

Diplôme national de master

Domaine - sciences humaines et sociales

Mention - sciences de l'information et des bibliothèques

Spécialité - cultures de l'écrit et de l'image

Le numérique au service de l'historien du livre

Alexis Burlot

Sous la direction de Dominique Varry
Professeur des Universités – École Nationale Supérieure des Sciences de
l'Information et des Bibliothèques

Remerciements

Un grand merci à Dominique Varry pour le suivi de ce travail de recherche et pour l'énergie qu'il déploie à transmettre sa passion pour le livre ancien et la bibliographie matérielle.

Mes remerciements vont également à Éric Guichard pour son introduction aux humanités numériques et son enthousiasme pédagogique et intellectuel. Je saurai faire mien son combat pour l'interdisciplinarité.

Enfin, je tiens à saluer l'énergie que savent véhiculer collègues et proches qui, au-delà même de la sphère intellectuelle, rendent la vie plus douce.

Résumé :

Le numérique fait aujourd'hui une entrée remarquée au sein des pratiques historiennes. Il bouleverse en profondeur le travail du chercheur en ce qu'il est le média de sources scripturales d'un genre nouveau : les livres imprimés anciens numérisés. Ces derniers se multiplient à grande vitesse sur la Toile et forcent à une réflexion sur la nature de l'offre numérique et sur le travail historique qu'elle rend possible et conditionne. Elle nous conduit finalement à interroger les affinités disciplinaires qui existent aujourd'hui entre sciences humaines et sociales et sciences informatiques et questionne l'être naissant des humanités numériques.

Descripteurs : numérisation – livre ancien – bibliothèque numérique – fichier numérique – océrisation – indexation – métadonnées – conservation patrimoniale – bibliographie matérielle – art typographique – culturomique – quantophrénie – SHS – sciences des données – informatique – humanités numériques

Abstract :

Digital technology is today penetrating the field of historian's practices. It deeply disrupts scientific research for it has become the medium of new written historical sources : the digitized old printed books. The latter are rapidly multiplying on the Internet and force us to think about the nature of the digital offer and about the historical work it could determine. This digital offer finally leads us to assess the ties which bring the social and human sciences closer to the computer sciences, it questions the emerging notion of digital humanities.

Keywords : Digitization – Old Books – Digital Library – Digital File – Optical Character Recognition (OCR) – Indexing – Metadata – Patrimonial Preservation – Bibliography – Typography – Culturomics – Quantophrenia – Social and Human Sciences – Data Sciences – Information Technology – Digital Humanities

Droits d'auteurs



Cette création est mise à disposition selon le contrat :
« **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France** »
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou
par courrier postal à *Creative Commons*, 171 Second Street, Suite 300, San
Francisco, California 94105, USA.

Sommaire

SIGLES ET ABRÉVIATIONS	9
INTRODUCTION.....	11
L’OFFRE NUMÉRIQUE	15
LA BIBLIOTHÈQUE UNIVERSELLE S’APPELLERA-T-ELLE	
GOOGLE LIVRES ?.....	15
<i>Le mythe de la bibliothèque universelle</i>	<i>15</i>
<i>Une offre pléthorique.....</i>	<i>17</i>
<i>La suprématie googléienne.....</i>	<i>22</i>
QUE RESTE-T-IL DE NOS BIBLIOTHÈQUES ?	24
<i>Le « vrac »</i>	<i>25</i>
<i>Existe-t-il une « bibliothèque numérique » ? Pour quels publics ?</i>	<i>27</i>
<i>Ressources thématiques, les nouvelles collections ?.....</i>	<i>30</i>
<i>Une politique de numérisation ?.....</i>	<i>40</i>
LA NATURE DE L’OFFRE	43
À LA SOURCE DU TRAVAIL DE L’HISTORIEN, UN FICHIER	
NUMÉRIQUE ?	43
<i>La valeur de la source : sa stabilité ? Lutte contre l’instabilité</i> <i>ontologique du fichier numérique</i>	<i>43</i>
<i>La searchability du fichier, condition de son existence ?</i>	<i>45</i>
<i>En amont : l’indexation et les métadonnées</i>	<i>48</i>
<i>En aval : conservation et pérennisation.....</i>	<i>52</i>
LE LIVRE NUMÉRISÉ.....	54
<i>Quid de l’original ?</i>	<i>54</i>
<i>Le processus de numérisation.....</i>	<i>57</i>
<i>Mode texte / Mode image</i>	<i>58</i>
<i>Quel rendu ?.....</i>	<i>63</i>
TRAVAIL DES SOURCES	75
RÉALITÉ MATÉRIELLE DE L’IMMATÉRIEL	
<i>Quid de la matérialité du livre ?</i>	<i>75</i>
<i>Bibliographie (im)matérielle, fondements d’une archéologie du livre</i> <i>numérisé</i>	<i>77</i>
<i>Les espoirs du numérique.....</i>	<i>79</i>
QUE RESTE-T-IL DE L’OBJET-LIVRE ?	85
<i>La reconnaissance optique de caractères, mémoire de l’art</i> <i>typographique</i>	<i>85</i>
<i>La recherche du fragment, la découverte du Tout-venant.....</i>	<i>91</i>
<i>La mort de l’objet-livre, essai de philologie numérique</i>	<i>92</i>
TRAVAILLER AVEC CES SOURCES	95

FAUT-IL SE JOUER DES DONNÉES ?	95
<i>Le traitement des données</i>	95
« <i>Culturomique</i> » et « <i>quantophrénie</i> »	99
<i>L'affrontement entre « quantitatif » et « qualitatif »</i>	103
<i>Une crise historiographique ?</i>	105
DE LA SCIENCE HISTORIQUE À LA SCIENCE INFORMATIQUE :	
FONDEMENTS D'UNE NOUVELLE DISCIPLINE ?	108
<i>L'avènement du « cyberhistorien » ?</i>	108
<i>Nouvelles contraintes disciplinaires</i>	110
<i>Les « humanités numériques », aux fondements d'une nouvelle</i>	
<i>épistémologie ?</i>	112
« <i>Transdiscipline</i> » et « <i>Data Sciences</i> »	115
CONCLUSION	117
BIBLIOGRAPHIE	121
Le livre, les bibliothèques et le numérique	121
<i>Monographies</i>	121
<i>Articles et revues</i>	123
L'historien, ses outils et le numérique	126
<i>Monographies</i>	126
<i>Articles et revues</i>	128
TABLE DES ILLUSTRATIONS	137
TABLE DES MATIÈRES	139

Sigles et abréviations

BAnQ : Bibliothèque et Archives nationales du Québec

BBF : Bulletin des Bibliothèques de France

BmL : Bibliothèque municipale de Lyon

BnF : Bibliothèque nationale de France

BNM : Bibliothèque numérique mondiale

CHR : Centre de recherches historiques

EHESS : École des hautes études en sciences sociales

Enssib : École nationale supérieure de l'information et des bibliothèques

JDN : Journal du Net

RFN : Réseau francophone numérique

SHS : Sciences humaines et sociales

INTRODUCTION

« La révolution sera numérique »¹. En ces termes, le lanceur d'alerte anonyme des « *Panama Papers* », sous le pseudonyme de John Doe, intitule en mai 2016 un manifeste destiné à justifier l'acte de piratage informatique par lequel il s'est procuré quelque onze millions de fichiers d'archives du cabinet d'avocats Mossack Fonseca. Ces documents, qu'il a transmis au journal allemand *Süddeutsche Zeitung*, ont donné lieu à un travail journalistique international d'analyse informatique de données visant à dénoncer une « corruption massive et généralisée » et d'innombrables pratiques de dissimulation de fonds et d'évasion fiscale dans le monde entier.

Si le scandale médiatique, politique et sociétal qui en a découlé est peu en lien avec le petit monde du livre ancien et de l'histoire de l'écrit, l'intitulé du manifeste n'en est pas moins pertinent, et si actuel. Car au sein de la discipline historique et des pratiques de l'écrit, la révolution est également numérique. Les avancées techniques liées au numérique bouleversent en profondeur les outils d'approche et d'appropriation de l'information scripturale, qu'elle soit contemporaine ou passée. Nul ne peut nier la rupture épistémique que marque l'irruption des pratiques numériques dans les processus de construction et d'assimilation de la connaissance en ce qu'elle est médiatisée par l'écrit.

En quelques décennies, et particulièrement avec la naissance du Web dans les années 1990, le numérique a révolutionné – au sens de brusque changement de paradigme – le rapport de l'individu à la lecture et à l'écriture. Ont été ainsi reléguées au second plan, d'abord avec l'avènement maintenant ancien des éditeurs de texte puis avec la naissance plus récente de l'Internet, nombre de pratiques d'écriture et de lecture attachées à la matérialité de l'objet-*papier*. L'ordinateur et son raccord avec le réseau du *World Wide Web* sont à l'origine d'une dématérialisation de l'écrit, non plus figé par l'immuabilité de son support physique mais sujet au contraire à une labilité constante qui interroge le cadre, la valeur et la portée de la littérature au sens premier de « chose écrite ».

Il est ici à distinguer deux types de cette forme nouvelle d'écrit, à savoir tout d'abord un écrit nativement numérique, contemporain, qui naît du clavier de l'ordinateur et n'a qu'occasionnellement vocation à être imprimé, et l'écrit nativement manuscrit ou typographié devenu *a posteriori* numérique par un processus technique dit de numérisation.

¹ À ce sujet, voir le manifeste de John Doe, dont *Le Monde* a proposé une traduction le 6 mai 2015, consultable à l'adresse suivante : http://www.lemonde.fr/evasion-fiscale/article/2016/05/06/la-revolution-sera-numerique-le-manifeste-de-john-doe-le-lanceur-d-alerte-des-panama-papers_4915025_4862750.html.

L'écrit nativement numérique, fluide et reproductible à l'infini, soulève de très nombreuses questions relatives à la valeur de l'information contemporaine et induit une nouvelle littérature, c'est-à-dire une relation nouvelle au lire-écrire considéré dans sa dimension pratique, concrète, appliquée finalement à la réalité quotidienne des contacts sociaux². Il pose par ailleurs nombre de problèmes d'ordre culturel, économique et juridique, en lien avec la propriété intellectuelle et la marchandisation de l'information.

Ce ne sont pourtant pas ces enjeux qui retiendront ici notre attention, laquelle se portera sur la seconde forme que peut prendre l'écrit dans le monde numérique : l'écrit numérisé. Distinguons donc le *livre numérique* ou *électronique*, aussi appelé *e-book* d'après le terme anglais, et le *livre numérisé* issu de la numérisation d'un ouvrage papier.

Parmi ces livres numérisés, pour les problématiques qu'ils soulèvent et parce qu'ils constituent un ensemble relativement déterminé de l'histoire du livre, nous choisirons de considérer tout particulièrement ceux que l'on dit « anciens », dénomination admise pour qualifier les productions livresques d'une période historique qui s'étend de la naissance de l'imprimerie en Europe, vers 1450, au début du XIX^e siècle ; on parle alors de livre imprimé de la période artisanale. Ces livres anciens, depuis la *Biblia latina* à quarante-deux lignes sortie en 1455 de la première presse mise au point par Johannes Gutenberg, sont aujourd'hui objets de conservation et de collection et intéressent hautement bibliophiles, historiens du livres et acteurs patrimoniaux divers.

L'électronique, qui a rendu possible la transmission de l'information par signaux électriques, a conduit au développement de nouveaux moyens de constitution et de conservation de l'information : la photographie numérique et les appareils de scan – on parle en français de « numériseurs » – permettent maintenant de reproduire numériquement page par page un ouvrage papier, non plus sous un format analogique comme les premiers microfilms, mais sous la forme d'un document numérique. C'est donc sous cette forme que sont désormais reproduits et conservés les livres anciens qui nous sont parvenus, et c'est également sous cette forme qu'ils sont transmis, *via* un média omniprésent : l'Internet, où ils voyagent d'autant plus facilement qu'ils sont libres de droits.

Préservés des dommages irrémédiables que peuvent causer le temps et les lecteurs physiques, ces ouvrages sont majoritairement conservés à l'abri du public dans les fonds anciens des bibliothèques ou chez des particuliers et ne s'offrent que rarement au regard et à la consultation de tous : la numérisation induit donc un rapport nouveau au livre ancien. Le public, et parmi lui de nombreux chercheurs en histoire du livre, doivent désormais bien souvent se contenter d'une reproduction numérique de l'ouvrage à considérer. Il ne s'agit pas ici de juger de ces politiques

² Voir à ce sujet la page Wikipédia « Littérature » qui, bien que « jargonante » d'après l'évaluation des internautes-encylopédistes, offre une définition éclairante d'un terme que nous empruntons à Éric Guichard, page consultable à l'adresse suivante : <https://fr.wikipedia.org/wiki/Litt%C3%A9rature>.

de conservation, sur lesquelles nous reviendrons plus en détails, mais d'en extraire une idée principale, selon laquelle le numérique intervient aujourd'hui à tous les niveaux et à tous les stades de la recherche historique sur le livre comme support de l'écrit.

Loin d'être réductible à une simple photographie, le document numérisé s'offre à un maniement nouveau : sa reproductibilité et sa souplesse en font un objet neuf de recherche et de travail médiatisés par l'outil informatique. En cela, il faut bien voir le numérique comme un média nouveau au service de l'historien du livre.

Une telle problématique interroge les pratiques disciplinaires dans toute leur contemporanéité, relativement à des objets d'étude faussement anciens : car si les ouvrages abordés sont en effet antérieurs au XIX^e siècle, leur version numérique a bien souvent moins de dix ans d'âge. La numérisation s'avère ainsi un sujet d'étude et de questionnement des plus récents et des plus actuels, d'où la mobilisation de publications imprimées et électroniques tout aussi récentes, dans des formats divers qui vont de la monographie au post de blog. Par ailleurs, l'actualité de ces questionnements induit certainement, non seulement des biais historiques dus à un manque de recul sur le sujet, mais également des creux informationnels, comme les données chiffrées qui font parfois défaut.

Quoi qu'il en soit, le numérique et les pratiques de numérisation engendrent une démultiplication de l'offre informationnelle mise à la disposition de l'historien. Il s'agira d'en étudier la nature afin d'évaluer les potentialités qu'offre ce nouveau support de l'écrit en termes d'étude historique, au regard du poids épistémique et philologique de l'objet-livre comme lourd héritage de la culture occidentale. Enfin, la pratique de l'informatique, nécessaire au maniement le plus basique des documents numériques, pénètre au plus profond des pratiques scientifiques et se creuse progressivement une place de choix parmi les outils de traitement de l'information historique. Il nous semble important d'évaluer cette place nouvelle des sciences de l'informatique au cœur du métier d'historien, afin de ne pas se laisser dépasser par le déploiement contemporain d'une épistémologie nouvelle qui tend à faire trembler les fondations des sciences humaines et sociales.

L'OFFRE NUMÉRIQUE

LA BIBLIOTHÈQUE UNIVERSELLE S'APPELLERA-T-ELLE GOOGLE LIVRES ?

« Le rêve, jusque-là chimérique, d'une Bibliothèque universelle, librement accessible à tous, en tout lieu et à tout moment devient, grâce aux ordinateurs et à Internet, un objectif réaliste dans un délai raisonnable, pour peu que les acteurs publics et privés capables de le réaliser parviennent à harmoniser leurs efforts au bénéfice de l'intérêt général. La perspective d'une numérisation massive, rapide et systématique des œuvres laisse espérer la mise à disposition progressive sur le réseau de l'ensemble du patrimoine écrit de l'humanité, y compris les œuvres nouvelles qui viendront aujourd'hui ou demain l'enrichir. Nul ne contestera le formidable progrès démocratique et culturel que représente la possibilité pour chacun de lire gratuitement, sur son écran, n'importe quel ouvrage du domaine public ou encore de se procurer à moindre coût, en raison des économies de papier, de stockage et de distribution, les œuvres récentes protégées par le droit d'auteur. »³

Le mythe de la bibliothèque universelle

« Merveille du numérique. L'accès à tout, tout de suite. L'immédiat, l'exhaustif. Des catalogues et des fonds d'une richesse inégalée. La mise à jour permanente. La démocratisation de l'accès. L'écriture de tous et l'écriture pour tous. »⁴ En ces termes, l'économiste de la culture et des médias Françoise Benhamou résume les espoirs que nombre d'historiens formulent face au bouleversement scientifique et bibliothéconomique que représentent les dernières avancées technologiques en matière de numérisation.

La numérisation d'une incroyable quantité d'ouvrages, qui alimente un accès en ligne croissant au patrimoine mondial, marque pour beaucoup l'avènement d'un temps nouveau, celui de la réunion progressive du savoir commun et son accessibilité permanente pour tous grâce à l'Internet. En témoignent plusieurs

³ Luc FERRY, « Introduction » dans *La révolution du livre numérique : état des lieux, débats, enjeux*, Marc TESSIER, Bruno RACINE, Jean-Noël JEANNENEY, et al., Paris : Odile Jacob, 2011, p. 7

⁴ Françoise BENHAMOU, *Le livre à l'heure numérique : papier, écran, vers un nouveau vagabondage*, Paris, Ed. du Seuil, 2014, p. 211

interventions de l'historien Robert Darnton⁵, la multiplication et l'accroissement des bibliothèques numériques sur le Web portent à croire qu'un rêve transhistorique formulé par les Lumières du XVIII^e siècle est en passe de se concrétiser : la constitution d'une « bibliothèque universelle » qui réunirait tous les savoirs du monde.

Cette bibliothèque « totale » et organisée offrirait ses services à tous, une idée chère à la pensée humaniste que l'on traduit aujourd'hui par la perspective d'une « démocratisation du savoir » et que l'on rêvait alors dans une « république des Lettres », « un territoire sans police ni frontières et sans inégalités autres que celles des talents. »⁶. Pour les Lumières, ce travail de réunion des savoirs existants s'est concrétisé à travers la constitution de l'*Encyclopédie*, qui répondait alors à une ambition encore modeste d'exhaustivité. Le numérique est cause d'un brusque changement d'échelle dans l'appréhension de l'encyclopédisme : la masse d'information croissante disponible sur le Web dépasse l'entendement et se soustrait à toute appréhension organisationnelle. Le « tout numérique »⁷ ne se soumet que difficilement à des démarches de classement et de sélection de l'information ; au contraire, il met en exergue une contradiction inhérente à la fameuse utopie des Lumières, qui rêvaient d'une exhaustivité organisée, d'un Tout classifié comme s'il était lui-même constitutif d'un ensemble plus grand à partir duquel il pouvait être pensé. Évoquons ainsi les réserves qu'émet Jean-Noël Jeanneney à ce sujet : « Il ne peut exister [...] de bibliothèque universelle, tout au plus des regards spécifiques sur l'universel. »⁸

Il est à croire qu'il existe une différence fondamentale entre la masse informationnelle en elle-même et son appréhension, qui nécessite des principes d'organisation *a priori* ; cette différence n'étonnera nullement les historiens dont la tâche consiste justement à évaluer l'information pour travailler à la production du savoir. Il est cependant remarquable de voir combien le numérique brouille la frontière qui sépare « information » et « savoir », « volume » et « exhaustivité ». L'« infosphère d'une richesse écrasante » et son « paysage d'apparence chaotique » qu'évoque Frédéric Barbier⁹ substituent ainsi à l'exhaustivité et à la

⁵ « La démocratisation du savoir est désormais à portée de main, du moins en matière d'accès aux sources. L'idéal des Lumières ne pourrait-il pas devenir réalité ? » (Robert DARNTON, « Accès public, contrôle privé. La bibliothèque universelle, de Voltaire à Google. », *Le Monde diplomatique*, 1^{er} mars 2009, p. 24-25)

« Sans doute l'idéal de la diffusion des Lumières du XVIII^e siècle nous paraît-il archaïque aujourd'hui, mais il peut acquérir un lustre du XXI^e siècle si on l'associe avec Internet, qui transmet des messages à un coût voisin de zéro » (Robert DARNTON, « La chandelle de Jefferson, une bibliothèque numérique nationale », dans « Le livre, le numérique », *Le Débat*, n°170, 2012, p. 112)

⁶ Robert DARNTON, « Accès public, contrôle privé. La bibliothèque universelle, de Voltaire à Google. », *Le Monde diplomatique*, 1^{er} mars 2009, p. 24-25

⁷ « La numérisation des manuscrits médiévaux à la Bibliothèque nationale de France : programmes et état des pratiques », Charlotte Denoël, dans *L'historien face au manuscrit : du parchemin à la bibliothèque numérique*, Fabienne HENRYOT, Louvain-la-Neuve : Presses universitaires de Louvain, 2012, p. 333

⁸ Jean-Noël JEANNENEY, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, 3e éd., Paris : Éd. Mille et une nuits, 2010, p. 11

⁹ Frédéric BARBIER, *Histoire des bibliothèques : d'Alexandrie aux bibliothèques virtuelles*, Paris : A. Colin, 2013, p. 289

représentativité rêvées des grandes bibliothèques¹⁰ le volume et la variété d'une masse informationnelle non filtrée. Finalement, d'un contrôle *a priori* des connaissances offertes – c'est le travail des encyclopédistes comme celui des bibliothécaires –, on passe à une évaluation *a posteriori* de l'information : c'est à celui qui se confronte au Tout-accessible d'apprendre à démêler la trame si dense des connaissances offertes sur la Toile, au risque de s'y perdre. Car si l'offre est universelle, en matière d'ouvrages libres de droits du moins, les pratiques historiennes et informatiques le sont bien moins.

Indépendamment de ces craintes relatives à la croissance exponentielle des ressources disponibles sur l'Internet et à la position plus qu'instable du lecteur qui s'y aventure, il est certain que le numérique bouleverse présentement le monde du livre et que, pour reprendre une nouvelle fois les mots de Robert Darnton, « rien de tel n'a encore jamais existé. Une bibliothèque sans murs qui s'étendra partout et contiendra la quasi-totalité de ce qui se trouve dans les dépôts emmurés de la culture humaine. *E pluribus unum* ! Jefferson eût apprécié. »¹¹

Une offre pléthorique

Il faut dire, en des termes empruntés à Roger Chartier, qu'il est une « puissance propre du média électronique pour le projet encyclopédique »¹². L'offre de documents numérisés disponibles sur la Toile est pléthorique et ne cesse de croître : depuis le projet Gutenberg en 1971¹³, et tout particulièrement depuis le début du XXI^e siècle, les projets de bibliothèques numériques se sont multipliés. Si la définition de « bibliothèque numérique » est problématique – nous y reviendrons –, nous considérons ici les cas pour lesquels le regroupement des ressources numérisées ne répond pas à des critères thématiques mais à une volonté assumée de constituer un fonds pluridisciplinaire aussi vaste que possible.

