques, la généralisation de l'usage des moteurs de recherche donne à cette dimension une prégnance plus forte. Le fait que les algorithmes des principaux moteurs soient brevetés et secrets empêche cependant d'analyser avec finesse les rapprochements qu'ils effectuent entre résultats.

Pourtant, si l'on veut comprendre le processus de découverte dans certaines disciplines, il faudra en passer par l'interrogation du fonctionnement des moteurs de recherche, dont l'usage a envahi notre quotidien de chercheurs. Qui n'a pas recours à Internet pour vérifier une référence ? trouver un exemple ou obtenir un aperçu rapide sur un sujet annexe ? Lors de ces recherches, combien de fois ne tombe-t-on pas sur une piste nouvelle ? L'article de C. Ginzburg en est une belle illustration. Si les catalogues de bibliothèques sont transparents dans la manière dont ils rapprochent les résultats, il en va autrement de l'outil « Recommandés pour vous » d'Amazon ou d'une recherche Google. Une histoire des découvertes scientifiques actuelles pourra se trouver enrichie de l'étude de ces logiques de rapprochements qui conditionnent notre accès à l'information.

4. Transmission des pratiques

Le fruit que l'on peut retirer des nouvelles pratiques jusqu'ici évoquées ne doit pas faire oublier que, même chez les utilisateurs qui semblent les plus avertis, certains mauvais usages des documents numériques pourraient apparaître. D'où la nécessité d'un enseignement ambitieux dans ce domaine. Il faut en effet se garder du mythe selon lequel les « *digital natives* » auraient, de manière innée, un savoir ou un savoir-faire concernant les ordinateurs. Contrairement à cette croyance, la numérisation des documents, loin de concrétiser l'utopie de l'accès pour tous au savoir du monde, risque de renforcer les inégalités dues aux disparités d'éducation ou, plus précisément, aux difficultés que P. Bourdieu et J.-C. Passeron (1964 & 1970) repéraient déjà dans le système scolaire français, à savoir l'inexistence d'une véritable « égalité des chances » face à l'apprentissage. La formation aux bonnes pratiques du document numérique ne peut pas demeurer en dehors de l'université. C'est à celle-ci de prendre conscience des enjeux considérables en la matière et de se donner les moyens d'une transmission efficace et réfléchie des pratiques approfondies de ces documents. On l'a vu, la recherche s'en est emparée — ou devrait le faire. L'université n'est-elle pas le lieu d'un enseignement de la recherche ?

Comme exemple, on citera l'usage d'une source documentaire comme Wikipedia. Portée aux nues ou vouée aux gémonies, l'encyclopédie collaborative a suscité diverses pétitions de principe de la part d'enseignants, déclarations peu efficaces pour nourrir un débat raisonné sur les usages et les limites de l'outil. Pourtant, cette base de connaissances ne peut être simplement ignorée. J'y vois pour ma part un magnifique terrain d'expérimentation (au sens fort) pour la critique historique. L'historique des articles ainsi que la page de discussion à leur propos sont autant de passages obligés pour se faire une idée de la valeur de ce qui y est affirmé. Parfois, les débats entre contributeurs, le profil de ceux-ci ou les sources extérieures apprennent autant sur le sujet que l'article lui-même. C'est que le savoir consensuel tel que le fabrique Wikipedia ne se résume pas au texte : il est aussi — voire surtout — affaire de contexte de production.

5. Conclusion

Ces quelques exemples montrent que la pratique des documents numériques n'a rien d'évident. Elle ouvre sur le vaste champ des « *digital humanities* »

qui ont le vent en poupe depuis quelques années. De discipline-gadget pour chercheurs en marge, elles sont en passe de s'imposer comme un domaine légitime. Si les études numériques sont l'objet de quelques « pôles d'excellence », elles sont à ce jour peu présentes dans les universités, et moins encore dans les cursus de formation des étudiants. À partir de ce constat, il m'a semblé utile de montrer combien il est nécessaire que chaque chercheur mobilisant des documents numériques réfléchisse à sa propre pratique.

Négliger cette part de réflexivité revient à se condamner à subir le sort de Funes, le personnage de Borges qui possède une mémoire eidétique : encombrés de nos immenses mémoires externalisées, nous pourrions en oublier de généraliser, c'est-à-dire de départiculariser, et donc de théoriser. Les documents numériques ne peuvent nous faire oublier qu'exhaustivité ne vaut pas explication et que toute connaissance (même pratique) est fondée sur différentes opérations de conceptualisation. Fin observateur des usages du numérique, Chr. Anderson (2008) constate également la nécessité de repenser nos pratiques de recherche quand nous sommes confrontés à une telle massification des sources. Sa rhétorique provocante et la conception fort peu estimable qu'il semble avoir d'une certaine théorie pourraient faire penser qu'il croit à l'émergence d'une logique immanente du magma des données. À le lire, on découvre au contraire que c'est le rapport même du chercheur avec ses sources qui a changé du fait de la numérisation : ce constat est une évidence en sciences dures ; il me semble indispensable d'en prendre conscience en sciences humaines et sociales.