Il va sans dire qu'une présentation exhaustive des ressources du Web et de leur histoire n'est pas envisageable, d'autant que manquent les chiffres précis et récents

¹⁰ Pensons à la Bibliothèque d'Alexandrie qui réunissait les ouvrages les plus importants de l'Antiquité, jusqu'à 700 000 à l'époque du règne de César, aux dires de l'historien antique Ammien Marcellin, qui évoque « *septingenta uoluminum milia* » (XXII, 16), référence à retrouver sur Wikipédia : https://fr.wikipedia.org/wiki/Biblioth%C3%A8que_d%27Alexandrie

¹¹ Robert DARNTON, « La chandelle de Jefferson, une bibliothèque numérique nationale », dans « Le livre, le numérique », *Le Débat*, n°170, 2012, p. 119

¹² Roger CHARTIER, *Le livre en révolutions : entretiens avec Jean Lebrun*, Paris : Textuel, 1997, p. 134

¹³ En 1971, Michael Hart fonde à l'université de l'Illinois ce qui deviendra en 2000 le Projet Gutenberg, qui consistait en la retranscription sur ordinateur du texte d'ouvrages existants. Si une grande attention était et est encore accordée à la conservation de la mise en page générale, les ouvrages ne sont proposés qu'en version nativement numérique et ne sont pas accompagnés d'une image scannée de l'original papier. En mai 2016, l'offre affichée sur le site Internet du projet avoisine les 52 000 livres électroniques à télécharger.

Le site est consultable à l'adresse suivante : <https://www.gutenberg.org/>.

sur la constitution de l'offre. Les quelques projets évoqués ci-après ne sont donc que les grands moteurs d'une dynamique de numérisation qui touche l'ensemble du monde de l'imprimé, des géants comme Google jusqu'aux structures patrimoniales de petite taille. Cette vague de numérisation est ainsi illustrée par Gallica, la bibliothèque numérique de la Bibliothèque nationale de France, qui regroupe, outre ses propres ouvrages numérisés, un nombre non négligeable de livres numérisés par ses bibliothèques partenaires, soit 30 % du fonds offert par les « plusieurs centaines d'institutions (bibliothèques, centres de recherche, sociétés savantes) » qui « contribuent à faire de Gallica une bibliothèque numérique collective », comme cela est présenté sur leur site Internet.



Le projet **Gallica**¹⁴, né dès 1997, consistait en la numérisation simple – c'est-à-dire en un scan – d'œuvres littéraires et de dossiers documentaires, à destination des « gens de culture »¹⁵. S'y ajoutèrent ensuite progressivement livres et périodiques préalablement sélectionnés. Ce n'est qu'à partir de 2007 qu'est lancée une nouvelle phase de cette politique de numérisation, baptisée *Gallica 2*, qui a encore cours aujourd'hui et qui consiste en une numérisation de masse par tranche de côte.¹⁶ Ce changement de politique est bien sûr directement lié au lancement de Google Livres, la bibliothèque numérique du géant Google. Aujourd'hui, *Gallica* offre un accès à plus de 650 000 ouvrages numérisés. Si le chiffre global est tenu à jour sur leur site Internet, il n'y a cependant pas de chiffre sur la part exacte qu'y constituent les ouvrages libres de droits, les livres anciens ou encore les ouvrages consultables en mode texte.

Selon un schéma comparable à celui que nous fournit la Bibliothèque nationale de France et avant que ne naissent des politiques de numérisation que l'on pourrait dire inter-structurelles – qui regroupent donc des structures patrimoniales diverses, nationales ou internationales –, il est à noter les deux importants projets de numérisation longuement étudiés par l'auteure Gaëlle Béquet¹⁷ que sont ceux menés par la British Library située à Londres et l'Österreichische Nationalbibliothek de Vienne, en Autriche.



Le projet **Europeana**¹⁸ est quant à lui une réponse européenne au projet de numérisation de Google, sur lequel nous reviendrons. Europeana, née en 2008 à l'initiative de la

¹⁴ Le site est consultable à l'adresse suivante : <http://gallica.bnf.fr/>

¹⁵ Christian LACOMBE, « La numérisation ou le grand sommeil du livre imprimé », *Documentation et bibliothèques*, Montréal, Asted, vol. 55, n°1, 2009, p. 39

¹⁶ Les Anglais distinguent ainsi la « boutique digitalisation » de la « mass digitalisation ».

¹⁷ Gaëlle BÉQUET, *Trois bibliothèques européennes face à Google : aux origines de la Bibliothèque numérique (1990-2010)*, Paris : École des Chartes, 2014, 473 p.

¹⁸ Le site est consultable à l'adresse suivante : <http://www.europeana.eu/portal/>

France¹⁹, est une bibliothèque numérique collective à l'échelle de l'Europe qui rassemble le travail d'une très importante quantité de partenaires²⁰ et qui est coordonnée par un consortium d'une trentaine de bibliothèques nationales. Son rôle n'est pas d'archiver les ressources qu'elle rassemble mais de les cataloguer et de renvoyer le lecteur vers les sites des institutions partenaires. Loin de restreindre son activité aux seuls livres imprimés, ce projet concerne, tout comme Gallica, un vaste panel d'objets constitutifs du patrimoine européen. Elle regroupe aujourd'hui plus de cinquante-trois millions de documents. Ainsi résumant-ils leur but sur la page d'accueil de leur site Internet : « *We transform the world with culture! We want to build on Europe's rich heritage and make it easier for people to use, whether for work, for learning or just for fun.* »²¹. Europeana, parce qu'elle est une plate-forme plus qu'une bibliothèque virtuelle et parce qu'elle brasse des ouvrages numérisés par d'autres institutions, rend compte de la difficulté qu'il y a à définir le concept de bibliothèque numérique, le projet n'étant rattaché à aucune infrastructure déterminée. Cependant, parce que son fonds s'avère *a minima* structuré, parce que les ressources proposées sont des plus conséquentes et parce qu'un travail de supervision œuvre à leur homogénéisation, il semble difficile de l'évincer d'un tel état des lieux.



HathiTrust²² est une bibliothèque numérique née dans le cadre du *Committee on Institutional Cooperation* et supervisée par les universités de l'Indiana et du Michigan. Ce projet rassemble une cinquantaine de bibliothèques

américaines ainsi que l'université de Madrid depuis 2010²³. HathiTrust propose aujourd'hui plus de quatorze millions d'ouvrages au format numérique dont plus de cinq millions cinq cent mille tombés dans le domaine public, soit près de 40 % des ouvrages consultables par qui le souhaite.



L'**Internet Archive**²⁴, supporté par Yahoo depuis 1996, est un organisme américain à but non lucratif qui a pour mission d'archiver le Web, par la capture d'une quantité astronomique de pages Internet mais également par la conservation des documents véhiculés sur la Toile, qu'il s'agisse de logiciels, de films,

¹⁹ Le projet est énergiquement porté par la BnF.

²⁰ La liste des institutions partenaires est consultable à l'adresse suivante :

<http://www.europeana.eu/portal/browse/sources.html>

²¹ « On transforme le monde par la culture ! Nous voulons construire sur le riche héritage culturel européen et donner aux gens la possibilité de le réutiliser facilement, pour leur travail, pour leur apprentissage personnel ou tout simplement pour s'amuser. », affiche la version française du site.

²² Le site est consultable à l'adresse suivante : <https://www.hathitrust.org/>

²³ Thierry CLAERR et Isabelle WESTEEL (dir.), *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la librairie, 2013, p. 28

²⁴ Le site est consultable à l'adresse suivante : <https://archive.org/>

d'enregistrements sonores ou de livres²⁵. Il sert donc également de bibliothèque numérique, une bibliothèque riche de plus d'un million de livres numérisés, majoritairement anglo-saxons, appartenant au domaine public.²⁶



En 2009 ouvre la **Bibliothèque numérique mondiale**²⁷, la *World Digital Library*, à l'initiative de l'UNESCO et de la Bibliothèque du Congrès américain. Si de nombreuses institutions ont signé un partenariat avec la BNM, l'offre culturelle de nature très variée reste relativement limitée. L'institution est du moins une des rares à tenir à jour des statistiques très précises sur le contenu de son offre²⁸, dont nous retiendrons en mai 2016 qu'elle affirme proposer au téléchargement quelques 373 000 livres sur un total de presque 671 000 fichiers. Parmi les 131 partenaires issus de 58 pays différents annoncés, dont 103 bibliothèques, on retrouve notamment la BnF et la British Library. La Bibliothèque du Congrès reste cependant le principal acteur de cette bibliothèque mondiale qui brasse du contenu plus qu'il n'en numérise. À son sujet, Thierry Claerr et Isabelle Westeel déclarent que la BNM « s'apparente davantage à une vaste banque de données culturelles et multilingues »²⁹ qu'à une véritable bibliothèque numérique.



Plus récemment encore, en 2013, naît la **Digital Public Library of America**³⁰ (DPLA), la bibliothèque numérique publique américaine, à destination d'un public universel. Le projet est à l'initiative de Robert Darnton, directeur de la *Harvard University Library*, avec le soutien de la Alfred P. Sloan Foundation, et a pour but de centraliser l'objet des politiques de numérisation américaines.³¹ Une nouvelle fois, une telle « bibliothèque universelle » fonctionne comme un agrégateur de contenus, qui moissonne et référence les ouvrages numérisés par des institutions distinctes. Remercions la DPLA de tenir également à jour une page Internet relative à son contenu³² qui affiche un fonds de presque 2 500 000 livres et périodiques, constitué par les apports de cent-cinquante *contributing institutions*³³ et de six partenaires, parmi lesquels HathiTrust et les plus grosses universités américaines. Notons enfin que si presque 1 500 000 documents sont de langue anglaise, les langues allemande et française n'en sont pas moins correctement

²⁵ Voir Wikipédia : https://fr.wikipedia.org/wiki/Internet_Archive

²⁶ Thierry CLAERR et Isabelle WESTEEL (dir.), *Manuel de constitution de bibliothèques numériques...* p. 28

²⁷ Le site est consultable à l'adresse suivante : <https://www.wdl.org/fr/>

²⁸ La page « Statistiques » est consultable à l'adresse suivante : <https://www.wdl.org/fr/statistics/>

²⁹ Thierry CLAERR et Isabelle WESTEEL (dir.), *Manuel de constitution de bibliothèques numériques...* p. 28

³⁰ Le site est consultable à l'adresse suivante : <https://dp.la/>

³¹ Voir Wikipédia : https://fr.wikipedia.org/wiki/Digital_Public_Library_of_America

³² La page « Bookshelf » est consultable à l'adresse suivante : <https://dp.la/bookshelf>

³³ « institutions contribuant au projet »

représentées, avec plus de 288 000 fichiers pour la première et 242 000 pour la seconde, l'ensemble des langues concernées atteignant les cent cinquante.³⁴



Ces différents projets alimentent une véritable *course à la numérisation*³⁵ démarrée en 2004 par le géant du Web Google qui, avec ce qui est devenu Google Books, ou **Google Livres**³⁶ en français, s'est lancé dans une campagne de numérisation sans précédent. Ce projet, consistant à numériser en quelques années entre six et quinze millions d'ouvrages issus des gigantesques collections des bibliothèques de Harvard, de Stanford, du Michigan, de Oxford et de la New York Public Library, a fait l'effet d'un véritable « tremblement de terre »³⁷ dans le monde des bibliothèques et, dans des proportions moindres, chez les chercheurs. Le projet est d'autant plus écrasant pour des structures nationales déjà en course comme la Bibliothèque nationale de France que la numérisation est prévue en mode texte, permettant ainsi une recherche d'une précision nouvelle dans ces œuvres *via* le moteur de recherche Google. Rapidement, les partenariats se multiplient, jusqu'à la Bibliothèque Municipale de Lyon qui en 2008 fut la première bibliothèque française à adhérer au projet³⁸. Cependant, si le projet est colossal et sa réalisation rapide, les acteurs du projet sont discrets et les chiffres précis manquent, depuis la promesse de Google d'atteindre les quinze millions d'ouvrages numérisés.

La réaction de plusieurs pays européens à un tel projet – américain qui plus est – est virulente : indépendamment des grands problèmes juridiques soulevés par la numérisation par Google d'ouvrages non libres de droits, et notamment des œuvres dites orphelines qui ne sont pas tombées dans le domaine public mais dont l'auteur est inconnu, c'est devant cette « menace » pour l'intégrité de la culture européenne que quelques grandes bibliothèques nationales insistent auprès de la Commission européenne et du Conseil de l'Europe pour que soit examiné un projet de Bibliothèque Numérique Européenne (BNE) qui deviendra Europeana. Par ailleurs, l'ambition de Google force les politiques de numérisation courante comme celles de la BnF ou de la British Library à emprunter une voie nouvelle : « Google Print », ancêtre de Google Books, marque ainsi la fin de la numérisation sélective, dite « de niche », et le début de la *numérisation de masse*. Si les débats relatifs à l'entreprise de Google ont été et sont encore des plus vifs, Google Books

³⁴ Avec des distinctions historiques toutefois, ex : Middle French (ca. 1400-1600) ou Old English (ca. 450-1100)

³⁵ Pensons au journaliste du *Monde* Philippe Escande qui dans son article : *La numérisation poursuit sa course effrénée*, daté du 5 décembre 2014, s'exclamait : « La numérisation du monde est en marche, ou plutôt en course folle. »

³⁶ Le site est consultable à l'adresse suivante : <https://books.google.fr/>

³⁷ Alain JACQUESSON, *Google livres et le futur des bibliothèques numériques : historique du projet, techniques documentaires, alternatives et controverses*, Paris : Éd. du Cercle de la librairie, 2010, p. 51

³⁸ Le partenariat entre Google et la Bibliothèque municipale de Lyon a donné lieu à la constitution de la base Numelyo, consultable à l'adresse suivante : <http://numelyo.bm-lyon.fr/>. Au sujet du fonds ancien, la BmL explique que « La collection des livres anciens [...] comprend 500 000 documents de l'incunable au livre imprimé jusqu'en 1920. Elle va être progressivement mise en ligne au fur et à mesure de l'avancée des opérations de numérisation. ». En mai 2016, plus de 10 000 livres anciens sont déjà disponibles au format numérique.

a du moins forcé l'Europe et le reste du monde à prendre place dans la fusée du numérique où se retrouvent tous les partisans d'une culture et d'un patrimoine mondial globalisés. « Au-delà de la diabolisation dont Google continue de faire l'objet, il faut le créditer d'un violent effet d'accélération de nos particules mentales »³⁹.

Ce « tournant Google Livres », qui a brusquement transformé les modalités de constitution d'une bibliothèque numérique « universelle », marque également l'abdication du savoir sélectionné devant l'information neutralisée, la victoire de la ressource sur la source. Jamais cependant le rêve de voir se constituer une bibliothèque exhaustive du patrimoine humain n'a été plus palpable ; comme le note Robert Darnton, « si les rêves d'une méga-méta-macro bibliothèque remontent aux Anciens, la possibilité d'en construire effectivement une est récente. Elle date de la création d'Internet (1974) et du Web (1991) »⁴⁰. Car la bibliothèque universelle se doit d'abolir les frontières, comme le résume Lucien Xavier Polastron dans *La Grande Numérisation* : « Transformer en vases communicants toutes les ressources existant *au monde* constitue la seule voie raisonnable vers une Très Grande Bibliothèque dont on peut rêver qu'elle agirait, sous l'impulsion soignée de ses usagers professionnels et l'enthousiasme des curieux, comme une métastase d'intelligence dans le tissu poisseux du Web. »⁴¹

La suprématie googléienne

Si le rôle de Google dans la constitution d'une bibliothèque mondialisée a été des plus discutés et a donné lieu à de houleux débats, il n'est pas le lieu ici de s'attarder sur les modalités de ces conflits interculturels. Deux points cependant ont de quoi retenir notre intérêt en ce qu'ils questionnent l'accès de l'historien aux sources qui fondent son travail de recherche. Le premier concerne l'*objectivité* du savoir moissonné par Google Livres, le second la *privatisation* du domaine public et du patrimoine mondial.

Le véritable pamphlet contre Google rédigé par le directeur de la BnF Jean-Noël Jeanneney en 2005 intitulé *Quand Google défie l'Europe*⁴², l'écho qu'il a pu trouver dans le milieu bibliothéconomique et intellectuel européen et les réponses qu'on a pu lui opposer résument les opinions divergentes qui animent partisans et détracteurs d'une bibliothèque universelle qui s'appellerait Google Livres.

³⁹ Jean SARZANA et Alain PIERROT, *Impressions numériques : quels futurs pour le livre ?*, Paris : Éd. du Cerf, 2011, p. 31

⁴⁰ Robert DARNTON, « La chandelle de Jefferson... », p. 112

⁴¹ Lucien Xavier POLASTRON, *La Grande Numérisation : y a-t-il une pensée après le papier ?*, Paris : Denoël, 2006, p. 121

⁴² Jean-Noël JEANNENEY, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, 3e éd., Paris : Éd. Mille et une nuits, 2010, 219 p.

Une première critique, fortement teintée d'anti-américanisme, anime les opposants à la position hégémonique de Google dans le domaine du livre numérisé, qu'est la crainte de voir s'établir un monopole sur la connaissance. Ce monopole consisterait à privilégier la production livresque anglo-saxonne au détriment d'autres aires civilisationnelles, notamment le monde francophone. Existente pourtant *a contrario* un partenariat entre la BnF et la Bibliothèque et Archives nationales du Québec (BAnQ), ou encore le Réseau francophone numérique (RFN)⁴³, créé dès 2006, qui réunit vingt-cinq des plus grandes institutions documentaires de la francophonie et destiné à la sauvegarde numérique des contenus de langue française.

Des partisans d'une entente possible avec Google, comme le successeur de Jean-Noël Jeanneney à la tête de la BnF : Bruno Racine, opposent à cette accusation des chiffres relatifs à la constitution des fonds aujourd'hui numérisés par la firme américaine : en 2009, sur les deux millions d'ouvrages du domaine public proposés par Google, une moitié ne sont pas en anglais et représentent quarante-deux langues différentes. « Le reproche adressé à Google de favoriser l'hégémonie de la culture anglo-saxonne apparaît dès lors comme un faux procès. »⁴⁴ Ces chiffres rendent en effet compte d'une hégémonie très relative du contenu anglophone, que des projets comme *Gallica* viennent largement contrebalancer. De même, la multiplication des partenariats de Google Livres avec des institutions nationales de pays très divers, comme l'université de Keiō à Tokyo, concorde plutôt avec son projet affiché d'embrasser la totalité de la culture mondiale.

Un second point inquiète les détracteurs de Google, qui concerne le risque de voir la production écrite internationale tombée dans le domaine public redevenir privée si Google décidait soudainement de rendre payant l'accès aux œuvres numérisées. Certains vont jusqu'à affirmer qu'il s'agit là à terme du destin de Google Livres⁴⁵, d'autres craignent que l'idéal humaniste affiché du projet ne soit battu en brèche si l'entreprise en arrive à changer de mains. Enfin, et c'est peut-être la crainte que partage une majorité des « googlo-sceptiques », la peur de voir Google faire faillite, l'entreprise privée quoique des plus fructueuses restant soumise aux lois du marché et de la capitalisation. Qu'advierait-il alors des fichiers numériques concentrés par Google qui se charge de leur diffusion ? Il est évident que la privatisation, sinon des ressources patrimoniales publiques elles-mêmes, du moins de leur gestion, est problématique. Il s'agit donc pour les partenaires de Google d'assurer leurs arrières lors de la signature du contrat, partant du constat que les modalités d'une collaboration avec Google ne sont pas figées, en témoignent notamment les débuts d'un dialogue qui s'est établi entre

⁴³ Le site est consultable à l'adresse suivante : <http://www.rfnum.org/pages/>

⁴⁴ Bruno RACINE, *Google et le nouveau monde*, Paris : Plon, 2010, p. 52

⁴⁵ « La gratuité des œuvres du domaine public n'est donc pas assurée à long terme dans l'univers numérique. » (Gaëlle BÉQUET, *Trois bibliothèques européennes face à Google...*, p. 254)

Google et la BnF en 2009, qui n'a cependant pas abouti⁴⁶. Marc Tessier, dans son *Rapport sur la numérisation du patrimoine écrit* remis au Ministre de la culture et de la communication le 12 janvier 2010⁴⁷, propose ainsi de nouvelles formes envisageables de partenariat avec la firme américaine. Ces propositions vont cependant à l'encontre des investissements importants de l'État français dans les politiques de numérisation et de l'énergie avec laquelle les autorités politiques et culturelles françaises luttent contre l'hégémonie googléienne.

Dans tous les cas, l'emprise de cette entreprise privée sur le monde de l'écrit continue de s'étendre et, quelle que soit la sincérité des motivations humanistes affichées par les dirigeants de Google, le moteur d'une entreprise n'en reste pas moins la recherche du profit. Comme Roger Chartier a pu le déclarer : « l'appropriation privée d'un patrimoine public, mis à disposition d'une entreprise commerciale, peut apparaître comme choquante. »⁴⁸ Marc Tessier, qui pourtant admet la nécessité de traiter avec le géant numérique, est catégorique : « L'Europe ne saurait confier l'avenir « numérique » de son patrimoine écrit à un seul opérateur privé ! »⁴⁹

Depuis que de nombreuses institutions patrimoniales se sont lancées, au niveau national comme international, dans une véritable course au tout-numérisé, l'offre numérique disponible sur le Web s'est ainsi considérablement accrue. En quelques années, avec en tête de course le géant américain Google et à sa suite de grandes infrastructures comme la BnF, les bases de données numériques à vocation universelle se sont multipliées au point d'impacter le rôle des bibliothèques physiques.

QUE RESTE-T-IL DE NOS BIBLIOTHÈQUES ?

Les bibliothèques physiques à vocation patrimoniale qui participent pleinement à ces campagnes de numérisation de masse contribuent à enrichir le World Wide Web de ressources constitutives d'une forme nouvelle de bibliothéconomie. Quoique les fonds ainsi constitués puissent être dénués d'apparent principe organisationnel, les bases de données marquent aujourd'hui

⁴⁶ Le 19 août 2009, *La Tribune* titre « Livre en ligne : Google a gagné » avec pour surtitre « La Bibliothèque nationale de France ferait appel aux services du géant ». Si en réalité le processus n'en était qu'à la négociation, l'affaire fut à l'origine d'un tel tollé médiatique, dans lequel Jean-Noël Jeanneney avait belle place, que l'affaire prit rapidement fin.

⁴⁷ TESSIER, Marc, *Rapport sur la numérisation du patrimoine écrit*, remis par Marc Tessier au ministre de la Culture et de la Communication le 12 janvier 2010, Paris, Ministère de la Culture et de la Communication, 2010, 64 p.

⁴⁸ Robert CHARTIER, « L'avenir numérique du livre », *Le Monde*, mardi 27 octobre 2009, p. 20

⁴⁹ Titre d'un entretien avec Marc Tessier publié dans *La révolution du livre numérique* ..., p. 17

l'avènement des « bibliothèques numériques », une notion vague qui sert à qualifier les ensembles de documents numériques et numérisés mis en ligne sur Internet. Ce « vrac » apparent, qui est l'objet d'une réelle politique de numérisation, contient ainsi les ressources et les collections de demain.

Le « vrac »

Quelles que puissent être les critiques adressées au fonctionnement de la bibliothèque numérique de Google Livres, et quels que soient les efforts d'autres bibliothèques comme Gallica, parce que la logique de numérisation de masse a supplanté un système de sélection ciblée, le principe même de la bibliothèque numérique induit une offre non triée dont la valeur est directement liée à la quantité d'ouvrages numérisés. Comme l'indiquent Jean Sarzana et Alain Pierrot dans *Impressions numériques, quel futur pour le livre ?*, « La numérisation est un processus industriel »⁵⁰ ; or tout processus industriel s'inscrit dans une visée de rendement qui passe par une mécanisation croissante et, ici, par une simplification du processus de sélection qui précède la numérisation, d'autant plus parce qu'elle requiert la mobilisation de compétences humaines. Aussi la numérisation de masse fait-elle parfois fi de toute considération bibliothéconomique ou scientifique.

Ainsi Jean Sarzana exprime-t-il le mode de fonctionnement de l'entreprise Google : « Google s'affiche depuis l'origine comme un outil de progrès universel, sa démarche et son langage sont de type messianique. [...] Google est un fabricant d'algorithmes, pas un éditeur. À l'origine, la substance ne l'intéresse pas, simplement la mise en relation, la chaîne qui devient réseau, puis rhizome, davantage la capacité à générer des flux que les flux eux-mêmes, et encore moins ce qu'ils charrient. »⁵¹ Partant de l'idée selon laquelle la largeur du fonds prévaut sur sa constitution, que l'importance du flux informatif prévaut sur la nature de l'information elle-même, une telle affirmation – dont il serait difficile d'admettre en toute bonne foi qu'elle est pleinement fautive – corrobore l'utilisation récurrente par Jean-Noël Jeanneney du terme de « vrac » pour qualifier le fondement de la bibliothèque numérique googléienne. Jean Sarzana, sur ce point, s'interroge : « Le recours à la numérisation massive implique-t-elle qu'elle soit pratiquée en vrac, la sélection intervenant dans un second temps, ou bien peut-on opérer des choix dès l'origine du programme ? »⁵²

Peut-être faut-il cependant réévaluer la valeur du « vrac » comme mode de présentation de l'information et lui ôter son caractère péjoratif, qui tend à associer au désordre apparent une inutilité fondamentale. Rappelons d'ailleurs, comme

⁵⁰ Jean SARZANA et Alain PIERROT, *Impressions numériques...*, p. 35

⁵¹ *Id.*, *op. cit.*, p. 30

⁵² *Id.*, *op. cit.*, p. 34

Bruno Racine, qu'un tel vrac est *de facto* conditionné par la pré-sélection des ouvrages opérée par les bibliothèques partenaires qui sont parmi les plus prestigieuses du monde : on peine alors à croire qu'un tel vrac puisse être de qualité si médiocre⁵³. À ce sujet, Jean-Noël Jeanneney et Bruno Racine s'affrontent par ouvrages interposés sur l'usage qu'il peut être fait du « vrac » numérique, au regard des pratiques de recherche contemporaines.

Lorsque Jean Sarzana se demande s'il est utile d'envisager une classification antérieure à toute numérisation, il répond : « Pour le bibliothécaire de métier, proche du livre, le choix doit précéder le passage au scanner. Tous les livres ne se valent pas en termes de discours. »⁵⁴ Est-ce à dire que l'ordre et le tri informationnels sont du goût des seuls bibliothécaires et qu'il est bon de proposer du « vrac » à d'autres publics ? Il est à croire en tout cas que Jean-Noël Jeanneney tient des propos dignes d'un « bibliothécaire de métier », puisqu'il s'oppose catégoriquement à l'offre en vrac et défend la légitimité des experts à opérer une sélection hiérarchisée des œuvres à numériser⁵⁵. Face à lui, son successeur à la tête de la BnF, Bruno Racine, défend au contraire l'offre en « vrac » qu'il juge d'utilisation plus souple et plus intuitive. Dans le second cas, il devient cependant nécessaire pour le lecteur numérique de s'appuyer sur des outils techniques destinés à le seconder dans ses recherches. Dans un ouvrage qui confronte ces différents points de vue, Marc Tessier exprime un avis proche de celui de Bruno Racine qu'il formule ainsi : « Sur le Net, l'ordre émerge souvent du désordre et ce n'est pas toujours en imposant un ordre *a priori* qu'on surmontera le désordre. Il me semble qu'il vaudrait mieux plutôt combiner de façon dynamique la façon de chercher des universitaires, plus ordonnée, et celle du grand public, plus aléatoire. »⁵⁶

Finalement, il s'agit bien d'opposer en amont le quantitatif au qualitatif, sans toutefois les opposer : l'exhaustivité qui caractérise la bibliothèque universelle signifie-t-elle « le-plus-gros-volume-de-données-possible » ou bien induit-elle une sélection fondée sur un jugement de valeur qui a cela de problématique qu'il tient de la considération subjective d'un individu, subjectivité qui dès lors est critiquable ? Si la question se pose, notamment pour les structures patrimoniales dont les moyens et les ressources sont plus limitées et qui ne sont pas en mesure de numériser à tout-va, il n'en est pas vraiment de même pour des structures importantes comme la BnF qui numérisent maintenant par tranches de cote et qui visent, à plus ou moins long terme, un passage au « tout numérique ».

Aussi l'offre à laquelle l'historien a aujourd'hui accès est-elle pléthorique mais relève, pour les plus grosses bibliothèques numériques, d'une sélection en

⁵³ Bruno RACINE, *Google et le nouveau monde...* p. 55

⁵⁴ Jean SARZANA et Alain PIERROT, *Impressions numériques...*, p. 34

⁵⁵ Voir la partie VI « Organiser le savoir » de *Quand Google défie l'Europe...* de Jean-Noël Jeanneney

⁵⁶ Marc TESSIER, Bruno RACINE, Jean-Noël JEANNENEY *et al.*, *La révolution du livre numérique...* p. 23

« vrac » qui ne sera exhaustive que lorsque tout aura été numérisé. Avant cet âge d'or de l'écrit numérisé, qui portera ou non le nom de Google Books, le chercheur court le risque ou bien de ne pas trouver en ligne l'ouvrage qui l'intéresse, auquel cas il lui est possible de se reporter aux catalogues des bibliothèques physiques, ou bien d'effectuer une recherche de sources biaisée par un creux informationnel et de constituer un corpus numérique incomplet. D'ici à ce que le numérique recoupe l'ensemble de la production écrite conservée de l'humanité, il s'agit donc de rester critique face à un « vrac » de ressources qui a pu, ou pas, faire l'objet d'une sélection antérieure à sa numérisation ; cette organisation de l'information relève de la responsabilité des institutions patrimoniales et, peut-être, des « experts » qui en seraient chargés.

Existe-t-il une « bibliothèque numérique » ? Pour quels publics ?

Cette question du « vrac » informationnel représentatif de l'offre numérique des grandes structures évoquées précédemment, qu'il s'agisse d'ailleurs de Google Livres ou de structures bibliothéconomiques comme la BnF ou la British Library, questionne la pertinence même du terme de « bibliothèque numérique » utilisé pour qualifier les projets qui travaillent à la mise à disposition sur la Toile du patrimoine écrit. « Bibliothèque ou fournisseur de contenu en ligne ? », s'interroge ainsi Gaëlle Béquet dans le titre d'un chapitre de son ouvrage *Trois bibliothèques européennes face à Google*⁵⁷. C'est qu'on est loin de l'infrastructure physique, caractérisée par un bâtiment solide et un fonds palpable, qui depuis l'Antiquité caractérise le lieu qu'est la bibliothèque. Deux caractéristiques peut-être sont les plus soumises au changement : la notion de *collection*, d'une part, battue en brèche par le moissonnage numérique ; les *publics*, d'autre part, qui légitiment et donnent vie à la bibliothèque.

La notion de « collection » a-t-elle ainsi encore un sens, à l'heure du numérique ? Il est certain que l'incroyable quantité de documents véhiculée par le Web contrevient au principe figé de la collection comme ensemble organisé de documents, documents qui font justement sens par leur promiscuité. Cela dit, deux différents principes d'organisation peuvent aujourd'hui redonner vie à la collection dans le monde du numérique.

⁵⁷ Chapitre III de la seconde partie de *Trois bibliothèques européennes face à Google...*, Gaëlle BÉQUET, p. 245

Le premier est un principe humain et relève d'un travail antérieur à la numérisation. Il consiste à donner une cohérence *a priori* aux ensembles numérisés qui conservent ainsi les caractéristiques des collections.

D'abord, il est à remarquer que de nombreuses institutions patrimoniales chargées de numériser une part de leur fonds proposent les œuvres numérisées dans des ensembles thématiques et qui présentent à l'écran un rapport de contiguïté visuelle semblable à celui qui existe entre les ouvrages voisins d'un même rayonnage de bibliothèque « réelle ». Pour exemple, la British Library s'est attachée à numériser deux cent cinquante-trois *festival books* de la Renaissance, choisis parmi les quelques deux mille de ce genre que détient la bibliothèque, relatifs aux grandes fêtes qui purent avoir lieu en Europe entre 1475 et 1700 : mariages, funérailles royales ou nobles, couronnements, prises de cités et autres jours de liesse.⁵⁸

La cohérence des collections peut ainsi être préservée lorsque le travail de numérisation tient compte de la réalité physique des ensembles documentaires. Cela relève de la politique documentaire adoptée par les autorités patrimoniales.

Dans le même ordre d'idées, une collection peut naître du processus de numérisation, non pas en tant qu'elle lui préexiste, mais parce que la dématérialisation des supports se prête particulièrement au remaniement des fonds documentaires. Il est certain que la réorganisation des ensembles thématiques et leur transport sous forme de fichiers numériques restent infiniment plus faciles et plus rapides qu'un quelconque mouvement de fonds matériel en bibliothèque. Cela est d'autant plus vrai que le moissonnage par diverses banques de données d'ouvrages libres de droits, numérisés par des établissements divers, permet la mise en relation de fonds physiquement distincts qui, sur la Toile, deviennent accessibles depuis un même fonds numérique. La numérisation se prête ainsi aisément à un travail de fond sur la constitution des collections et il relève des bibliothèques de savoir faire appel à des connaisseurs – pensons aux historiens – pour les accompagner dans un travail réfléchi de réorganisation des fonds et des collections qui les constituent.

L'autre principe d'organisation est un principe d'ordre technique qui relève des outils mis à la disposition des lecteurs pour la recherche documentaire. Si l'on peut imaginer une machine capable d'organiser de manière automatique ou semi-automatique les ouvrages numérisés à partir des informations normalisées qui les caractérisent – les fameuses métadonnées, dont nous reparlerons –, considérons plutôt les outils qui servent justement à constituer, instantanément et continuellement, les ensembles documentaires d'aujourd'hui.

Les *moteurs de recherche*, qui brassent une quantité phénoménale d'informations pour offrir une sélection répondant à une requête de l'utilisateur, avec tous les biais qu'ils peuvent induire, semblent maintenant faire office de « machine-à-

⁵⁸ Voir la page consacrée sur le site de la British Library : <http://www.bl.uk/treasures/festivalbooks/homepage.html>

constituer-des-collections ». La cohérence thématique à laquelle ces collections éphémères répondent est donnée *a priori* par le lecteur-utilisateur qui constitue en une fraction de seconde un ensemble de documents choisis, censés répondre à la question formulée ou du moins proposer une offre sélectionnée et représentative de documents numériques. Ainsi la technique supplée-t-elle le bibliothécaire dans ce rôle de sélection de l'information.

En un sens, on peut dire que les moteurs de recherche qui moissonnent l'information numérique ont pour fonction de créer, avec une pertinence qu'ils ne sont pas en mesure d'évaluer eux-mêmes, des « collections » à l'obsolescence programmée.

Alain Jacquesson distingue de la même manière les corpus stables, c'est-à-dire les collections des bibliothèques physiques, des corpus ponctuellement dimensionnés par les moteurs de recherche⁵⁹. Aussi puissants que soient ces moteurs de recherche, ils sont cependant loin de fournir les services qui font la spécificité des bibliothèques et que liste Marie Lebert pour définir la bibliothèque numérique :

« Appelée également bibliothèque électronique ou bibliothèque virtuelle, une bibliothèque numérique est une entité utilisant les technologies numériques pour acquérir, stocker, préserver et diffuser des documents. [...] Une collection numérique devient une bibliothèque numérique lorsqu'elle répond aux quatre conditions suivantes : (a) elle peut être créée et produite dans un certain nombre d'endroits différents, mais elle doit être accessible en tant qu'entité unique ; (b) elle doit être organisée et indexée pour un accès aussi facile que possible ; (c) elle doit être stockée et gérée de manière à avoir une existence assez longue après sa création ; (d) elle doit trouver un équilibre entre le respect du droit d'auteur et la diffusion libre du savoir. »⁶⁰

L'efficacité d'une organisation bibliothéconomique de l'information sur l'Internet requiert donc une organisation préalable de l'information par les acteurs de la numérisation que les outils actuels de moissonnage et de tri de l'information ne peuvent remplacer *a posteriori*. Outre une conservation de la constitution elle-même des fonds physiques existants – constitution qui peut intéresser les historiens, notamment pour leurs travaux sur la circulation historique de l'écrit – il est nécessaire d'employer les technologies informatiques actuelles pour mettre en ordre la surabondance de l'écrit numérisé, et ceci afin de pouvoir répondre aux attentes de publics variés et notamment, dans le domaine du livre ancien, à celles des historiens. Car, ne l'oublions pas, « les bibliothèques doivent regarder au-delà

⁵⁹ Alain JACQUESSON, *Google livres et le futur des bibliothèques numériques...* p. 47

⁶⁰ Marie LEBERT, « Le livre numérique : chronologie 1971-2007 », dans *Dossier du Net des études françaises*, 2007, disponible à l'adresse suivante : <http://www.etudes-francaises.net/dossiers/chronologie.htm#01> et la définition qu'elle donne de la « bibliothèque numérique » dans le dictionnaire : <http://www.etudes-francaises.net/dico/>, citée par Gaëlle BÉQUET dans *Trois bibliothèques européennes face à Google...*, p. 14

de leurs missions de collecte et d'organisation de l'information et s'intéresser aux personnes auxquelles l'information est destinée. »⁶¹.

Ressources thématiques, les nouvelles collections ?

Il s'avère que de nombreux projets ont vu le jour qui s'adressent spécifiquement aux historiens, et particulièrement aux historiens du livre, et qui usent des progrès actuels dans le monde du numérique pour constituer des fonds thématiques très divers.

Ces projets thématiques sont de plusieurs sortes : certains groupements de spécialistes profitent ainsi de l'accessibilité et de la gratuité de l'information numérique sur l'Internet pour constituer des bases spécialisées de fichiers regroupés ou bien par « discipline », comme Europeana Regia, ou bien par « genre », comme les projets Fleuron-Vignette, Môriâne, Maguelone ou Passe-partout, pour ne citer que les plus gros projets francophones, consacrés aux ornements d'imprimerie.

Certains projets, enfin, concernent des fonds réunis pour leur cohérence linguistique ou géographique, notamment dans le domaine de la francophonie.

Ces quelques grands projets numériques thématiques illustrent les perspectives que peut offrir le numérique relativement au travail historique.

Projets « disciplinaires »

Sur le modèle du défunt projet **Debora** (Digital accEss to the BOoks of the RenAissance), qui naquit en 1999 à l'initiative d'un professeur de l'École nationale supérieure des sciences de l'information et des bibliothèques (Enssib), Richard Bouché⁶², et qui consistait en la constitution, sur trente mois, d'une station de travail donnant accès à des ouvrages rares du XVI^e siècle⁶³, nous pouvons envisager, à terme, la multiplication des projets de collections numériques centrées sur des périodes historiques déterminées. De manière à illustrer l'idée selon

⁶¹ Jia MI et Frederick NESTA, « Marketing library services to the Net generation », dans *Library Management*, t. 27, n° 6-7, 2006, p. 412, cités par Gaëlle BÉQUET dans *Trois bibliothèques européennes face à Google...*, p. 257

⁶² Voir Richard BOUCHÉ, Ibadjadene MAJID, *Le Project DEBORA (Digital accEss to Books of the renAissance)*, consultable à l'adresse suivante : http://www.webreview.dz/IMG/pdf/le_projet_DEBORA_digital_access_to_books_of_the_renaissance_.pdf

⁶³ Pour une étude plus poussée des teneurs et aboutissements de ce projet, voir Nicolas BARBEY, Jean GUILLEMAIN, Géraldine PEOC'H et Patrice RACT, *La renaissance du livre ancien : bilan du projet DEBORA et perspectives d'avenir*, Mémoire de recherche pour la validation du diplôme de conservateur de bibliothèques, Enssib, 2002, 71 p., disponible sur le Web : <http://www.enssib.fr/bibliotheque-numerique/documents/995-la-renaissance-du-livre-ancien.pdf>

laquelle les collections numériques existent et se pensent relativement à un public, évoquons rapidement quelques projets que l'on dira « disciplinaires ».

Le portail Ménéstrel détaille ainsi l'offre numérique qui existe sur l'Internet à destination des médiévistes.

Ménéstrel⁶⁴ est un portail francophone de référence en études médiévales qui date de 1997. « Il a, d'une part, vocation à contribuer au développement et à la publicité des ressources et travaux francophones pour l'étude du Moyen Âge. Il se propose, d'autre part, d'offrir gratuitement un répertoire critique de ressources disponibles dans les domaines les plus variés des études médiévistes. »⁶⁵ Il s'agit donc d'un projet de longue date destiné à servir d'outil à la recherche dans un domaine précis de l'histoire, par le biais des ressources informatiques. Ménéstrel propose notamment des liens vers les différentes bibliothèques numériques et virtuelles à destination des médiévistes.

Profitons-en pour noter une distinction établie sur le site Ménéstrel par Amandine Postec, rédactrice de la rubrique « Catalogues de manuscrits » du projet et conservatrice au service des manuscrits médiévaux du département des manuscrits de la BnF⁶⁶. Celle-ci distingue en effet deux expressions, « bibliothèque numérique » et « bibliothèque virtuelle » : la première sert à qualifier les bibliothèques « à vocation généraliste, offr[ant] un accès en ligne à des documents numérisés qui peuvent présenter une grande diversité dans leur contenu, leur provenance, leurs thèmes ou les supports. » ; la seconde qualifie les bibliothèques « à vocation thématique, [qui] entend[ent] reconstituer une collection spécifique, le plus souvent un fonds homogène ou les collections d'une bibliothèque ancienne. ». Nous avons choisi d'établir cette même distinction précédemment en usant des termes de « bibliothèque numérique » et de « collection numérique », quoique leur portée soit la même.

Si l'on retrouve parmi les « bibliothèques numériques » listées de grandes bibliothèques pluridisciplinaires telles que la BnF, la Bibliothèque Sainte-Geneviève ou encore Europeana, Ménéstrel présente également les « bibliothèques virtuelles » qui peuvent servir d'outils aux historiens médiévistes ; notons pour exemples des projets comme Europeana Regia ou la Bibliothèque virtuelle de Clairvaux. **Europeana Regia**⁶⁷ rassemble quelques neuf-cents manuscrits royaux européens du Moyen Âge à la Renaissance. La **Bibliothèque virtuelle de Clairvaux**⁶⁸ a quant à elle pour but de reconstituer virtuellement l'ancienne

⁶⁴ Le site est consultable à l'adresse suivante : <http://www.menestrel.fr/>

⁶⁵ Mathieu OLIVIER, « Ménéstrel : portail francophone de référence en études médiévales, et partenaire de l'IFHA », *Revue de l'IFHA*, 1/2009, p.113

⁶⁶ La page « Bibliothèques numériques et virtuelles » est consultable à l'adresse suivante : <http://www.menestrel.fr/spip.php?rubrique1935&lang=fr>

⁶⁷ Le site est consultable à l'adresse suivante : <http://www.europeanaregia.eu/fr>

⁶⁸ Le site est consultable à l'adresse suivante : <https://www.bibliotheque-virtuelle-clairvaux.com/>

bibliothèque de l'abbaye de Clairvaux et réunit plus d'un millier de manuscrits numérisés par une dizaine de bibliothèques partenaires françaises et européennes.

Ces deux exemples servent à illustrer l'intérêt que peut offrir Internet dans la constitution de fonds thématiques : le fichier numérique n'étant sur la Toile rattaché à un aucun lieu déterminé, et étant par ailleurs, d'un point de vue technique, multipliable à l'infini, des institutions diverses peuvent constituer ce qui s'apparente à de réelles collections en faisant simple œuvre de référencement et de catalogage, sans que la collection ait d'existence matérielle réelle.

Projets « genrés », l'exemple des ornements d'imprimerie

Certains projets numériques s'attachent à répondre à des thématiques d'une grande précision. Citons pour exemple une grande base de données consacrée aux marques typographiques : la **Printers' Devices**⁶⁹, constituée par l'Université de Barcelone, qui s'attache à référencer les marques typographiques d'imprimeurs. Lancé en 1998, cette base constituée à partir du fonds ancien de la Bibliothèque de l'Université de Barcelone s'intéresse à tous les pays européens sur une période qui s'étend du XVI^e siècle au XVIII^e siècle, avec une attention particulière portée sur les ouvrages espagnols, français, italiens, portugais, allemands et néerlandais. Les marques typographiques sont ainsi référencées et illustrées, et y sont adjointes par ailleurs des informations relatives à la situation de l'imprimeur et à son activité, à la période et au lieu concernés. Les marques sont quant à elle indexées à partir de mots-clés catalans, espagnols et anglais, permettant une recherche thématique des images. D'après le site de la base étaient référencées, en mars 2015, 1653 imprimeurs et 2776 images de marques⁷⁰.

Si l'on s'intéresse aux quelques 258 imprimeurs parisiens recensés, on trouve ainsi Charles Angot (**figure 1**), imprimeur d'une édition de la *Géométrie* de René Descartes que nous considérons plus loin⁷¹ et dont nous reconnâtrons le monogramme à motifs végétaux⁷².

⁶⁹ Printers' Devices, ou « Marques d'impressors », est consultable à l'adresse suivante : http://www.bib.ub.edu/fileadmin/impressors/cerca_eng.htm.

⁷⁰ Voir la page « Home » du site Printers' Devices, consultable à l'adresse suivante : http://www.bib.ub.edu/fileadmin/impressors/home_eng.htm.

⁷¹ Voir partie « Quel rendu ? », p. 63.

⁷² Voir **figure 18**, p. 67.



Centre de Recursos
per a l'Aprenentatge i
la Investigació



UNIVERSITAT DE
BARCELONA

Printers' Devices

[Search in Catalogue](#)

[Persistent link to this record](#)

Printer **Angot, Charles**

Alternative name *Angot, Carolus*

Alternative name *Angot, Ch. (Charles)*

Country [França](#)

City [Paris, França](#) [1657-1696]

Ensign *Au lion d'Or*

Source LC/NAF, 17 febr. 2015 (enc.: Angot, Charles)

Source Laurent-Vibert, R. Les marques de libr. et d'impr. France ..., 1925: f. 7

Device/Motto Monograma amb motius vegetals

Mark term [monograma](#)

Image



Figure 1. Sur le site « Printers' Devices » de l'Université de Barcelone, la fiche technique de l'imprimeur Charles Angot avec l'image de sa marque d'imprimerie, un monogramme de la majuscule lettre A ornée de motifs végétaux.