Du point de vue d'une institution universitaire, une pratique réflexive des documents numériques par les chercheurs ne peut être que bénéfique. Les « *digital humanities* » sont encore un secteur de niche, mais elles bénéficient d'une reconnaissance grandissante ; il n'est pas trop tard pour l'investir et jouer un rôle de pionnier.

Bibliographie

- CHR. ANDERSON (2007), *The Long Tail: Why the Future of Business is Selling Less of More*, New York, Hyperion, 2006, trad. fr. *La longue traîne. La nouvelle économie est là !*, Paris, Pearson Education.
- (2008), *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, dans *Wired magazine* (23/6/2008), http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (consulté le 18 avril 2009).

- J. BAND (2008), *A Guide for the Perplexed: Libraries and the Google Library Project Settlement* (13/11/2008), <http://www.arl.org/bm%7Edoc/google-settlement-13nov08.pdf> (consulté le 15 mars 2009).
- J. BATTELLE (2005), *The Search. How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*, New York, Portfolio.
- G. BOISMENU & G. BEAUDRY (2002), *Le nouveau monde numérique : le cas des revues universitaires*. Montréal, PUM.
- Fr. BON (2009), *Édition numérique : rue Sébastien-Bottin, Québec*, dans *Le blog|journal* (18/03/2009), <http://www.tierslivre.net/spip/spip.php?article1698> (consulté le 15 mars 2009).
- P. BOURDIEU & J.-C. PASSERON (1964) *Les Héritiers. Les étudiants et la culture*, Paris, Minuit.
- (1970), *La Reproduction : Éléments d'une théorie du système d'enseignement*, Paris, Minuit.
- St. CASTLE (2008), *France Dominates Europe's Digital Library*, dans *The New York Times* (20/11/2008), http://www.nytimes.com/2008/11/20/business/worldbusiness/20digital.html?_r=4&partner=rss&emc=rss&pagewanted=print (consulté le 15 mars 2009).
- B.-O. DOZO & Fr. PROVENZANO (2007), *Google patron de l'édition*, dans *La Libre Belgique* (31/08/2007), http://www.lalibre.be/index.php?view=article&art_id=367429 (consulté le 15 mars 2009).
- O. ERTSCHEID (2008a), *Le marché du livre comme algorithme* (28/11/2008), http://affordance.typepad.com/mon_weblog/2008/11/le-tr%25C3%25A8s-r%25C3%25A9cent-accord-historique-pass%25C3%25A9-par-google-avec-les-%25C3%25A9diteurs-am%25C3%25A9ricains-continue-de-faire-causer-sur-la-toile-et.html.
- (2008b), *Le marché du livre comme algorithme (suite)* (30/11/2008), http://affordance.typepad.com/mon_weblog/2008/11/le-march%25C3%25A9-du-livre-comme-algorithme-suite-.html.
- A. GALLIMARD (2009), *Défendons le libre commerce des idées et de la création*, dans *Le Monde* (12/3/2009), http://www.lemonde.fr/opinions/article/2009/03/12/defendons-le-libre-commerce-des-idees-et-de-la-creation-par-antoine-gallimard_1167065_3232.html.
- C. GINZBURG (2006), *Conversation avec Orion*, dans *Matériaux pour l'histoire de notre temps* 82, p. 129-132.
- A. GUNTHERT (2008), *La lecture exportable ou la mort du copyright* (6/7/2008), <http://www.arhv.lhivic.org/index.php/2008/07/06/762-la-lecture-exportable-ou-la-mort-du-copyright> (consulté le 15 mars 2009).

- J.-N. JEANNENEY (2005), *Quand Google défie l'Europe : plaidoyer pour un sursaut*, Paris, Mille et une Nuits.
- Fr. MORETTI (2008), *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*, Paris, Les Prairies ordinaires, coll. « Penser/croiser ».
- P. MOUNIER (2003), *Édition scientifique et universitaire : un nouveau monde numérique ?* (23/1/2003), dans *Homo Numericus*, <http://homo-numericus.net/spip.php?article195> (consulté le 19 mars 2009).
- (2008), *L'édition en ligne : un nouvel eldorado ?*, dans *Blogo Numericus* (22/5/2008), <http://blog.homo-numericus.net/spip.php?article154> (consulté le 25 mai 2008).
- M. PIERSSENS (2005), *Revue savantes : quel avenir ?*, dans B. MELANÇON [dir.], *Le savoir des livres*, Montréal, PUM, p. 71-94.
- J. SUTHERLAND (2006), *The ideas interview: Franco Moretti John Sutherland meets a great iconoclast of literary criticism*, dans *The Guardian* (9/1/2006), <http://www.guardian.co.uk/books/2006/jan/09/highereducation.academicexperts> (consulté le 15 mars 2009).