Partant de la difficulté qui est celle des spécialistes du livre ancien à se procurer des données fiables sur l'art typographique depuis l'invention de l'imprimerie, des sites sont également nés qui sont consacrés au matériel ornemental des imprimeurs. Selon un bref état des lieux dressé par Rémi Jimenes en 2013 : « À la fin des années 1990, l'arrivée d'internet permet la mise en ligne de certains de ces projets. La base de lettres ornées parisiennes du XVI^e siècle, développée par Stephen Rawles à la bibliothèque universitaire de Glasgow, constitue à ce titre une œuvre pionnière⁷³. La bibliothèque universitaire de Barcelone administre depuis 1998 une importante base de marques typographiques européennes⁷⁴. Les bases Môriâne, Fleuron et Maguelone (les deux dernières fusionnant au sein du projet Passe-Partout) constituent de très utiles répertoires pour traquer les fausses adresses et les contrefaçons dans les publications du

⁷³ *University of Glasgow, Special Collections Department, Special Collections* : <http://special.lib.gla.ac.uk/initials/search>

⁷⁴ *Centre de Recursos per a l'Aprenentatge i la Investigació, Marques d'impressors* : <http://www.bib.ub.edu/fileadmin/impressors>

XVIII^e siècle. » Fort de ce constat, l'historien présente un nouveau projet, conduit par le Centre d'études supérieures de la Renaissance de Tours, qui est **BaTyR**⁷⁵ (Base de Typographie de la Renaissance) et qui cherche à constituer une base de données consacrée aux matériels d'imprimerie des XV^e-XVII^e siècles.⁷⁶

Les travaux centrés sur ces diverses bases de données existant déjà⁷⁷, nous ne ferons donc que revenir rapidement sur les trois derniers projets évoqués par Rémi Jimenes afin qu'ils puissent servir de support à une analyse postérieure des bases de données et des bouleversements épistémologiques qu'elles peuvent causer dans la discipline historique.

La base de données **Môriâne**⁷⁸ est créée en 1990 avec pour but de recenser le matériel ornemental utilisé dans les ouvrages imprimés du XVIII^e siècle. On peut donc parler de *répertoire typographique*⁷⁹. Un tel projet offre des perspectives nouvelles d'étude bibliographique sur la matérialité du livre et est destiné à faciliter le travail des chercheurs qui tenteraient d'identifier l'atelier d'imprimerie à l'origine d'une publication donnée, à partir de ses ornements typographiques, pensons aux bandeaux⁸⁰ ou aux vignettes⁸¹. Le moteur du site propose deux critères de recherche : ou bien le nom de l'imprimeur, ou bien une *description*, qui fonctionne donc par mots-clés.

Ces deux critères ont ceci de désavantageux qu'ils nécessitent d'une part une connaissance *a priori* du sujet de la recherche, soit une capacité à identifier le sujet de l'ornement, et d'autre part que la recherche ainsi formulée se fonde sur la description subjective et fermée des participants du projet, relevant donc d'un regard et d'un vocabulaire orientés notamment propres à une discipline déterminée.⁸²

Est proposée à l'issue de la recherche une liste des ornements possiblement recherchés : l'image, généralement associée à une réglette millimétrée, est accompagnée de sa fiche technique (**figure 2 et 3**).

⁷⁵ Le site est consultable à l'adresse suivante : http://www.bvh.univ-tours.fr/batyr/beta/Index_libraires.php

⁷⁶ Voir Rémi JIMENES. « La base de Typographie de la Renaissance (BaTyR) ». *Bulletin des bibliothèques de France*, n° 5, 2013, p. 18-22, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2013-05-0018-004.pdf>

⁷⁷ Voir notamment Magali SOULATGES, Claudette FORTUNY et Silvio CORSINI, « Quels nouveaux développements pour les bases de données d'ornements d'imprimerie ? », Synthèse de la journée d'étude internationale organisée le 29 mai 2012 par l'IRCL, juillet 2012, disponible sur le Web : <http://www.enssib.fr/quels-nouveaux-developpements-pour-les-bases-de-donnees-dornements-dimprimerie>

⁷⁸ Le site est consultable à l'adresse suivante : <http://promethee.philo.ulg.ac.be/moriane/ornSearch.aspx>

⁷⁹ Daniel DROIXHE, Stéphane STIENNON et Nadine VANWELKENHUYZEN, « Le projet Môriane », dans les Actes du colloque *Vers une nouvelle érudition : la numérisation au service de l'histoire du livre*, codirigé par VARRY, Dominique, BAUDRY, Guylaine et CHARON, Annie, à l'occasion des 12e entretiens du Centre Jacques Cartier, Lyon-Villeurbanne, Enssib, 6-7 décembre 1999, disponible sur le Web : <http://www.enssib.fr/bibliotheque-numerique/notices/1510-le-projet-moriane>

⁸⁰ Les bandeaux, *headpieces* en anglais, qualifient les ornements, souvent rectangulaires, placés en tête de chapitre.

⁸¹ Si le terme de *vignette* est parfois ambigu, il sert ici à nommer les ornements placés au titre ou en fin de chapitre, que l'on appelle parfois *culs-de-lampe* ; *tailpieces* en anglais. Bien souvent, ces ornements servent à combler une absence de texte sur la page, les imprimeurs ayant longtemps eu peur du vide. Notons qu'il existe enfin les *fleurons*, terme qui remplace parfois celui de *vignette* mais qui peut également désigner les caractères typographiques mobiles.

⁸² Une solution plus adaptée, le traitement automatique d'image, est évoquée dès 1999 lors du colloque des 6-7 décembre (voir note 69), mais n'est pas effective dans le cas de la base Môriâne. Nous reviendrons sur ce point en page 49.

Ornement n° 939

Imprimeur :

Description : Bandeau constitué d'une urne ou bonbonnière (au centre)posée sur un socle rayé verticalement. tout autour, figurent des branches feuillues, des rinceaux et des rameaux..

Provenance : fonds Weil

Nr Weil: 1122

Lieu réel:Hollande



L'ornement n° 939 apparaît dans les ouvrages suivants:

Numéro	Auteur	Titre	Lieu d'édition et/ou imprimeur	Parution	Page(s)	...
562	Levassor	Histoire du règne de Louis XIII	Amsterdam	1717		Détails

Figure 2. Exemple de bandeau sur la base Môriâne.

Ornement n° 447

Imprimeur : Collette (Colette)

Description : Un soleil avec des yeux, un nez, une bouche est entoure par deux branches (couronne) de laurier serrees par un noeud; la vignette est signee n.



L'ornement n° 447 apparaît dans les ouvrages suivants:

Numéro	Auteur	Titre	Lieu d'édition et/ou imprimeur	Parution	Page(s)	...
310	CARRONT	L'art de bien bâtir	Liège, Colette	1749	42	Détails

Figure 3. Exemple de vignette sur la base Môriâne.

La base de données étant exclusivement vouée aux ornements, elle ne s'intéresse donc aux ouvrages anciens que pour les détails typographiques qu'ils peuvent contenir. Le référencement des ornements, d'une grande précision, en fait, malgré les limites évoquées ci-dessus, un outil de travail précieux pour l'identification de provenance des ouvrages du XVIII^e siècle. On déplore cependant que le référencement de la base sur Internet soit des plus discrets, ce qui condamne l'outil à n'être connu et utilisé que par un nombre très restreint de spécialistes.

Fleuron⁸³, selon un principe identique, est une base de données d'ornements typographiques née d'une initiative de la Bibliothèque universitaire et cantonale de Lausanne. Bien qu'essentiellement centrée sur la Suisse romande, la base s'est étendue à quelques imprimeurs d'autres régions, notamment de Suisse italienne et de France, avec Lyon, Paris et Avignon.⁸⁴ La recherche peut cibler les imprimeurs, les livres ou les ornements eux-mêmes, avec dans ce dernier cas des critères tels que leurs dimensions, leur nature (bloc⁸⁵, taille-douce⁸⁶, composition⁸⁷, mixte) ou encore leur genre (fleuron-vignette, bandeau, lettrine, encadrement, pagination). La base, documentée avec grande précision, présente pour chaque imprimeur des notes relatives à sa biographie et à son travail typographique, des notes qui témoignent d'un riche travail de constitution de métadonnées. Si l'on additionne les 3 542 ornements en bloc, les 32 en taille-douce, les 3 671 en composition et les 28 ornements mixtes qui s'affichent à l'issue d'une recherche sans autre critère de sélection, on peut en déduire que Fleuron référence en mai 2016 pas moins de 7 272 ornements typographiques (**figures 4, 5, 6 et 7**).

❁ (113) ❁



Accueil>ChoixOrnements>ListeOrnements>LivresOrnement

FLEURON

1 livre trouvé pour l'ornement 000937-b

Auteur	Titre	Édition	Partie	Adresse	Année	Format	Pages
auteur inconnu	Psaumes de David, mis en vers, revus et approuvés	Nouvelle édition		A Lausanne, chez J. Pierre Heubach	1771	12°	prélim.

Figure 4. Exemple de pagination ornementée dans la base Fleuron et l'ouvrage dans lequel elle apparaît (nature : composition, genre : pagination). Notons que la base Fleuron référence cinquante-six ornements de pagination.

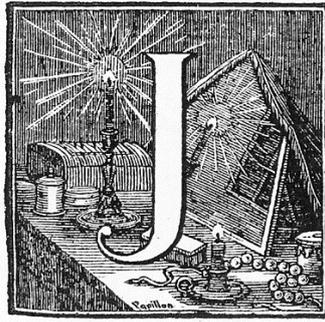
⁸³ Le site est consultable à l'adresse suivante : <http://dbserv1-bcu.unil.ch/ornements/scripts/>

⁸⁴ La page « Info » est consultable à l'adresse suivante : <http://dbserv1-bcu.unil.ch/ornements/scripts/Info.html>.

⁸⁵ Le *bloc* qualifie un ornement d'une pièce, gravé sur bois ou en fonte.

⁸⁶ La *taille-douce* qualifie un ornement gravé reconnaissable au cadre creux entourant le motif.

⁸⁷ La *composition* qualifie un assemblage de petites vignettes typographiques



Accueil>ChoixOrnements>ListeOrnements>VarianteOrnement>LivresOrnement

FLEURON

3 livres trouvés pour l'ornement 000350-j

Auteur	Titre	Édition	Partie	Adresse	Année	Format	Pages
Newton	Optice sive de reflexionibus, refractionibus, inflexionibus et coloribus lucis libri tres	Editio novissima		Lausannae & Genevae, sumpt. Marci-Michaelis Bousquet & Sociorum	1740	4°	p. II
Beausobre	Supplément à l'Histoire de la guerre des Hussites de Mr. Lenfant			A Lausanne & à Genève, chez Marc-Michel Bousquet & Comp.	1745	4°	prélim.
Voltaire	Pièces curieuses et intéressantes concernant la famille Calas qui ont été fournies par Mr. de Voltaire			A Lausanne, chez François Grasset & Comp.	1768	4°	1

Figure 5. Exemple de lettrine dans la base Fleuron et les ouvrages dans laquelle elle a été référencée (nature : bloc, genre : lettrine, lettre : J).



Accueil>ChoixOrnements>ListeOrnements>LivresOrnement

FLEURON

1 livre trouvé pour l'ornement or5585

Auteur	Titre	Édition	Partie	Adresse	Année	Format	Pages
Aristote	Aristotelis stagiritaee peripateticorum principis organon			Morgiis, excudebat Guillelmus Laimarius	1584	4°	Titre

Figure 6. Exemple d'encadrement dans la base Fleuron et l'ouvrage dans lequel apparaît (nature : bloc, genre : encadrement). Notons que Fleuron ne référence que six encadrements, les cinq autres étant de genre composé.

Accueil > Choix de l'imprimeur

FLEURON

Info

ORNEMENTS

IMPRIMEURS

LIVRES

Log in

RECHERCHE par IMPRIMEUR

Nom

Cliquez pour sélectionner

- Couvent
- Glessier, Jean-David - Neuchâtel
- Heilmann, Johann Christoph - Biel
- Heller, Jean-Baptiste - Yverdon
- Hugon, Louis - Lausanne
- Hutin, Emmanuel - Bern
- Hutin, Victor Emmanuel - Bern
- Jacob, Pierre - Yverdon
- Jacob, Philippe - Orbe
- Koehler, Niklaus - Basel
- Kupffer, Samuel - Bern
- Lamine, Charles - Yverdon
- Lambert, Michel (v.veuve) - Morges
- Lambert, Antoine (v.veuve) - Paris
- Le Blanc, Antoine-François (+veuve) - Paris
- Le Mercier, Pierre-Gilles (v.veuve) - Paris
- Le Prieux, Jean - Lausanne

7 critères

ORNEMENTS

LIVRES

IMPRIMEURS

Remarque

Source

Source: boward-heubach

Accueil > Choix de l'imprimeur

FLEURON

Info

ORNEMENTS

IMPRIMEURS

LIVRES

Log in

295 livres trouvés pour l'imprimeur Heubach, Jean-Pierre

Auteur	Titre	Édition	Partie	Adresse	Année	Format
Conde	Chambre de fil, à l'usage des religieux, ou règlement de la chambre de fil, à l'usage des religieux de la ville de Fribourg	Tome 3		A. Lausanne, chez J. P. Heubach	1787	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 4		A. Lausanne, chez J. P. Heubach	1787	12°
Sevan	Discours dans la cause d'une femme protestante au tribunal or pour enlever ses filés de la terre	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1787	8°
Stevens	L'opium, badre vertable, tete des papiers du pape Clément	Tome 4		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Volzke	Catalogue des livres de la Bibliothèque de la préfecture évangélique de sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
auteur inconnu	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 4		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Ochla Nabo	P. Ochi Nabo, Metaphysique, Bul IV	Tome 1		Lausanne, chez J. P. Heubach et Comp.	1788	12°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 2		A. Yverdon	1788	8°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 3		A. Yverdon	1788	8°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 4		A. Yverdon	1788	8°

Accueil > Choix de l'imprimeur

FLEURON

Info

ORNEMENTS

IMPRIMEURS

LIVRES

Log in

287 ornements trouvés pour l'imprimeur Heubach, Jean-Pierre

000279

001502

31 x 125

36 x 67

Source

Source: boward-heubach

Accueil > Choix de l'imprimeur

FLEURON

Info

ORNEMENTS

IMPRIMEURS

LIVRES

Log in

295 livres trouvés pour l'imprimeur Heubach, Jean-Pierre

Auteur	Titre	Édition	Partie	Adresse	Année	Format
Conde	Chambre de fil, à l'usage des religieux, ou règlement de la chambre de fil, à l'usage des religieux de la ville de Fribourg	Tome 3		A. Lausanne, chez J. P. Heubach	1787	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 4		A. Lausanne, chez J. P. Heubach	1787	12°
Sevan	Discours dans la cause d'une femme protestante au tribunal or pour enlever ses filés de la terre	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1787	8°
Stevens	L'opium, badre vertable, tete des papiers du pape Clément	Tome 4		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Volzke	Catalogue des livres de la Bibliothèque de la préfecture évangélique de sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
auteur inconnu	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 4		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Nardin	Le précurseur évangélique du sermo pour tous les branches et les principaux filés de France	Tome 5		A. Lausanne, chez J. P. Heubach et Comp.	1788	8°
Ochla Nabo	P. Ochi Nabo, Metaphysique, Bul IV	Tome 1		Lausanne, chez J. P. Heubach et Comp.	1788	12°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 2		A. Yverdon	1788	8°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 3		A. Yverdon	1788	8°
Quenry	Physiocratie, ou constitution naturelle du human	Tome 4		A. Yverdon	1788	8°

Figure 7. Exemple de recherche sur la base Fleuron. a) Choix de l'imprimeur ; b) Fiche de l'imprimeur ; c) Livres référencés de l'imprimeur ; d) Ornements référencés de l'imprimeur.

Sur un principe identique, la base **Maguelone**⁸⁸, créée en 2001, référence les ornements typographiques des imprimeurs du XVIII^e siècle. « Son objet vise à identifier des ateliers français et européens qui ont imprimé, sous de fausses adresses ou sans indication de lieu. »⁸⁹ Les types d'interrogations de la base de données sont les mêmes que pour la base Fleuron évoquée précédemment. Remarquons enfin que les deux bases Fleuron et Maguelone voient leur contenu moissonné par un projet plus large qui est la base Passe-Partout, dont la finalité est la même quoique les modalités de recherche divergent.

Le projet **Passe-Partout**⁹⁰ consiste en la constitution d'une banque internationale d'ornements d'imprimerie.⁹¹ Il a pour mission de coordonner les recherches menées par divers centres d'étude dans le domaine des ornements d'imprimerie.⁹² Les partenaires du projet sont invités à constituer des bases de données thématiques dont le contenu est ensuite centralisé dans la banque centrale Passe-Partout qu'il est possible d'interroger ou bien par mots-clés, telles les bases précédemment évoquées, ou bien à travers une comparaison informatique des formes, un projet novateur restreint à un nombre limité d'ornements. Le projet travaille en effet à la conception d'un programme de reconnaissance automatique des formes assistée par ordinateur, le logiciel TODAI⁹³, destiné à pallier les insuffisances du langage dans la description textuelle objective des formes visuelles. « Rien de plus subjectif que l'interprétation textuelle d'une image, aussi simple soit-elle ! », s'exclame à ce sujet Silvio Corsini.⁹⁴

Une journée d'étude européenne consacrée au livre ancien, des points de vue épistémologique, institutionnel et technique, s'est tenue le 29 mai 2012 où les participants ont été amenés à questionner la pertinence et l'obsolescence potentielle de la base Passe-Partout comme outil technique contemporain à disposition des chercheurs ; il en est ressorti une critique unanime relative à la complexité du processus technique de validation des ornements et à la lenteur du procédé de leur comparaison, bien que nul n'ait remis en cause l'utilité d'une telle base de données pour le travail d'analyse des livres anciens.⁹⁵

⁸⁸ La site est consultable à l'adresse suivante : <http://maguelone.enssib.fr/>

⁸⁹ La page « Présentation » est consultable à l'adresse suivante : <http://maguelone.enssib.fr/info.php>

⁹⁰ Le site est consultable à l'adresse suivante : <https://bcutodai.unil.ch/BCUTodai/app/Todai.do>

⁹¹ Silvio CORSINI, « Passe-Partout, banque internationale d'ornements d'imprimerie », *Bulletin des bibliothèques de France*, n°5, 2001, p.73-79, disponible à l'adresse suivante : <http://bbf.enssib.fr/consulter/bbf-2001-05-0073-010>

⁹² La page « Introduction » est consultable à l'adresse suivante : <https://bcutodai.unil.ch/BCUTodai/app/todaiGetIntro.do?uri=todaiInfo&page=todaiIntro.html>

⁹³ Le projet TODAI (*Typographical Ornaments Database and Indexation*) a été développé dans le cadre du fonds de recherche UNIL-EPFL.

⁹⁴ Silvio CORSINI, « Passe-Partout, banque internationale d'ornements d'imprimerie »..., p. 74

⁹⁵ Magali SOULATGES, « Les bases de données d'ornements typographiques du XVIII^e siècle », *Bulletin des bibliothèques de France (BBF)*, n° 6, 2012, p. 63-64, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2012-06-0063-005>

Projets linguistiques

Nous ne ferons que constater la constitution progressive de fonds scripturaux numériques ayant pour caractéristique leur rattachement à une aire géographique ou linguistique commune. C'est en particulier le cas du monde francophone : nous avons déjà évoqué le partenariat numérique qui unit la BnF et la BAnQ ainsi que le Réseau francophone numérique né de l'Organisation internationale de la Francophonie et qui rassemble vingt-cinq institutions documentaires.

Ces projets peuvent être le fait d'institutions patrimoniales : ils servent alors à la constitution de véritables *collections numériques*, comme nous l'avons vu. D'autres sont le fait de chercheurs en histoire et en informatique qui s'entendent avec pour objectif d'user du Net pour mettre à la disposition du monde de la recherche des bases de données utiles à des disciplines aussi variées que l'histoire culturelle ou la bibliographie matérielle. Ils rendent compte de la nécessité qui peut exister d'organiser l'information numérique afin de ne pas offrir en vrac des fonds si conséquents d'ouvrages numérisés qu'ils en deviennent insondables. Cette organisation, souple lorsqu'elle est le fait des moteurs de recherche, peut être préfigurée en amont par les acteurs de la numérisation dans le cadre de ce qui relève d'une *politique de numérisation*.

Une politique de numérisation ?

Il semble difficile de nier l'utilité des politiques de numérisation qui organisent *a priori* la constitution des bibliothèques numériques et de leurs fonds. Si elles diffèrent des « politiques d'acquisition » habituelles, au sens où elles concernent la mise en valeur et la mise à disposition d'ouvrages déjà acquis et déjà conservés dans des bibliothèques « en dur », elles se pensent néanmoins également en termes de budget et de public. Elles forcent à s'interroger de façon très pragmatique sur la pertinence des programmes de numérisations et sur leur réalisation.

Les différents projets abordés précédemment illustrent donc l'importance qu'il y a à penser la numérisation autrement que comme une entreprise industrielle de scan de masse : elle se doit de prendre en compte les préoccupations et les exigences de ses différents publics, et notamment celles des historiens qui doivent prendre part activement à la constitution de ces fonds documentaires de demain. Les structures patrimoniales ont ainsi élaboré de nouveaux services destinés à répondre aux besoins de ses usagers : la *numérisation à la demande* est notamment mise en place par certaines bibliothèques, pour que l'offre réponde à la demande

lorsque les politiques de numérisation en cours ne sont pas en mesure de satisfaire cette dernière. Le principe est intéressant pour le chercheur qui souhaite étudier sans contrainte un ouvrage conservé en fonds ancien et qui ne serait pas encore numérisé. Bien entendu, il faut cependant que le chercheur ait connaissance de l'existence de cet ouvrage, car une telle pratique ne permet pas de « flâner » parmi les fonds conservés pour y découvrir un objet de recherche non référencé, inconnu ou non encore identifié. Cela dit, les contraintes inhérentes à la conservation des livres anciens et leur fragilité forcent les bibliothèques à restreindre fortement l'accès à des ouvrages qui ne « sortent » qu'occasionnellement et sur demande, ce qui revient un peu au même.

C'est là peut-être le point fort de la numérisation de masse qui « dépoussière » les ouvrages conservés dans les fonds anciens et qui dorment, à défaut d'être examinés. Elle peut ainsi offrir aux chercheurs de nouveaux objets d'étude qui, sans leur mise à disposition sur le Net, seraient condamnés à ne pas exister. Il leur est d'ailleurs toujours possible, si besoin est, de chercher ensuite à consulter l'exemplaire matériel dans sa structure de conservation, dès lors que la démarche de numérisation s'ajoute, et non se soustrait, à la conservation de l'original papier. Enfin, un tel procédé de numérisation à la demande s'avère particulièrement intéressant pour le spécialiste de livres anciens qui souhaiterait comparer différents exemplaires d'un même ouvrage dont il ne serait, en théorie, numérisé qu'un seul exemplaire, si les politiques de numérisation faisaient l'objet d'une supervision globale efficace. Il est en effet acté qu'il est pour le moment préférable de ne numériser qu'un exemplaire d'un même ouvrage, afin d'éviter les doublons à l'heure où les bibliothèques numériques répondent en priorité à des exigences de variété et d'exhaustivité. Il est de toute manière difficile, pour le moment, de reprocher aux bibliothèques de rechercher l'efficacité, au fondement de cette course à la numérisation.

Restent cependant plusieurs problèmes : il faudrait d'une part que ce principe de numérisation à la demande se généralise ; d'autre part, un tel système requiert un catalogage précis et complet des fonds matériels existants. Enfin, un tel procédé engendre des coûts. À ce sujet, deux écoles existent : ou bien la bibliothèque numérise à ses frais le document demandé et la « demande » du lecteur numérique n'est qu'indicative, prise en compte car représentative d'un besoin scientifique – on peut imaginer un système d'évaluation de la demande qui serait confiée à une commission scientifique. Ou bien la numérisation est aux frais du demandeur, ce qui représente un frein certain pour des domaines de recherche peu subventionnés ou dont les représentants ne seraient rattachés à aucun laboratoire de recherche à même de prendre en charge les frais occasionnés. De même, cela pourrait freiner les travaux étudiants qui représentent pourtant une part non négligeable de la recherche contemporaine.

Considérons ainsi la politique de numérisation de la Bibliothèque municipale de Lyon, qu'elle détaille sur le site de sa bibliothèque numérique, Numelyo⁹⁶ : s'y distinguent la *numérisation par objectifs*, subventionnée, qui concerne des ouvrages ciblés du vaste fonds de la bibliothèque ; la *numérisation de masse*, ayant pour objet quelque 400 000 ouvrages en cours de numérisation par Google ; enfin la *numérisation à la demande* : « celle-ci concerne les documents à l'unité ou de petits ensembles qui n'ont pas encore fait l'objet de campagne de numérisation. Elle permet de fournir des images numériques aux chercheurs, auteurs, journalistes, éditeurs, scénaristes, établissements culturels, curieux... à des fins privées ou de publication. Ce service est payant et soumis à délai. » Une telle politique résume bien les enjeux qui sont ceux des institutions patrimoniales, à l'heure de la numérisation.

Un système de ce type a également été mis en place par la BnF sous la rubrique « Adoptez un document » et organisé par l'Association des amis de la Bibliothèque nationale de France, qui permet de choisir un ouvrage du catalogue de la bibliothèque et d'en financer la numérisation. La British Library propose un service identique à un tarif fixe de £25 : « *Adopt a Book* ». L'existence d'un tel service rend compte de l'importance d'un contact entre les infrastructures patrimoniales et leurs publics et pourrait pallier, durant un temps, l'impossible exhaustivité des campagnes de numérisation.

Ce bref regard porté sur l'offre numérique aujourd'hui disponible dans les méandres de l'Internet met en exergue l'importante quantité de ressources scripturales désormais dématérialisées, fruits de vastes campagnes de numérisation de masse accélérées par l'ambition démesurée de l'entreprise Google. Fleurissent ainsi des ensembles documentaires d'un nouvel ordre, les bibliothèques virtuelles, dont le contenu mouvant est de nature numérique.

⁹⁶ La page « Projet et partenaires » est consultable à l'adresse suivante : <http://numelyo.bm-lyon.fr/projet.php>

LA NATURE DE L'OFFRE

Il s'agit donc de définir les conditions d'existence du fichier numérique, à savoir la nature de l'offre numérique comme ensemble mouvant d'œuvres numérisées. Une telle caractérisation nous conduit à nous questionner sur la légitimité du fichier numérique à servir de fondement à un travail historique, c'est-à-dire à faire source. Cette légitimation, nous le verrons, passe en amont par un travail méthodique de constitution du document numérique qui dépasse de loin la seule numérisation.

À LA SOURCE DU TRAVAIL DE L'HISTORIEN, UN FICHER NUMÉRIQUE ?

Au regard de la science historique, le fichier numérique est une *source* d'un ordre nouveau. Deux caractéristiques du document-source s'avèrent problématiques dans le cas où l'ouvrage est numérisé : d'une part, l'*accessibilité* du document conditionne sa légitimité à « faire source » ; d'autre part, sa *stabilité* en détermine la valeur et la fiabilité. Accessibilité et stabilité, qui caractérisent avec une relative évidence le document papier, font au contraire défaut à la source numérique, y compris lorsqu'elle est directement issue de l'original papier. À ces deux critères, il faut ainsi opposer *indexation* et *conservation*, deux points incontournables d'une politique de numérisation efficace.

La valeur de la source : sa stabilité ? Lutte contre l'instabilité ontologique du fichier numérique

Un document qui « fait source », c'est-à-dire qu'il peut servir de matériau d'étude à l'historien, doit afficher une stabilité matérielle qui garantisse son authenticité et sa fiabilité. Il faut également, bien entendu, que l'historien ait un accès concret à la source, afin d'en évaluer la réalité physique – la source physique étant peut-être plus dure à falsifier qu'un fichier numérique dont l'origine reste éminemment obscure.

Le principal problème de la source numérique tient au fait que le fichier numérique se caractérise par une instabilité fondamentale due à son être immatériel.

L'instabilité est de plusieurs sortes : l'adresse url⁹⁷ qui définit le chemin du document dans les méandres du Net est ainsi rarement stable et, dès lors qu'elle est modifiée, devient caduque. Autre source d'instabilité : la labilité du document numérique, modifiable à souhait, anonymement, et dont il est difficile de mesurer l'intégrité. Enfin dans le cas des ouvrages numérisés, il est difficile sinon impossible de remonter précisément à la source du document, à moins que cette origine soit précisée. Et encore, l'information n'en sera pas moins douteuse. Philippe Rygiel, dans *Les sources de l'historien à l'heure d'Internet*, l'explique en ces termes : « la fiabilité des données accessibles depuis le réseau pose véritablement problème dans un certain nombre de cas. Deux raisons à cela. D'une part, nous ne disposons pas des repères physiques et institutionnels (sa présence dans un dépôt d'archives par exemple) qui nous permettent d'habitude d'identifier le document que nous utilisons. D'autre part, du fait de sa nature même, l'information numérique est assez facile à copier et à modifier de façon indécélable par l'utilisateur – et s'introduire dans un serveur afin d'en modifier les données est à la portée de nombre d'informaticiens et de bricoleurs. »⁹⁸

Il s'agit donc à l'heure des bibliothèques numériques de mettre en place des moyens efficaces d'étiquetage et de marquage de l'information afin d'en garantir la provenance, l'intégrité et donc la fiabilité. Cette lutte contre l'instabilité ontologique de l'information numérique passe par une standardisation des ressources, qui relève du travail de numérisation et du rôle des acteurs en jeu – qu'ils soient issus du monde des bibliothèques ou de celui de la recherche scientifique –, ainsi que d'un apprentissage de la part des usagers des modalités de fonctionnement de l'Internet. Une fois de plus, un travail préalable à la production des ouvrages numérisés doit s'accompagner d'un travail *a posteriori* sur les documents produits. Le travail sur les sources en histoire a toujours été essentiel pour qu'en soit garantie la fiabilité : il en est de même dans le monde numérique. Ce ne sont finalement que les modalités de ce travail qui ont évolué.

« On se plaît à rêver d'un bouton sur les ordinateurs qui permettrait de croiser la puissance d'un moteur de recherche avec la rigueur d'un « authenticomètre ». Hélas, l'instrument reste à inventer... »⁹⁹

Pour juger de l'authenticité du document numérique, existe-t-il une forme de « diplomatique » numérique ? Prenons la définition que donne Marie-Anne Chabin de ce domaine d'étude : la diplomatique est une « méthode d'analyse de l'authenticité, de la véracité, de la sincérité, de la fiabilité des documents

⁹⁷ L'URL (*uniform resource locator*) est l'adresse d'un site ou d'une page web. Voir le dictionnaire en ligne de Marie LEBERT, « Le dictionnaire du NEF », version du 27 octobre 2007, disponible à l'adresse suivante : <http://www.etudes-francaises.net/dico/>

⁹⁸ Philippe RYGIEL, « Les sources de l'historien à l'heure d'Internet », *Hypothèses*, 2004/1 (7), p. 351-352

⁹⁹ Marie-Anne CHABIN, « Peut-on parler de diplomatique numérique ? », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.)..., p. 116-117

d'archives [...] dont le but est la critique historique, et l'origine, le souci d'une meilleure exploration des sources de l'histoire »¹⁰⁰. Appliquée au numérique, une telle démarche de diplomatique demande la connaissance d'un langage et de pratiques nouvelles, relevant de l'informatique. Elles permettent par exemple, en ce qui concerne la fiabilité des documents, d'accéder aux différentes versions d'un même document, une nouvelle version n'en écrasant jamais véritablement une autre. Et à Marie-Anne Chabin de citer Isabelle Renard : « Heureusement, la technologie numérique possède une autre façon de garantir l'authenticité et la maintenance de cette authenticité : c'est d'enregistrer systématiquement les éléments d'identification et de modification des fichiers. Ce qui a inspiré à Isabelle Renard cette superbe sentence : « L'original est mort, vive la trace numérique ! »¹⁰¹ ».

L'informatique conserve en effet des traces numériques de presque tout ce qui peut avoir lieu sur l'écran de nos ordinateurs ; c'est à l'utilisateur ensuite de posséder les clés permettant de circuler dans les sinuosités du monde informatique. Pour ce qui est de l'authenticité et de la traçabilité concrète de l'ouvrage numérisé, le travail de diplomatique est bien moins évident : il conjugue finalement une part certaine de confiance dans l'institution œuvrant à la numérisation des fonds et une connaissance poussée de son domaine de recherche pour l'évaluation de la justesse des métadonnées qui servent à l'identification du document. Encore faut-il, cependant, que de tout document on ait trace.

La *searchability* du fichier, condition de son existence ?

Il est à craindre, à l'heure où la tendance générale est à la massification des flux, qu'un torrent informationnel non endigué puisse conduire à une altération du patrimoine écrit sur lequel travaillent les historiens.

Nous avons déjà noté quelques avantages et inconvénients relatifs à l'énorme mouvement de fonds causé par la numérisation : s'il est déplorable que les fonds numériques actuels soient encore loin de l'exhaustivité, ils n'en permettent pas moins la découverte ou redécouverte possible de documents oubliés sur les étagères d'un fonds anciens. Mais c'est avant tout l'accessibilité du livre numérisé qui pose problème.

La numérisation partielle, et partielle, laisse craindre que les chercheurs ne se rabattent en premier lieu sur ce qui est disponible en ligne. Un tel risque concerne

¹⁰⁰ *Id.*, *op. cit.*, p. 105-106

¹⁰¹ Isabelle RENARD, « L'original est mort, vive la trace numérique ! », *Journal du Net (JDN)*, article daté du 27 février 2003, consultable à l'adresse suivante : http://www.journaldunet.com/solutions/0302/030227_juridique.shtml, citée par Marie-Anne CHABIN dans *Vers un nouvel archiviste numérique ...*, p. 122

d'ailleurs le monde de la recherche en général : les sujets les plus étudiés ne vont-ils pas donner lieu à encore plus d'études ? Par ailleurs, elle concerne également le travail d'analyse en lui-même : le chercheur qui n'aurait accès qu'à la numérisation partielle d'un ouvrage va-t-il prendre la peine de récolter les passages manquants ?¹⁰² Un catalogage partiel, qui ne tiendrait pas compte à la fois des ressources imprimées et des ressources numériques, pourrait également être à l'origine de graves biais bibliographiques, tout comme l'historien lui-même qui, par manque de pugnacité ou de chance, peut constituer un corpus biaisé. C'est pourtant là un trait constitutif du travail de l'historien, qui toujours doit évaluer la représentativité des documents et des informations qu'il considère.

S'indigner contre les lacunes des bases de données comme cause d'une altération des corpus documentaires, c'est donc oublier que les catalogues papier, les fonds documentaires, tout comme le travail sur les sources de l'historien, donnent lieu à des biais de recherche inévitables qui confirment une évidence qui est celle de la non-neutralité de l'information.

Et l'information numérique, plus que toute autre, n'est pas neutre. Les moteurs de recherche, qui permettent l'exploration des bases de données numériques, présentent une caractéristique qui n'existait pas avec le travail sur les catalogues et les fonds papier : c'est que toute recherche numérique induit une hiérarchisation *a priori* des réponses proposées et donc un classement de valeur dont les critères sont hautement critiquables. C'est ce que François Rastier appelle la *searchability* du fichier numérique¹⁰³, sa « trouvabilité », à savoir la place à laquelle il va apparaître dans un moteur de recherche, place en théorie relative à sa pertinence, mais en pratique conséquence de son indexation, de sa provenance et de l'intérêt qu'elle suscite sur l'Internet. Cette « pertinence » supposée est ainsi liée au nombre de liens qui dans le World Wide Web renvoient à ce document, ainsi qu'à la « qualité » des sites qui y font référence – « qualité » calculée selon un principe voisin, à savoir son référencement et sa fréquentation. Nombreuses sont ainsi les critiques qui visent les algorithmes des moteurs de recherche, en ce qu'ils discriminent les fichiers relativement à leur « popularité ». Il existe ainsi des moyens de rendre plus « populaires » des œuvres numériques via leur indexation et leur publicité sur le Net. La critique perd cependant de sa pertinence dans le monde du livre ancien, où les ouvrages sont si peu consultés que leur présence sur l'Internet reste relativement neutre.

Une conséquence possible de ce classement de l'information relativement à sa popularité sur l'Internet consiste en ce que Claudio Paolucci a appelé une « tiers-

¹⁰² Remarquons une nouvelle fois que la situation n'est pas moins problématique dans un cas non-numérique, l'ouvrage conservé en fonds anciens étant bien souvent d'une accessibilité tout aussi limitée.

¹⁰³ François RASTIER, « La sémiotique des textes, du document à l'œuvre », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.)..., p. 55

mondialisation du savoir »¹⁰⁴, qu'il définit comme un processus « où le plus visible tend à devenir de plus en plus visible et le moins visible tend à disparaître »¹⁰⁵. L'analyse qu'il propose est parlante :

« Les algorithmes des moteurs, qui construisent les hiérarchies des sites résultant d'une recherche par mots clés, sont fondés sur le concept du *PageRank*. Le *PageRank* est normalement représenté par un chiffre variable qui exprime l'évaluation du site et sa place dans la liste de l'indexation. Le *PageRank* est un algorithme d'analyse qui donne un poids numérique à chaque élément d'un lien hypertextuel d'un ensemble de documents avec le but de quantifier son importance relative à l'intérieur de l'archive. »¹⁰⁶

« Le système définit la valeur du *PageRank* sur les bases de clics et de la « clickabilité » d'un site : les sites forts deviennent de plus en plus forts et les sites faibles deviennent de plus en plus faibles. »¹⁰⁷

« L'algorithme du *PageRank* établit cette forme de paramètre par rapport auquel celui qui dévie devient *minorité*. Cette forme de paramètre est obtenue sur la base de critères statistiques fondés sur une popularité *de facto* (le nombre de visites) et de droit (la « visitabilité » à partir d'autres sites). Les moteurs de recherche matérialisent ainsi au niveau du savoir un mécanisme que nos sociétés pratiquent à d'autres niveaux comme le *gossip* ou les télé-réalité. »¹⁰⁸

Pour finir, citons Marie-Anne Chabin qui cite Jérôme Bondu : « Vous savez déjà que ce qui n'est pas googlisé n'existe pas. Mais une nouvelle tendance apparaît. Être googlisé ne suffit pas ! ... L'information qui n'est pas en mouvement n'est pas captée par les flux, et donc n'existe pas. »¹⁰⁹

Dans le cas de l'étude historique du livre ancien, une telle discrimination informative pourrait être la cause d'une altération profonde de l'offre documentaire et donc de sa représentativité historique. Elle condamne ainsi des ouvrages mal référencés et peu usités à tomber dans l'oubli et conduit à évincer des corpus historiques des documents possiblement représentatifs. Certains historiens craignent qu'à terme un tel processus de sélection par la popularité conduise à une véritable *altération du passé*.¹¹⁰ Sur ce point, Claudio Paolucci

¹⁰⁴ Claudio PAOLUCCI, « Archive, patrimoine et mémoire. Un regard sémiotique sur la tiers-mondialisation du savoir à l'ère de la numérisation », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.)..., p. 75-105

¹⁰⁵ *Id. op. cit.*, p. 94

¹⁰⁶ *Id. op. cit.*, p. 92

¹⁰⁷ *Id. op. cit.*, p. 93

¹⁰⁸ *Id. op. cit.*, p. 94-95

¹⁰⁹ Jérôme BONDU, « Impact du web 2.0. », *Société Inter-Ligere / Conseil en Intelligence Economique Newsletter*, n°19, 2009 : ce blog, en mai 2016, n'est plus consultable ; cité par Marie-Anne CHABIN dans *Vers un nouvel archiviste numérique...*, p. 114

¹¹⁰ Pour nuancer un tel propos, rappelons que la « popularité » des ouvrages est un fait discriminatoire qui n'est pas né avec le numérique et qui a toujours mis en avant des ouvrages dits « de référence » dont la popularité

conclut : « Il est paradoxal que la valorisation du document, son être patrimonial, soit une fonction d'un système qui rend mémorable ce qui l'est déjà. »¹¹¹. Ajoutons qu'il est paradoxal, *a contrario*, qu'un tel système condamne à l'oubli ce qui devrait faire l'objet d'un travail de mémoire.

En amont : l'indexation et les métadonnées

Ce « travail de mémoire » qui permet de lutter contre les écueils des moteurs de recherche passe par une démarche d'indexation des documents numérisés. Cette indexation, qui consiste à définir le document dans un langage que l'on dira commun ou universel¹¹², est essentielle pour un bon référencement des ressources dans les bases de données, afin qu'elles soient ensuite repérables par les moteurs de recherche qui moissonnent ces fonds numériques. Cette indexation s'accompagne d'un travail méticuleux de catalogage et par la constitution de métadonnées informatiques destinées à décrire selon plusieurs approches la constitution de l'ouvrage. Pour ces dernières, employons la définition qu'en donne la BnF sur son site Internet : « Une métadonnée est un ensemble structuré d'informations décrivant une ressource quelconque. Les métadonnées ne décrivent pas nécessairement des documents électroniques. »¹¹³ Bien au contraire, il s'agit ici de considérer les métadonnées d'ouvrages papier qui sont numérisés : elles donnent ainsi sens à la version numérique qui reproduit l'original décrit. On peut différencier ici, comme le fait François Rastier, les intra-données des extra-données, les premières caractérisant les informations contenues dans l'ouvrage lui-même, les secondes des informations externes destinées à apporter au lecteur des précisions non contenues dans l'ouvrage. « La question des métadonnées peut alors se reformuler radicalement en distinguant les *intradonnées*, relevant de la teneur des textes, et les *extradonnées*, relevant de leur portée. »¹¹⁴.

s'autoalimente. Si le travail de l'historien consiste à rechercher l'objectivité, il n'en fait pas moins régulièrement appel à l'autorité de ses prédécesseurs pour appuyer des faits admis.

¹¹¹ Claudio PAOLUCCI, « Archive, patrimoine et mémoire... », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.)..., p. 95

¹¹² « L'indexation consiste à attribuer à un document un code de classement et une liste de mots-clés. Le code de classement est destiné à la recherche physique (sur les rayonnages) ou la recherche informatique. La liste de mots-clés est destinée à la recherche dans une base de données informatisée, appelée catalogue en ligne ou encore OPAC (online public access catalogue). Nombre de bibliothèques francophones utilisent RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié), un ensemble hiérarchisé de mots-clés géré par la Bibliothèque nationale de France (BnF). », définition proposée par Marie LEBERT dans « Le dictionnaire du NEF », version du 27 octobre 2007, disponible à l'adresse suivante : <http://www.etudes-francaises.net/dico/>

¹¹³ Voir la page « Document numérique et métadonnées » sur le site de la BnF, consultable à l'adresse suivante : http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.metadonnees_doc_numerique.html.

¹¹⁴ François RASTIER, « La sémiotique des textes, du document à l'œuvre », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.)..., p. 67

Les « intra-données » sont de nature multiple : les plus fondamentales sont des informations de fond que l'on retrouve aujourd'hui dans tout catalogue de bibliothèque – titre, auteur, édition et autres – ; d'autres moins établies mais essentielles dans le domaine du livre ancien concernent la matérialité-même de l'ouvrage numérisé, telles que la nature de la couverture, le relevé de signature¹¹⁵ détaillant avec précision la constitution du livre, la présence de marques de lecture (figure 8), le sens des lignes de chaîne¹¹⁶, ou encore les filigranes¹¹⁷ qui permettent de retracer la provenance du papier utilisé pour l'impression (figure 9). Le site de la BnF distingue ainsi plusieurs types de métadonnées : parmi elles, qui sont en ce sens des intra-données, retenons les *métadonnées descriptives*, qui s'attachent à rendre compte de l'identité bibliographique d'un ouvrage, et les *métadonnées de structure*.¹¹⁸

Ces innombrables détails qui demandent une connaissance approfondie de l'histoire du livre restent, quoi que l'on puisse en penser, des informations essentielles si la version numérisée de l'ouvrage est destinée à « remplacer » l'ouvrage papier.

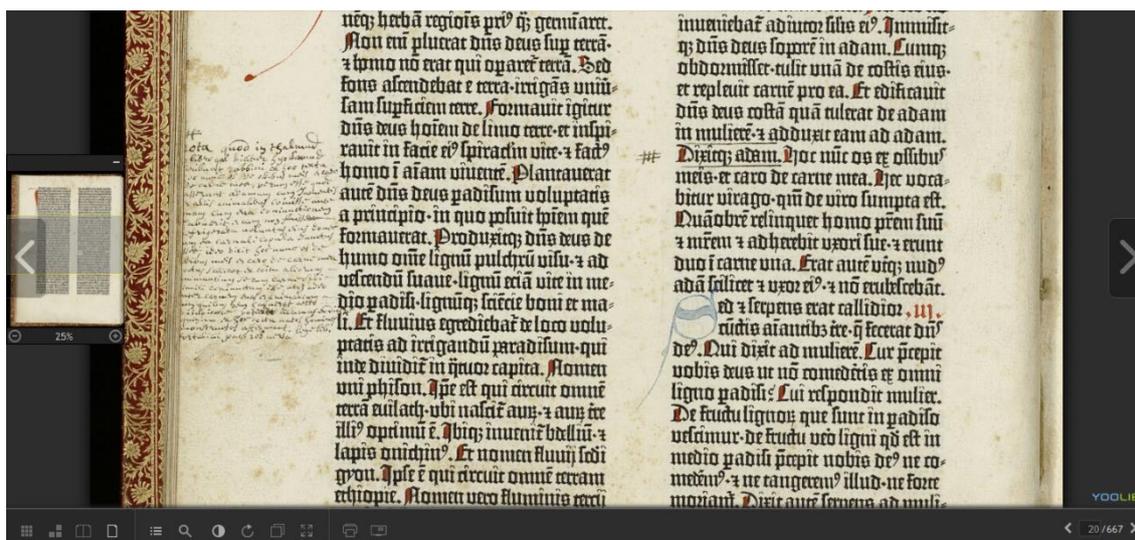


Figure 8. Marques de lecture sur la treizième page de la *Biblia latina* (1^{er} volume) publiée à Mayence vers 1455 par J. Gutenberg et J. Fust et conservé par la Bibliothèque Mazarine de Paris. La version numérisée est consultable sur le site Mazarinum à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/collection/7-biblia-latina/?n=2>.

¹¹⁵ Le relevé de signature, *formula* en anglais, est une formule à base de lettres et de nombres détaillant l'architecture et la composition d'un livre ancien. Pour plus de précision et des exemples de relevés, voir la page « Effectuer un relevé de signature » sur le site de bibliographie matérielle de Dominique VARRY, consultable à l'adresse suivante : <http://dominique-varry.enssib.fr/node/41>

¹¹⁶ Les lignes de chaînes, aussi appelées à tort *pontuseaux*, sont « les fils en métal qui forment le tamis avec lequel est fabriqué le papier. ». Ces derniers marquent le papier qui par transparence laisse voir des lignes régulières, verticales ou horizontales selon le nombre de pliages effectués à partir de la feuille initiale qui sert à la constitution des feuillets. Voir la page Wikipédia sur le papier vergé : https://fr.wikipedia.org/wiki/Papier_verg%C3%A9

¹¹⁷ « Un filigrane est un dessin qui apparaît sur certains papiers quand on le regarde par transparence ». Dans l'industrie papetière, il a servi longtemps de marque de fabrique, ce qui en fait un indice précieux pour les bibliographes qui cherchent à établir la provenance d'une édition papier. Voir la page Wikipédia sur le filigrane : <https://fr.wikipedia.org/wiki/Filigrane#Papeterie>

¹¹⁸ Voir la page « Document numérique et métadonnées » sur le site de la BnF, consultable à l'adresse suivante : http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.metadonnees_doc_numerique.html.



Figure 9. Image d'un filigrane et de lignes de chaînes visibles par transparence au travers du papier d'une page de garde. Référence : BRESSIEU, Maurice, *Grationopolitani Regii et ramei mathematicarum (...) metrices astronomicae libri quatuor. Haec maximam partem nova est rerum astronomicarum et geographicarum per plana sphericaque triangula dimensionis ratio, veterique impendio expeditior et compendiosior (...)*, Parisiis : apud Aegidium. Gorbinum, 1581.¹¹⁹

Ces informations standardisées, si elles peuvent être erronées, restent relativement objectives, dès lors que le travail d'analyse de l'ouvrage est précis et que les acteurs chargés d'indexer les fichiers possèdent une compétence poussée en la matière ; on sait malheureusement que ce n'est pas le cas, productivité et rentabilité obligent.

Les « extra-données » sont bien moins formelles et pour le moment bien peu travaillées. Elles sont destinées à enrichir le document par des informations absentes de son contenu mais qu'un travail d'analyse permet de constituer. Dans ce cadre, les outils informatiques apparaissent comme particulièrement pratiques pour ce qui est de contextualiser les œuvres : il est en effet possible de renvoyer, au moyen d'*hypertexte*, à d'autres sources d'information. L'*interactivité* permise par le média informatique offre ainsi des possibilités multiples et variées pour travailler sur le contexte informationnel d'un document donné. Bien loin de l'espace clos du livre papier, le livre numérisé peut se concevoir comme un espace ouvert, dans lequel les possibilités d'interactions ne sont limitées que par les

¹¹⁹ Le nom français de l'imprimeur est Gilles GOURBIN, qui travaillait à Paris entre 1551 et 1590. Ce livre ancien, possédé en plusieurs exemplaires par la BnF, fait partie du fonds pédagogique de l'Enssib.

moyens informatiques déployés : des liens hypertextuels renverraient alors ou bien à des bases biographiques, ou bien à des catalogues divers, ou bien encore à d'autres documents, aussi bien textuels que sonores et vidéos. Le numérique peut ainsi laisser rêver d'une version nouvelle et ô combien vivante des métadonnées, à la condition bien sûr qu'elles soient l'œuvre d'esprits éclairés à même de garantir la véracité et la relative objectivité des informations fournies.

Pour le moment, les détracteurs de Google s'échinent à repérer les nombreuses erreurs contenues dans le fonds numérique de Google Livres pour stigmatiser ce que Geoff Nunberg qualifie en 2009 de « désastre pour les chercheurs »¹²⁰.

Par ailleurs, si Alain Jacquesson s'exclame devant la capacité de certains serveurs de documents numériques à s'auto-cataloguer¹²¹, cette auto-indexation – qui peut fonctionner dans le cadre de documents nativement numériques dont le décodage est rendu aisé par l'harmonisation des normes de mise en page – semble bien loin d'être possible pour les livres anciens qui non seulement se prêtent mal à la reconnaissance automatique de caractères, mais qui présentent par ailleurs des mises en page bien différentes au cours des siècles qui ont suivi la naissance de l'imprimerie¹²². Aussi est-il peu probable que les machines en viennent à supplanter les bibliothécaires dans leur travail d'indexation et de constitution des métadonnées ; un travail qui, au contraire, doit répondre à des exigences croissantes de la part des chercheurs. Ce qui est une *médiatisation* du travail de constitution des métadonnées, c'est-à-dire l'impossibilité qui est celle des chercheurs de les collecter eux-mêmes en travaillant sur l'original papier du livre, pose cependant question relativement à la neutralité et à la fiabilité des informations transmises. Alain Jacquesson exprime ainsi cette crainte : « Encore ne faut-pas réifier les métadonnées, comme on l'a fait des « données » : elles sont aussi des résultats d'interprétation, et même la date du texte et le nom de l'auteur demandent à être examinées de manière critique. [...] Les métadonnées indexent certes les documents, mais concrétisent et anticipent des lectures et des interprétations. »¹²³

L'indexation rigoureuse de la documentation numérique apparaît donc comme une solution viable pour lutter contre la labilité de l'écrit numérisé, dès lors qu'elle est l'œuvre d'un personnel humain qualifié. Elle ne change en revanche rien à l'instabilité du document comme fichier numérique qu'il s'agit de conserver ; au contraire, les fichiers contenant les métadonnées relatives aux

¹²⁰ Geoff NUNBERG, « *Google's Book Search : A Disaster for Scholars* », *The Chronicle of Higher Education*, 31 août 2009, disponible sur le Web : <http://chronicle.com/article/Googles-Book-Search-A/48245/>, cité par Alain Jacquesson dans *Google livres et le futur des bibliothèques numériques...*, p. 74 et p. 166

¹²¹ Alain JACQUESSON, *Google livres et le futur des bibliothèques numériques...* p. 24

¹²² En témoigne pour exemple la page de titre visible sur la figure 19 d'une édition de la *Géométrie* de Descartes datée de 1683.

¹²³ *Id.*, *op. cit.*, p.64

données deviennent eux-mêmes des documents à conserver. Ainsi la BnF rappelle-t-elle qu'il existe une troisième catégorie de métadonnées, appelées *administratives*, pour gérer les droits d'accès, d'usage, de reproduction et de modification, pour préserver les informations techniques nécessaires à la lecture des fichiers et pour garantir l'intégrité des fichiers et le suivi de leurs éventuelles modifications¹²⁴. Summum des paradoxes actuels propres à la pérennisation des objets de l'écrit : ces métadonnées de conservation doivent elles-mêmes être conservées.

En aval : conservation et pérennisation

Si elle ne concerne que de loin l'historien lui-même qui ne peut que s'inquiéter de l'obsolescence de l'information proposée sur Internet, il importe cependant de considérer la *conservation* comme une problématique essentielle relative à la nature même de l'offre numérique.

Sans s'attarder sur des considérations techniques propres à toute entreprise de conservation, il faut du moins pointer du doigt la très courte espérance de vie des ouvrages numérisés, pour deux principales raisons : il existe dans le monde du numérique une obsolescence des langages et des supports, laquelle peut rendre illisible un document en un très court temps ; il existe également une obsolescence des machines destinées à lire ces supports et ces langages, ce qui oblige à une constante *migration* des documents d'un support à un autre, afin qu'ils demeurent lisibles pour les contemporains *via* leurs machines de lecture. La conservation de ces documents si peu pérennes doit être une priorité pour que la numérisation reste un processus légitime de préservation du patrimoine historique et qu'elle offre aux historiens une base fiable de travail de recherche. Quel paradoxe de voir l'historien travailler sur des supports d'une durée de vie ridiculement courte, censés au contraire véhiculer l'image d'objets d'une pérennité sans précédent, avec des supports aussi durables que la pierre ou le parchemin ! À côté du papier, qui est pourtant loin d'être le matériau le plus stable qui soit, la durée de vie d'un support comme le CD-ROM, qui dans de bonnes conditions de conservation dépasse rarement les dix ans, est tout bonnement insignifiante.

Aussi certains s'inquiètent-ils de voir le mot « numérisation » fleurir sur toutes les lèvres et remettre en cause la stabilité et la préciosité du livre imprimé ancien. Christian Lacombe, ancien bibliothécaire de la BnF, conservateur aux Archives des Jésuites au Canada et désormais expert en *consulting* relatif au livre

¹²⁴ Voir la page « Document numérique et métadonnées » sur le site de la BnF, consultable à l'adresse suivante : http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.metadonnees_doc_numerique.html.

rare comme placement financier pour les entreprises, se demande en 2009, dans *Documentation et bibliothèques*, si la numérisation est « sauvegarde ou mise à l'écart des documents ? ». Il le formule ainsi : « Ces inquiétudes nous amènent à se demander si la numérisation des collections imprimées des grandes bibliothèques européennes, anglo-saxonnes et autres ne constituerait pas une vaste entreprise de douce « mise à mort » du livre ou, tout au moins, une mise à l'écart, une relégation de l'imprimé vers un grand sommeil »¹²⁵. L'expression, parlante, sert d'ailleurs à titrer l'article et résume un questionnement important relatif à la multiplication de ces grands projets : la numérisation signe-t-elle le grand sommeil du livre imprimé ?

L'histoire rend compte d'une succession des supports de l'information qui marque l'obsolescence de ceux que la modernité relègue au placard. Il s'agit donc, afin de ne pas voir disparaître le patrimoine papier, de faire du livre imprimé un objet de conservation rigoureuse. Ainsi l'histoire des techniques pourra-t-elle se faire sans que les sociétés n'aient à craindre de perte. Citons Roger Chartier interrogé par Ivan Jablonka : « Aussi fondamental que soit l'accès à des textes sous une forme numérique, ce qui se trouve néanmoins renforcé par cette numérisation, c'est le rôle de conservation patrimoniale des formes successives que les textes ont eu pour leurs lecteurs successifs. La tâche de conservation, de catalogage et de consultation des textes dans les formes qui ont été celles de leur circulation devient une exigence absolument fondamentale, qui renforce la dimension patrimoniale et conservatoire des bibliothèques. »¹²⁶

Le fichier numérique cherche donc sa place en tant que source de travail pour l'historien. Sa légitimité à faire source est conditionnée par la mise en place, en amont et en aval, de procédés de traitements de l'information qui garantissent son intégrité, sa fiabilité et sa publicité. Aussi une politique efficace d'indexation peut-elle offrir à un document numérisé quelconque, indépendamment de la « popularité » qu'il pourra connaître ensuite, le moyen d'être accessible sur le Net et le droit d'être considéré comme une source historique viable. Cela est d'autant plus important qu'un ouvrage qui « dort », à l'abri du temps et des hommes, gagne en longévité autant qu'il perd en intérêt, car qu'est-ce qu'un livre sans lecteur ? Comme le souligne Roger Chartier, « Paul Ricœur a souvent souligné le fait qu'un monde de textes qui n'est pas saisi, approprié par un monde de lecteurs, n'est qu'un monde de textes possibles, inertes, sans véritable existence »¹²⁷. Le numérique, en cela, peut réveiller les livres qui dorment.

¹²⁵ Christian LACOMBE, « La numérisation ou le grand sommeil du livre imprimé », *Documentation et bibliothèques*, Montréal, Asted, vol. 55, n°1, 2009, p. 40

¹²⁶ Ivan JABLONKA, « Le livre, son passé, son avenir. Entretien avec Roger Chartier », Dossier « Le livre demain », *La vie des idées*, sept-oct 2008, consultable à l'adresse suivante : <http://www.laviedesidees.fr/Le-livre-son-passe-son-avenir.html>

¹²⁷ Roger CHARTIER, *Le livre en révolutions...*, p. 154

LE LIVRE NUMÉRISÉ

Les politiques rigoureuses de conservation tendant à protéger au maximum les œuvres anciennes des dommages du temps et des lecteurs, les fonds anciens font office de temples qui contiennent, hors de portée du public, les reliques du passé, consultables sous contrôle par quelques privilégiés. Toute la connaissance qu'ils rassemblent reste par là-même contenue, soumise en quelque sorte à l'inertie qui est celle de l'écrit dormant, tenu hors d'atteinte du regard qui par la lecture peut le rendre vivant. La numérisation est en cela une seconde naissance pour le livre ancien.

Quid de l'original ?

Mais livre dormant ne signifie pas livre inutile. Si la numérisation peut servir à la publicité de livres qui jusqu'alors ne touchaient aucun public – historiens-chercheurs compris – elle ne peut pourtant remplacer le livre papier lui-même, notamment parce que ce dernier reste l'« original » et que la *conservation*, mission dévolue aux structures patrimoniales, consiste justement en la préservation des originaux. Il est essentiel de garder en tête, lorsque l'on considère un document non nativement numérique, qu'il n'est que la reproduction d'une œuvre à conserver, et que cet original possède des caractéristiques mémorables. Il faut donc distinguer le document patrimonial papier de sa version numérique comme l'on distingue l'*original* d'une *reproduction*.

Le fichier numérique est symptomatique d'une « révolution des supports » contemporains, qu'il s'agisse du livre numérique ou du livre numérisé, puisque ce dernier consiste en la présentation d'un support ancien, le livre imprimé, sur un support nouveau : l'écran. Jean Sarzana et Alain Pierrot définissent ainsi la numérisation : il s'agit d'un « procédé de transformation d'un signal analogique en une suite de valeurs numériques, c'est-à-dire la conversion d'un langage dans un autre, avec de notre point de vue tous les biais qu'implique la traduction »¹²⁸. La notion de *traduction* employée par les deux auteurs est intéressante pour qualifier un processus qui consiste bien à produire à partir d'un original une copie qui se veut, malgré un langage différent, au plus près du texte original. Cependant, pour ce qui est du mode image qui est un *scan*, la numérisation traduit moins le texte que l'image même de la page sur laquelle figure ce texte ; la traduction dépasse alors le cadre du seul matériau langagier.

¹²⁸ Jean SARZANA et Alain PIERROT, *Impressions numériques...*, p. 32

Considérons cette remarque de Frédéric Barbier : « L'imprimé ne véhicule pas davantage un message qui se réduit au seul texte. Le lecteur lira le texte, en principe, mais la forme et la représentation de l'objet lui apporteront autre chose : le choix du format, de la couverture, de la mise en page, du contenu et de l'organisation du paratexte apporte autre chose du point de vue sémiologique, de même que l'environnement du livre »¹²⁹. Les ouvrages numérisés – dans le cas où la numérisation est de qualité – contreviennent à ces catégories sans pourtant s'en détacher franchement : l'imprimé n'est pas réduit au seul texte, le contenu reste le même quoique le rendu puisse différer en taille et en qualité visuelle, l'environnement du livre n'est plus le même mais le scan, comme les procédés photographiques, cherche bien à rendre compte de l'aspect matériel et visuel du livre.

Pour cette raison, le rapport Zelnik-Toubon-Cerutti, *Création et Internet*¹³⁰, introduit un terme que reprend Marc Tessier dans son *Rapport sur la numérisation du patrimoine écrit* : la notion de « livre homothétique », terme qui sert ainsi à qualifier le livre numérisé¹³¹. Du grec *homo-thesis*, l'*homothétie* signifie « même position » et sert à traduire la correspondance entre deux figures de même forme et de même orientation¹³². Patrick Zelnik, Jacques Toubon et Guillaume Cerutti expliquent ainsi l'usage du terme : « c'est-à-dire reproduisant à l'identique l'information contenue dans le livre imprimé, tout en admettant certains enrichissements comme un moteur de recherche interne, par exemple »¹³³. L'usage de ce terme rend compte du rapport étroit de correspondance qui existe entre l'œuvre papier et sa reproduction numérique. Mais sur cet usage technique, certains sont sévères : « L'édition homothétique rassure et valorise à la fois, on comprend qu'elle suscite un réel engouement. Elle nous apparaît néanmoins sans portée véritable. D'abord, elle met l'accent sur la technique beaucoup plus que sur l'œuvre elle-même, elle privilégie l'outil au détriment de ce à quoi on l'applique. »¹³⁴.

La numérisation n'est donc qu'une révolution partielle. Le mode d'affichage de l'écrit change avec son support, mais non l'écrit en lui-même. La reproduction n'a pas pour finalité de supplanter l'original : elle le médiatise et, par là-même, le pérennise ; c'est à l'historien ensuite de ne jamais prendre la copie pour l'original. Aussi les historiens et certains bibliothécaires mettent-ils en garde la communauté scientifique et bibliothéconomique quant à l'écueil qui consisterait en la destruction des ouvrages originaux suite à leur numérisation. Roger Chartier, dans

¹²⁹ Frédéric BARBIER dans *La bataille de l'imprimé, à l'ère du papier électronique*, Éric LE RAY et Jean-Paul LAFRANCE (dir.), Montréal, Presses de l'Université de Montréal, 2008, p. 19

¹³⁰ Patrick ZELNIK, Jacques TOUBON et Guillaume CERUTTI, *Création et Internet*, remis en janvier 2010 au Ministre de la Culture et de la Communication, 2010, disponible sur le Web : <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/104000006.pdf>

¹³¹ Cette distinction est évoquée par Jean SARZANA et Alain PIERROT dans *Impressions numériques...*, p. 202

¹³² Voir la page Wikipédia sur l'homothétie : <https://fr.wikipedia.org/wiki/Homoth%C3%A9tie>

¹³³ Patrick ZELNIK, Jacques TOUBON et Guillaume CERUTTI, *Création et Internet...*, p. 7

¹³⁴ Jean SARZANA et Alain PIERROT, *Impressions numériques...*, p. 202

L'écrit et l'écran, une révolution en marche, l'exprime en ces termes : « Le bonheur extravagant suscité par la bibliothèque universelle pourrait devenir une impuissante amertume s'il devait se traduire par la relégation, ou pire, la destruction des objets imprimés qui ont nourri au fil du temps les pensées et les rêves de ceux et de celles qui les ont lus. »¹³⁵ Et Christian Lacombe, qui le cite, d'ajouter : « Malheureusement, cette vision antagoniste de conservation/communication, loin d'être tombée en désuétude, est au contraire réactivée depuis que les éditions originales des documents imprimés ne sont plus communiquées en premier aux lecteurs, si une version numérique est disponible. Il est également facile d'imaginer que, pour des raisons de coût, un document imprimé quelque peu « fatigué » ne sera pas envoyé en priorité à la conservation s'il est numérisé, sachant que sa consultation est devenue possible à l'écran. [...] La numérisation d'un livre est un miraculeux moyen de sauvegarde de son contenu. Si, par contre, la version numérique devient le moyen privilégié de sa transmission, alors il nous faut rappeler que la lecture à l'écran impose une logique analytique et encyclopédique, « la violence faite aux textes, donnés à lire dans des formes qui ne sont plus celles où les rencontrèrent les lecteurs du passé »¹³⁶. En séparant le livre de sa forme originale issue du codex, la lecture à l'écran rend simplement impossible son interprétation et surtout sa signification historique. »¹³⁷.

De l'original, il ne faut donc pas se passer. La numérisation est un moyen technique de communication de l'information sans précédent mais qui n'offre qu'une reproduction d'une œuvre papier. Cette œuvre papier est à conserver au même titre que sa version numérique, parce que la consultation à l'écran ne permet pas l'obtention de toutes les informations que peuvent rechercher les historiens relativement à un livre, d'une part, mais également parce que la valeur patrimoniale et donc historique de l'œuvre réside encore dans sa forme originale. La reproduction numérisée ne remplace pas l'original papier, elle le complète, elle le double. De là le titre d'un article de Françoise Benhamou : « Le livre et son double. »¹³⁸, qui résume à merveille, non leur rapport de correspondance, mais leur rapport d'homothétie.

¹³⁵ Roger CHARTIER, « L'écrit et l'écran, une révolution en marche », *Le Monde*, 13 octobre 2007, p. 24

¹³⁶ *Ibid.*

¹³⁷ Christian LACOMBE, « La numérisation ou le grand sommeil du livre imprimé », p. 41.

¹³⁸ Françoise BENHAMOU, « Le livre et son double, réflexion sur le livre numérique », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 90-102

Le processus de numérisation

Le processus de numérisation consiste, sur les modèles maintenant courants de reproduction photographique, à constituer une image dématérialisée de la réalité physique d'une feuille de papier. À l'aide d'un scanner ou d'un appareil photographique, un personnel humain capture ainsi, page par page, l'image d'un ouvrage matériel. Le support est ainsi converti en données numériques, soit « une suite de caractères et de nombres qui représentent des informations » et « que des supports informatiques ou d'électronique numérique pourront traiter. »¹³⁹

Si l'on utilise encore parfois des appareils photographiques dans le cas où le document à dématérialiser est rare et précieux, l'usage des scanners – de plusieurs types – s'est généralisé pour les livres anciens. Initialement et pendant longtemps, les acteurs de la numérisation ont utilisé des scanners dits « à plat », peu satisfaisants au regard des contraintes à exercer sur le livre pour que le rendu de la page soit proportionné et uniforme. Ces scanners nécessitent en effet l'ouverture complète du livre pour que sa surface soit plane, occasionnant des dégâts préjudiciables sur la reliure et la jonction entre le dos et les pages de l'ouvrage. Depuis les années 1990, des scanners de plus en plus performants, associés à des logiciels de traitement d'images, permettent de numériser les pages d'un livre sans qu'il soit complètement ouvert, la capture d'image étant suivie d'une correction de la déformation visuelle due à la courbure des pages. Quelques spécialistes nous l'expliquent : « La numérisation de livres s'effectue en général par prise de vue par le dessus. L'ouvrage est déposé tel quel sur un plateau équilibrant, ou maintenu ouvert à 90° si son état le nécessite. Les déformations des images peuvent ensuite être traitées par logiciel, selon diverses heuristiques : rectification des bords, des lignes de texte, inclinométrie. »¹⁴⁰. Au début des années 2000, le tourne-pages automatique révolutionne l'industrie de la numérisation en mécanisant un geste jusque-là effectué par un personnel dédié. Cette « machinisation » du processus induit bien sûr une proportion relativement plus importante d'erreurs à corriger *a posteriori*, comme les pages manquantes ou pliées. Ajoutons également que Google a développé une machine plus spécialisée encore intégrant un processus de numérisation tenu aussi secret que possible, mais dont le fonctionnement est connu depuis que l'entreprise en a déposé un brevet, le brevet US 7508978, qui rend compte de l'utilisation conjointe d'un scanner et de caméras infrarouges destinées à calibrer automatiquement les images numérisées. Le système, d'une efficacité accrue, répond par ailleurs aux exigences de conservation puisqu'il n'exerce aucune pression mécanique sur les ouvrages traités¹⁴¹.

¹³⁹ Voir la page Wikipédia sur la numérisation : <https://fr.wikipedia.org/wiki/Num%C3%A9risation>

¹⁴⁰ Rodrigo ALMEIDA, Pierre CUBAUD, Jérôme DUPIRE et Alexandre TOPOL, « Interactions et métadonnées riches pour les bibliothèques numérisées », dans « Visualisation pour les bibliothèques numériques », Jean-Daniel FEKETE et Eric LECOLINET (dir.), *Document numérique*, vol. 9, 2006/2, p. 93

¹⁴¹ Alain JACQUESSON, *Google livres et le futur des bibliothèques numériques...* p. 57-58

Il s'agit là du déroulé global de la numérisation ; nous ne nous attarderons pas sur ses modalités techniques comme les appareils utilisés ou les coûts de réalisation. Pour l'historien du livre, le processus en lui-même n'importe que s'il est à même de conjuguer efficacité de traitement, c'est-à-dire la qualité des images obtenues, et préservation des ouvrages numérisés. Il s'agit là, une fois de plus, de faire confiance aux acteurs de la numérisation, du bibliothécaire qui sélectionne les livres à numériser – et qui est donc en mesure de protéger les livres trop fragiles, par exemple – à l'opérateur qui supervise les chaînes de numérisation. L'industrialisation numérique ne doit pas gagner en efficacité au détriment de la qualité du travail effectué.

Prenons pour finir l'exemple de la BnF qui détaille sur son site Internet le processus de numérisation adopté en fonction des ouvrages, un processus qui rend compte de la grande adaptabilité que requiert un tel procédé : « Les techniques de numérisation des documents dépendent de leurs conditions physiques : taille, volume, fragilité, etc. Les documents de la BnF sont numérisés sur une grande variété de machines, dans des ateliers internes ou par des prestataires extérieurs. »¹⁴²

Mode texte / Mode image

Cette « qualité » de la numérisation s'évalue selon différents critères. Il est d'abord à distinguer deux modes de la numérisation, distincts mais potentiellement complémentaires : le *mode texte* et le *mode image*. Le processus de numérisation simple, présenté précédemment, permet la constitution d'un fichier en mode image, c'est-à-dire qu'il scanne la page comme si elle n'était qu'un objet plane que l'on peut ensuite afficher sur écran. Le document produit est alors un simple *fac-similé* de l'original, une reproduction numérique.

Un traitement informatique de ces images peut en revanche offrir de nouvelles potentialités d'exploitation du document. Ainsi existe-t-il des logiciels qui localisent les caractères textuels, tentent de les reconnaître et reconstituent ainsi le texte présent sur la page dont ils présentent ensuite une traduction sous forme nativement numérique. On parle de Reconnaissance Optique de Caractère (ROC), ou *Optical Character Recognition (OCR)* en anglais, qui a donné naissance au terme français admis d'« OCRisation », ou « océrisation ». Si pour les typographies récentes, le taux de reconnaissance automatique de caractère est

¹⁴² Voir la page « Formats et techniques de numérisation en mode image » du site de la BnF consultable à l'adresse suivante : http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.outils_numerisation_mode_image.html

excellent, il n'en est pas de même avec les typographies plus anciennes dont la reconnaissance est bien plus problématique.

Il en est ainsi de tous les ouvrages anciens, et ce pour plusieurs raisons : tout d'abord, le vieillissement du papier est cause de « bruit », c'est-à-dire de taches que le logiciel d'océrisation voit comme un caractère et qu'il tente de traduire, perturbant ainsi le texte final (**figure 13**). Ensuite, les typographies qui naissent avec l'imprimerie de Gutenberg au XV^e ont fortement évolué au cours des siècles : certains caractères ne sont plus utilisés aujourd'hui et ne figurent donc plus dans les bases de données des logiciels (**figure 13**), ou bien leur utilisation n'était pas la même – pensons à l'accentuation des majuscules pour laquelle les imprimeurs devaient user de deux types différents : la majuscule, puis l'accent (**figure 10**). De même, les lettrines ornant les débuts de chapitre, non seulement sont inconnues au logiciel d'océrisation, mais perturbent par ailleurs beaucoup la transcription informatisée puisque s'étalant de par leur hauteur sur plusieurs lignes (**figure 10**). Il est à noter enfin que, avant que ne naissent l'informatique et les traitements de texte, l'imprimerie fonctionnait à partir de caractères mobiles qui pouvaient s'abîmer – notamment pour les gros caractères d'abord sculptés dans du bois et pour les ornements – ou encore être utilisés à l'envers, volontairement ou non (**figure 11**)¹⁴³. Des inégalités d'encre nous rappellent que l'imprimerie a longtemps été une technique artisanale (**figure 12**). La relative irrégularité du matériau textuel présent dans les ouvrages anciens et leur état de conservation freinent donc l'efficacité des logiciels d'océrisation ; Alain Jacquesson évoque une fiabilité de 60 à 70 %¹⁴⁴.

Citons ici le propos acerbe mais si humoristique de Julien Xavier Polastron à ce sujet :

« aucune de ces dispendieuses reproductions d'éditions admirables au demeurant n'est ni si utile au spécialiste ni si facile à déterrer : c'est une simple photographie de page qui se dévoile à un train de vieille tortue sous les yeux exclusifs de l'obstiné s'il connaît déjà le site de l'archive, et ce sans indexation ni citation possible sauf à copier à la main ou imprimer le livre au format A4 page après page pour le passer à la moulinette OCR, qui va trébucher sur chaque empattement un peu écrasé des Bodoni et compter soigneusement chaque chiure de mouche. Cet anachronisme bon marché s'appelle le « mode image ». »¹⁴⁵

¹⁴³ Nous verrons à ce sujet, page 89, les solutions proposées par Jacques André et Rémi Jimenes.

¹⁴⁴ Alain JACQUESSON, *Google livres et le futur des bibliothèques numériques...* p. 59

¹⁴⁵ Lucien Xavier POLASTRON, *La Grande Numérisation : y a-t-il une pensée après le papier ?*, Paris : Denoël, 2006, p. 16

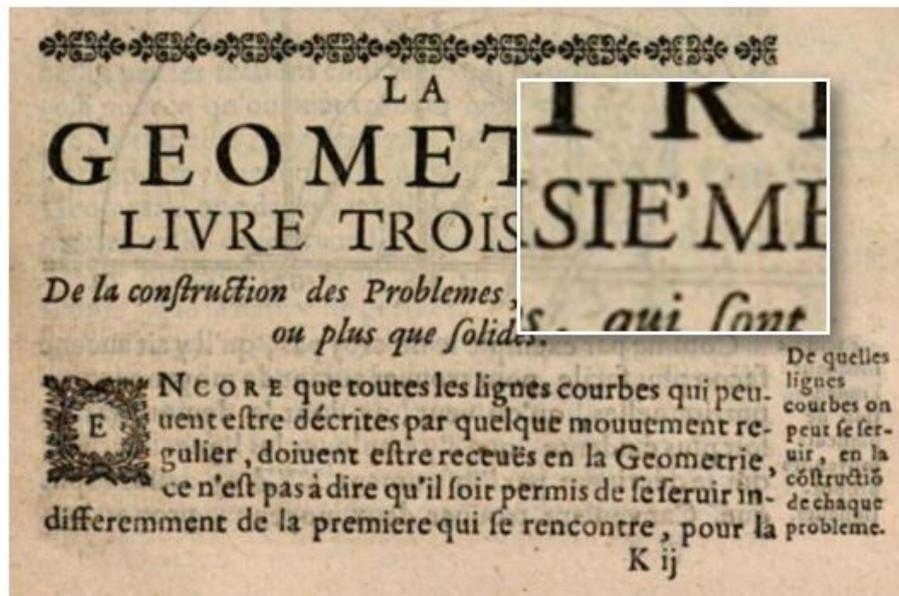


Figure 10. Zoom sur l'utilisation de deux types distincts, une lettre E grande capitale et une apostrophe, pour figurer un « E accent aigu ». Référence : *La Géométrie de René Descartes*, Paris, chez Charles Angot, 1664, p. 75 (Bibliothèque nationale d'Autriche).

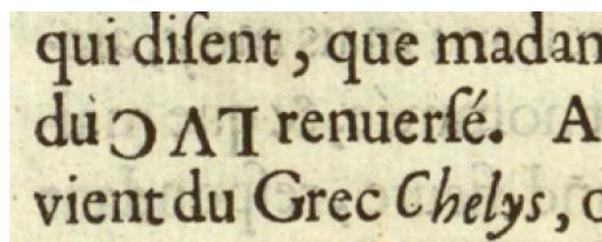


Figure 11. Exemple de lettres volontairement renversées. Photo empruntée à Jacques André et Rémi Jimenes¹⁴⁶ et tirée des *Discours non plus mélancoliques que divers*, Poitiers, Marnef, 1557, p. 97 (Poitiers, BM).

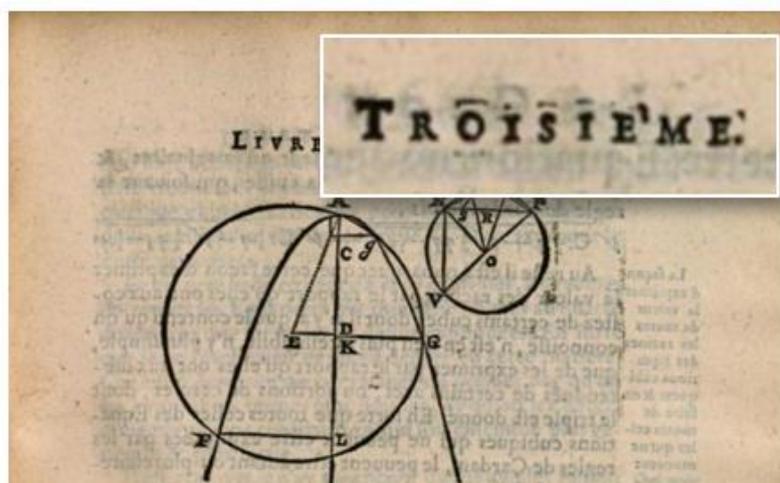


Figure 12. Zoom sur des marques d'épaulement laissées par un encrage trop important des caractères mobiles. Référence : *La Géométrie de René Descartes*, Paris, chez Charles Angot, 1664, p. 105 (Bibliothèque nationale d'Autriche).

¹⁴⁶ ANDRE, Jacques et JIMENES, Rémi, « Transcription et codage des imprimés de la Renaissance : réflexions pour un inventaire de caractères anciens », dans « Gestion informatisée des écritures anciennes », BENEVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 133.

...sont que
des **fi**ctions de mon Esprit. Qu'est-ce donc qui
pouir
estre estimé veritable **f**peut-estre rien autre
chose, sinon
qu'il n'y a rienaji monde de certain.

Mais que [ça y-je s'il n'y a point quelque autre
chose
différente de celles que ie viens de juger
incertaines, de
laquelle on ne puisse auoir le moindre doute ?
N'y a t'il
point quelque **Dipu** ou quelque autre
puissance, qui me
met en r eiprit ces pensees? Cela n'41 pas
necessaire ; car

Le texte affiché peut comporter un certain nombre d'erreurs.
En effet, le mode texte de ce document a été généré de façon
automatique par un programme de reconnaissance optique de
caractères (OCR). Le taux de reconnaissance estimé pour ce
document est de **93.49** %.

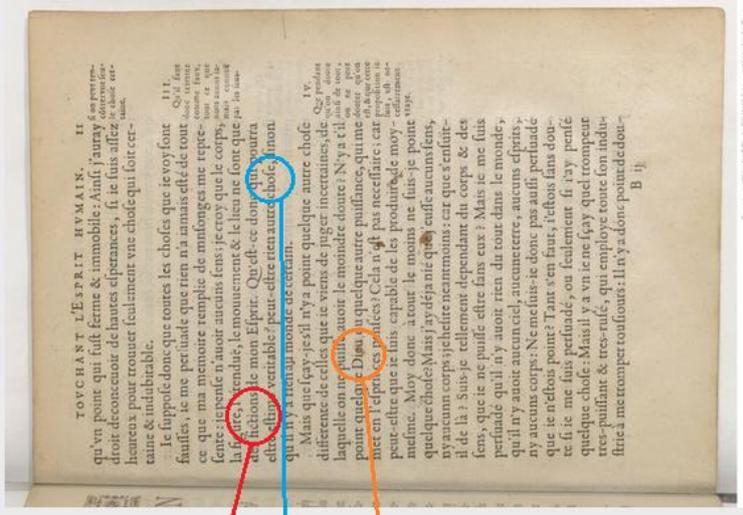


Figure 13. Onzième page du texte des *Méditations métaphysiques* de René Descartes, 3^e édition, édité chez M. Bobin et N. Le Gras à Paris en 1673, conservé à la Bibliothèque nationale de France sous la cote R-3503. La version en ligne est disponible sur Gallica à l'adresse suivante : <http://gallica.bnf.fr/ark:/12148/bpt6k9621169g.r=R-3503>.

On remarquera ainsi que les taches perturbent le bon fonctionnement du logiciel d'ocrisation (« Dieu » devient « Dipu ») et que certains caractères d'imprimerie, comme le « ct » ligaturé, ne sont pas reconnus (« fictions » devient fi&ions »). Notons cependant qu'à l'inverse, le logiciel traite très correctement le caractère « s » de l'époque.

Avant de conclure, il est un dernier point à aborder : nombre d'acteurs du livre, qu'ils soient issus du monde de la recherche ou du monde des bibliothèques, citent le Project Gutenberg de Mickael Hart comme le point de départ d'une histoire de la numérisation, en 1971. S'il s'agit bien là du premier projet de bibliothèque numérique de l'histoire, qui consistait, selon leur slogan, à briser « les barrières de l'ignorance et de l'illettrisme » en mettant en ligne des livres d'abord retranscrits à la main sur ordinateur, soulignons que les plus de 42 000 *e-books*¹⁴⁷ aujourd'hui proposés ne sont accompagnés d'aucune image de l'ouvrage original. La provenance et l'authenticité des ouvrages perdent d'autant plus en crédibilité que les métadonnées manquent parfois pour identifier l'exemplaire que l'on dira « tapuscrit » plutôt que « numérisé ». Impossible donc de s'assurer de la fiabilité du document numérique qui n'acquiert par là-même aucune légitimité à faire source.

Comment appréhender, alors, ces deux modes de traitement de l'image numérisée ? Il semble qu'ils ne concernent *a priori* pas le même public : pour le chercheur qui s'intéresse au contenu du livre, à la seule information qu'il véhicule, le mode texte accompagne sinon remplace le travail de lecture de l'œuvre. Il permet également une indexation automatique des mots du texte, dont le contenu devient « moissonnable » par les moteurs de recherche et permet une recherche dite « lexicométrique », sur laquelle nous reviendrons. Un tel outil ouvre des perspectives sans précédent en termes d'analyse textuelle et de recherche bibliographique approfondie. En revanche, le mode texte – même si la performance des logiciels d'océrisation s'améliorait – apparaît peu utile à l'historien du livre qui s'intéresse moins au texte qu'à l'objet-livre lui-même et à sa réalité matérielle. Le mode image permet alors une analyse approfondie quoique non exhaustive de l'ouvrage papier dans son unicité, l'épaisseur en moins.

Dans tous les cas, la correspondance entre la version image et la version texte d'un ouvrage sert la crédibilité du travail de numérisation et le travail de l'historien : complémentaires, mode texte et mode image sont gages de fiabilité et de qualité lorsqu'ils permettent une étude conjointe de la forme et du fonds d'un ouvrage numérisé. Mais pour que dans un tel cas, la numérisation soit réellement utile et permette un travail de recherche approfondi sur le document, encore faut-il que le rendu visuel du fac-similé soit de qualité.¹⁴⁸

¹⁴⁷ Par *e-book*, entendons donc « livre électronique », à savoir un ouvrage non pas scanné mais proposé en un mode texte nativement numérique.

¹⁴⁸ À ce sujet, nous conseillerons la lecture de la page « Techniques et formats de conversion en mode texte » du site de la BnF, consultable à l'adresse suivante : http://www.bnf.fr/fr/professionnels/numerisation_boite_ouils/a_num_conversion_mode_texte.html ; ainsi que les procédures de contrôle qualité effectuées par cette même bibliothèque : « Numérisation à la BnF : contrôle qualité »,

Quel rendu ?

L'utilité de l'offre numérique pour l'historien du livre dépend en effet étroitement de la qualité de la numérisation et de sa fidélité relative à l'original conservé. La correspondance visuelle entre l'original et la reproduction détermine l'usage possible du livre numérisé comme source historique.

Tentons ainsi d'évaluer quelques critères qui fondent la qualité du rendu d'une numérisation d'ouvrage, indépendamment du fait qu'il soit ensuite OCRisé ou non.

Un problème récurrent des images d'ouvrages numérisés tient non à la qualité de l'image elle-même mais à l'état de l'original papier. Pensons ainsi aux papiers fins et vieillis qui laissent voir, par transparence, le texte du verso de la page, et prenons pour exemple la page suivante (**figure 14**) :

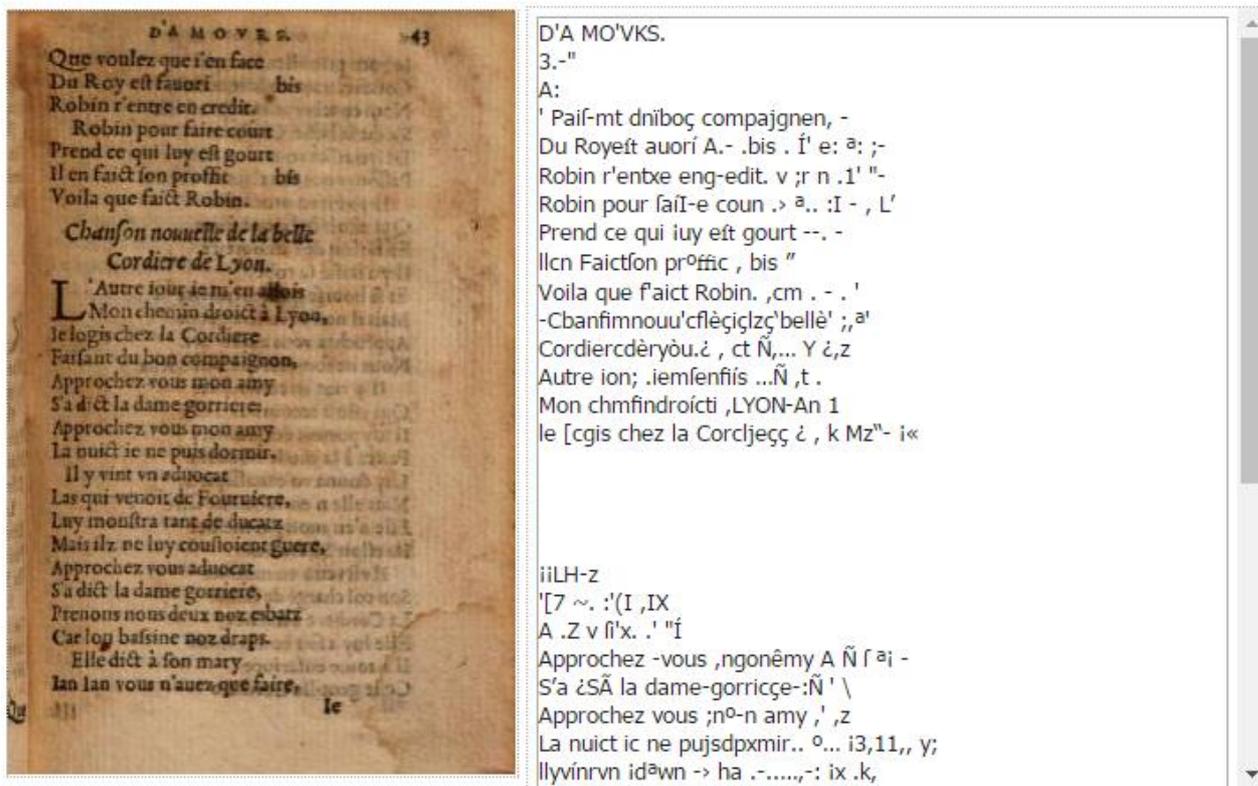


Figure 14. Exemple d'une page abimée, en mode image et en mode texte. Référence : *Recueil de plusieurs chansons divisé en trois parties*, Lyon, Benoist Rigaud et Jean Saugrain, 1557, p. 42 (Vienne, Österreichische Nationalbibliothek).¹⁴⁹

consultable à l'adresse suivante :

http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.num_controle_qualite.html.

¹⁴⁹ La version numérisée de cet ouvrage est consultable à l'adresse suivante : http://digital.onb.ac.at/OnbViewer/viewer.faces?doc=ABO_%2BZ180907702.

Remarquons, dans le cas de cette page, qu'elle illustre à la fois les problèmes de transparence du papier et des dégâts relevant de l'histoire du livre : inégalités de teinte, taches causées par le temps et dégradations naturelles ou humaines – en témoigne la tache d'eau dans l'angle inférieur droit.

Dans le même ouvrage, notons par ailleurs que la page 91 a été rognée et que l'opérateur n'a donc pas préservé l'intégrité de l'ouvrage, ou encore, avec les pages 167 et 192, que la courbure des pages rend le texte difficilement lisible, du moins pour une lecture computationnelle, c'est-à-dire automatisée (**figure 15**).

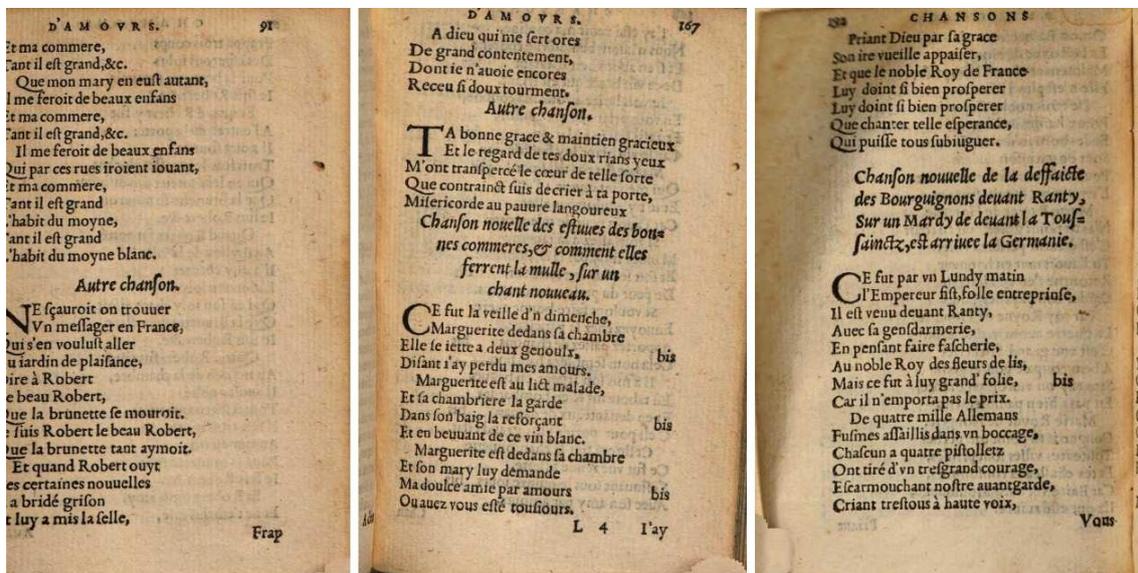


Figure 15. Exemple d'une page rognée à la numérisation, d'une page inclinée et d'une autre courbée. Référence : *Recueil de plusieurs chansons divisé en trois parties*, Lyon, Benoist Rigaud et Jean Saugrain, 1557, pages 91, 167 et 192 (Vienne, Österreichische Nationalbibliothek).

Un seul ouvrage numérisé permet donc d'illustrer quelques défaillances et lacunes de la numérisation, alors que la qualité de l'image est par ailleurs très correcte : un zoom réalisé via l'interface de consultation en ligne rend compte d'un scan de relativement bonne qualité (**figure 16**).

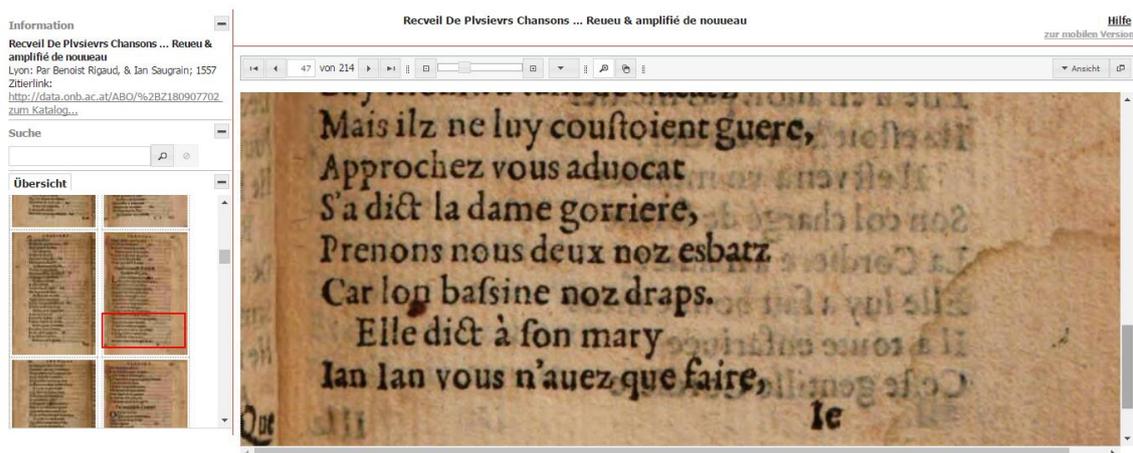
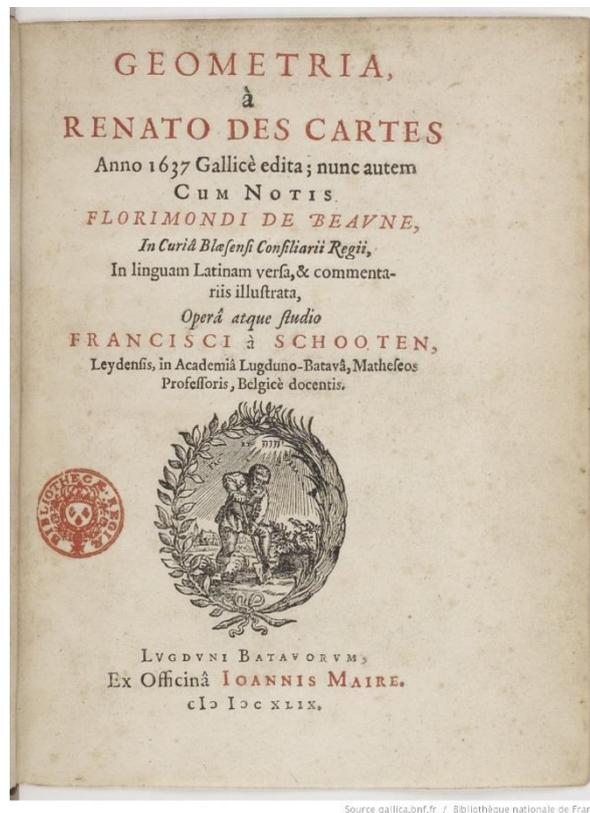


Figure 16. Zoom sur une page abimée de livre ancien. Référence : *Recueil de plusieurs chansons divisé en trois parties*, Lyon, Benoist Rigaud et Jean Saugrain, 1557, p. 42 (Vienne, Österreichische Nationalbibliothek).

Considérons maintenant différents exemplaires de la *Géométrie* de René Descartes pour comparer les rendus divers d'ouvrages numérisés datant du XVII^e siècle. La *Géométrie* est l'un des trois appendices publiés par Descartes en 1637 pour accompagner le *Discours de la méthode*, avec la *Dioptrique* et les *Météores*. Le plus ancien de ces exemplaires choisis, publié chez J. Maire, date de 1649 et est conservé au département Réserve des livres rares de la BnF qui l'a numérisé et mis en ligne en 2010 sur Gallica¹⁵⁰ (**figure 17**). Il s'agit de la première version latine de la *Géométrie*, traduite par le mathématicien hollandais Frans van Schooten ; ce dernier y a adjoint des commentaires et a ainsi permis une diffusion plus large du travail de Descartes chez ses contemporains, alors que la première version de cet essai est publiée en français en 1637, à Leyde. Notons que l'ouvrage a eu des répercussions énormes sur la discipline scientifique et a marqué l'entrée fracassante des mathématiques en physique.

¹⁵⁰ *Geometria à Renato Descartes anno 1637 Gallicè edita : nunc autem cum notis Florimondi de Beaune (...) Operâ atque studio Francisci à Schooten (...)*, Lugdunum Batavorum, J. Maire, 1649 (Paris, BnF). La version numérisée de l'ouvrage est consultable à l'adresse suivante : <http://gallica.bnf.fr/ark:/12148/btv1b8601510z>.



Source gallica.bnf.fr / Bibliothèque nationale de France

Figure 17. Page de titre du *Geometria à Renato Descartes (...)*, Lugdunum Batavorum, J. Maire, 1649 (Paris, BnF).

Une seconde édition datée de 1664 et publiée chez Charles Angot est conservée à la Bibliothèque nationale d'Autriche et a été numérisée par Google en 2014¹⁵¹ (figure 18).

¹⁵¹ *La Géométrie de René Descartes*, Paris, chez Charles Angot, 1664 (Vienne, Bibliothèque nationale d'Autriche). La version numérisée de l'ouvrage est consultable sur Google Livres à l'adresse suivante : https://books.google.fr/books?id=VtFcAAAaAAJ&printsec=frontcover&dq=la+g%C3%A9om%C3%A9trie+descartes&hl=fr&sa=X&redir_esc=y#v=onepage&q&f=false

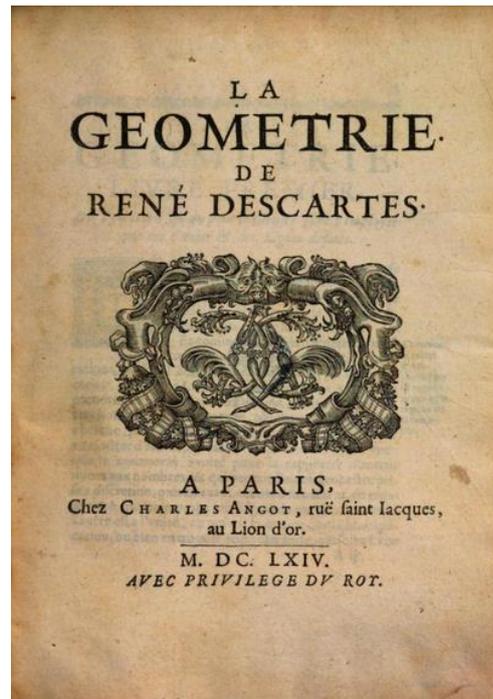


Figure 18. Page de titre de *La Géométrie de René Descartes*, Paris, chez Charles Angot, 1664 (Vienne, Bibliothèque nationale d'Autriche).

Considérons enfin un exemplaire de la troisième édition de 1683, mis en ligne par Gallica en 2007 à partir d'un microfilm¹⁵² (figure 19).



Figure 19. Page de titre du *Geometria à Renato Des Cartes (...)*, Amstelodami, ex typographia Blaviana, 1683 (Paris, BnF).

¹⁵² *Geometria à Renato Des Cartes : anno 1637 Gallicè edita ; postea autem una cum notis Florimondi de Beaune (...)* in latinam linguam versa, & commentariis illustrata opera atque studio Francisci à Schooten (...), ([Reprod.]), Amstelodami, ex typographia Blaviana, 1683 (Paris, BnF). La version numérisée est consultable à l'adresse suivante : <http://gallica.bnf.fr/ark:/12148/bpt6k57484n>.

Pour ce qui est d'une étude de ce dernier exemplaire microfilmé, il est évident qu'une numérisation contrastée en noir et blanc n'apporte pas les informations nécessaires à l'historien du livre pour une étude relevant de la bibliographie matérielle¹⁵³ : l'image, qui est par ailleurs de très bonne qualité, ne rend pas compte de l'apparence visuelle de l'ouvrage, de son papier et du livre-objet comme ensemble structuré signifiant : ni la couverture, ni les pages vierges, ni les marques du temps ne sont visibles.

Il est ainsi déplorable que le téléchargement de l'exemplaire de 1664 sur le site de Google Livres ne nous permette pas d'obtenir la version qui est consultable sur l'interface en ligne : le téléchargement terminé, le lecteur reçoit une version de l'ouvrage en noir et blanc, grossièrement contrastée, à laquelle les versions microfilmées n'ont rien à envier. Un tel fait nous permettra cependant de comparer la même page avant et après retouche automatique (**figure 20**), et donc de pointer du doigt le danger que peuvent représenter les logiciels de retouche d'image relativement à la qualité et à la fidélité de la numérisation. Car l'augmentation du contraste, si elle sert à accentuer le contour des caractères afin de réduire le « bruit » d'une image en niveaux de gris ou en couleur, tend paradoxalement à accentuer certaines taches et à rogner certains caractères.

Notons cependant que si une telle version numérisée ne peut être la source d'un travail approfondi sur la matérialité du livre, elle se prête en revanche tout particulièrement à une analyse typographique du texte, des caractères et des ornements qui le constituent.¹⁵⁴

¹⁵³ Nous traiterons ultérieurement de la bibliographie matérielle, dans « Bibliographie (im)matérielle, fondements d'une archéologie du livre numérisé », p. 78.

¹⁵⁴ A ce sujet, voir notre partie sur « La reconnaissance optique de caractères, mémoire de l'art typographique », page 86.

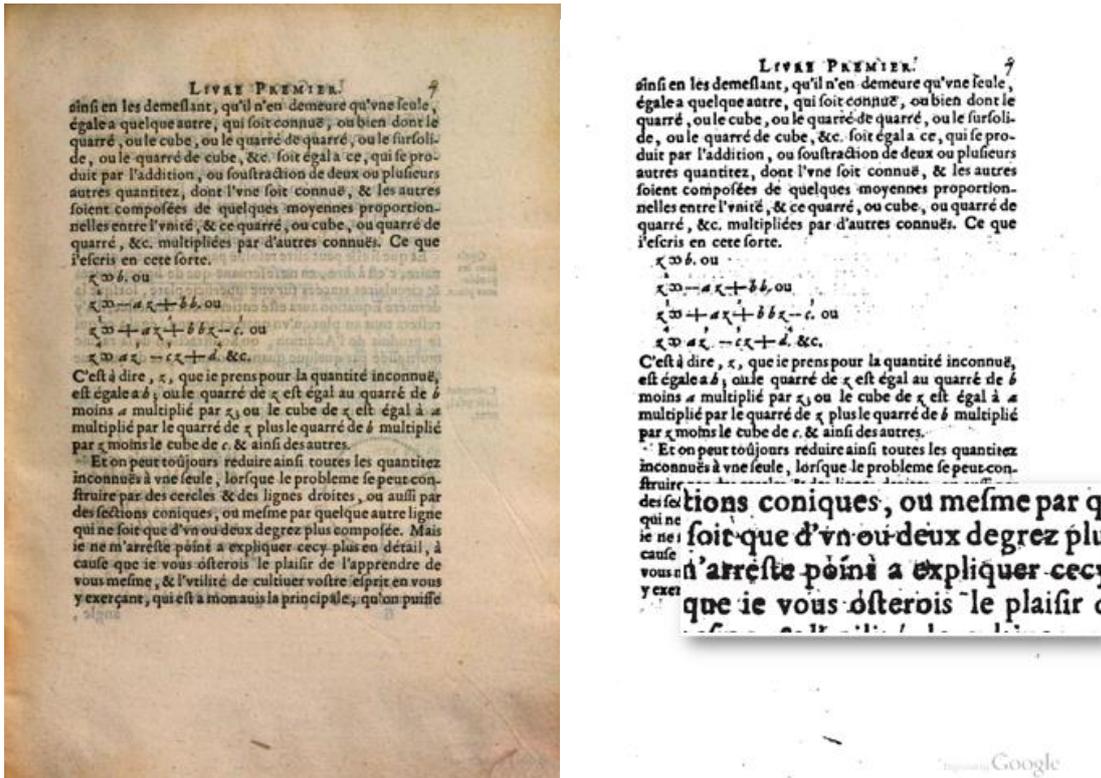


Figure 20. La page 7 de *La Géométrie* de René Descartes, Paris, chez Charles Angot, 1664 (Vienne, Bibliothèque nationale d'Autriche), version en ligne et même image une fois téléchargée.

La version de 1649, numérisée par la BnF en 2010, est plus complète et se porte ainsi mieux à une étude matérielle de l'ouvrage reproduit. On la comparera à l'édition de 1664 mise en ligne par Google Livres en 2014. Ces deux exemplaires s'avèrent représentatifs du rendu des numérisations contemporaines telles que les mènent à grande échelle des structures comme la BnF et Google. L'ensemble du livre est ainsi numérisé, avec sa couverture et ses pages de garde. La numérisation est en couleur et rend compte autant que se peut des teintes et des textures du papier. L'image, de qualité – en témoigne la possibilité de zoomer sans que le rendu en soit trop entamé –, rend compte d'un perfectionnement récent des techniques de numérisation et des logiciels qui l'accompagnent : les pages sont plus régulières, leur courbure plus correctement corrigée et la retouche d'image ne porte pas atteinte à l'intégrité du texte et des caractères. En revanche, un tel scan sans retouche violente ne sert pas la lisibilité du texte que l'on voit à nouveau entaché par l'irrégularité du papier, la vision par transparence du recto de la page ou encore les taches et les problèmes d'encre. C'est bien là, pourtant, que la numérisation sert à fournir une image réaliste et crédible de l'original qu'elle représente.

Pour finir, évoquons la version nativement *numérique* – et non pas *numérisée* – de la *Géométrie* élaborée dans le cadre du projet Gutenberg (**figure 21**) : cette version tapuscrite de l'édition de 1886 publiée à Paris chez A. Hermann¹⁵⁵ offre un texte peut-être retranscrit avec fidélité mais qui ne représente en rien un outil de travail pour l'historien du livre : un tel fichier n'est accompagné d'aucune image qui servirait à le crédibiliser et à l'évaluer. Un tel texte ne présente, du point de vue historique, aucune légitimité à faire source.



Figure 21. La même page en version nativement numérique constituée dans le cadre du projet Gutenberg. Référence : *La Géométrie de René Descartes*, nouvelle édition, Paris, A. Hermann, 1886.

Enfin, nous finirons par un exemple de numérisation remarquable qui vient illustrer les perspectives ouvertes par le numérique relativement à la transmission du savoir livresque et à la préservation de l'objet-livre : les travaux de la Bibliothèque Mazarine, à Paris.

La Bibliothèque Mazarine a en effet inauguré en mai 2015 sa bibliothèque numérique **Mazarinum** autour de la mise en ligne d'un corpus d'incunables¹⁵⁶

¹⁵⁵ *La Géométrie de René Descartes*, nouvelle édition, Paris, A. Hermann, 1886. La version tapuscrite proposée par le site du projet Gutenberg est consultable à l'adresse suivante : https://www.gutenberg.org/files/26400/26400-pdf.pdf?session_id=ae855c047ec15eac958bc7b51092c15771f9a4ca

¹⁵⁶ Par incunable est désigné tout livre imprimé en Europe entre l'invention de l'imprimerie au milieu du XV^e siècle et le début du XVI^e siècle. Du latin *incunabula*, « les langes d'un nouveau-né » les incunables sont les nouveaux-nés de l'imprimerie, héritiers d'un certain nombre de pratiques manuscrites et précurseurs des post-incunables des siècles

sélectionnés parmi les quelque deux-mille-quatre-cents incunables qu'elle conserve actuellement¹⁵⁷. À terme, Mazarinum a pour but de donner accès aussi bien à des livres imprimés du XV^e au XIX^e siècle qu'à des manuscrits, photographies et nombreux autres objets d'art¹⁵⁸. Le travail de numérisation, extrêmement soigné, concerne le livre dans son ensemble : couverture et dos sont numérisés au même titre que l'ensemble des pages de l'ouvrage (**figures 23, 24 et 25**). L'image, de très haute qualité, s'offre au manienement – informatique, bien entendu – de curieux qui jamais, sûrement, n'auraient reçu l'autorisation de consulter des ouvrages d'une telle rareté et d'une valeur si grande. La prise de vue, distanciée, offre une mise en perspective de l'ouvrage qui rendrait presque compte de l'épaisseur du livre (**figure 22**). La très haute définition des images permet un grossissement important sans que la qualité visuelle n'en pâtisse. Les couleurs, enfin, sont d'une fidélité surprenante (**figure 26**).

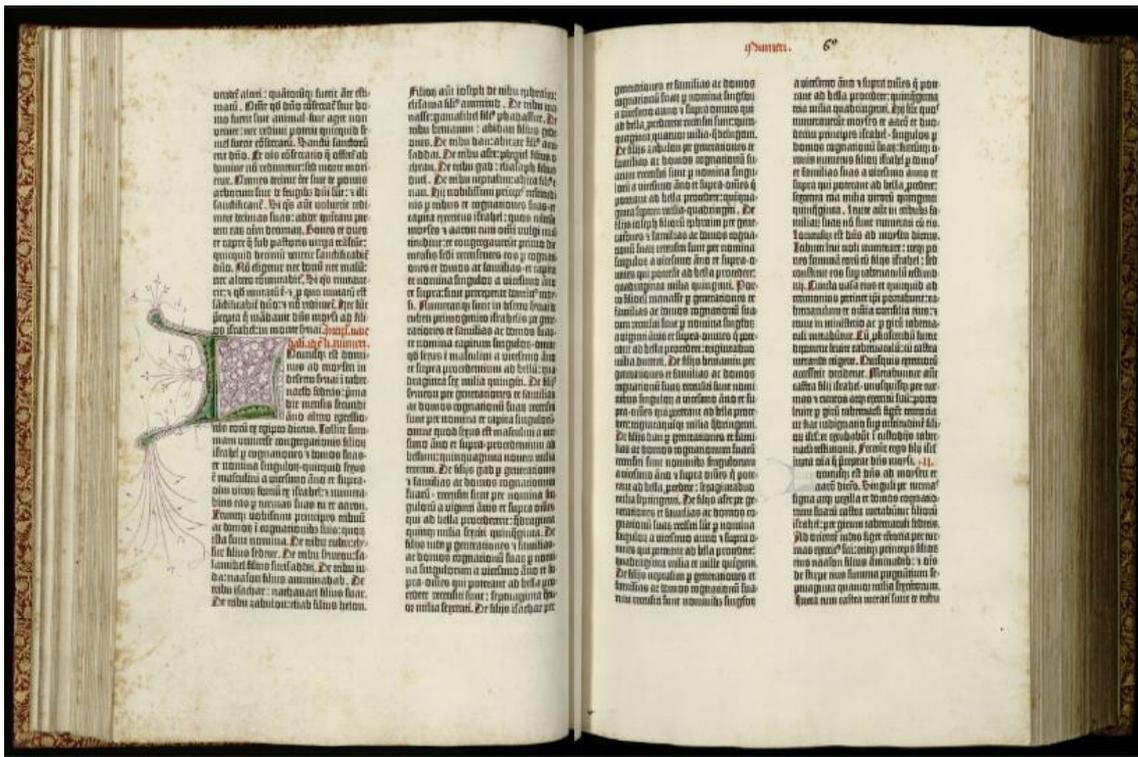


Figure 22. Pages 136 et 137 de la *Biblia latina* (1^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1-1), présentées en double page.¹⁵⁹

suivants. Voir la page Wikipédia consacrée à l'incunable, consultable à l'adresse suivante : <https://fr.wikipedia.org/wiki/Incunable>.

¹⁵⁷ En mai 2016, Mazarinum compte cent-quatorze incunables.

¹⁵⁸ Le site est consultable à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/>.

¹⁵⁹ La version numérisée de cet ouvrage est consultable sur le site Mazarinum à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/collection/7-biblia-latina/?n=2>.

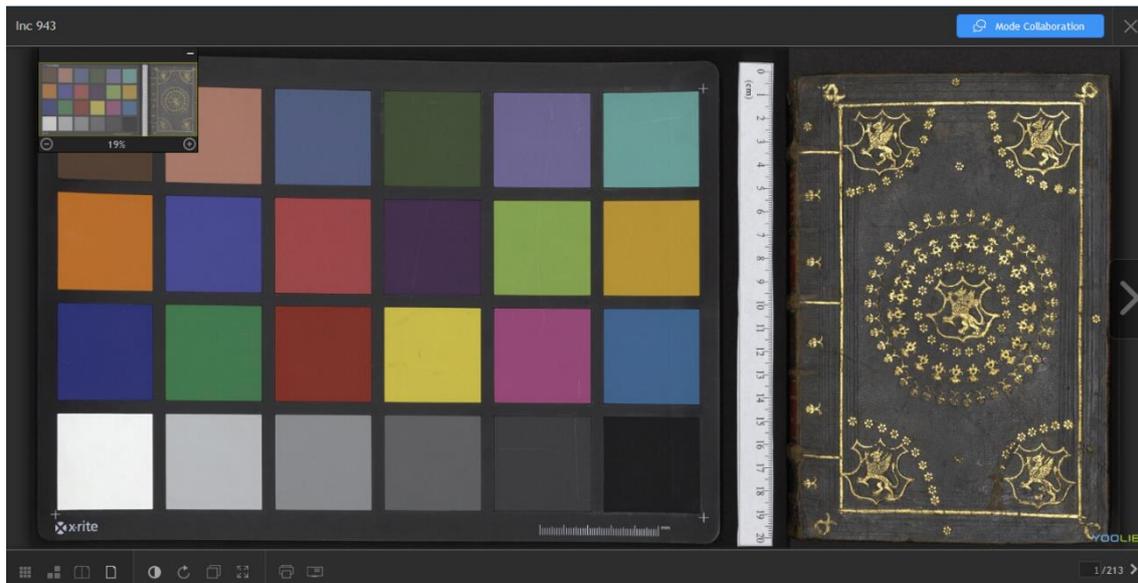
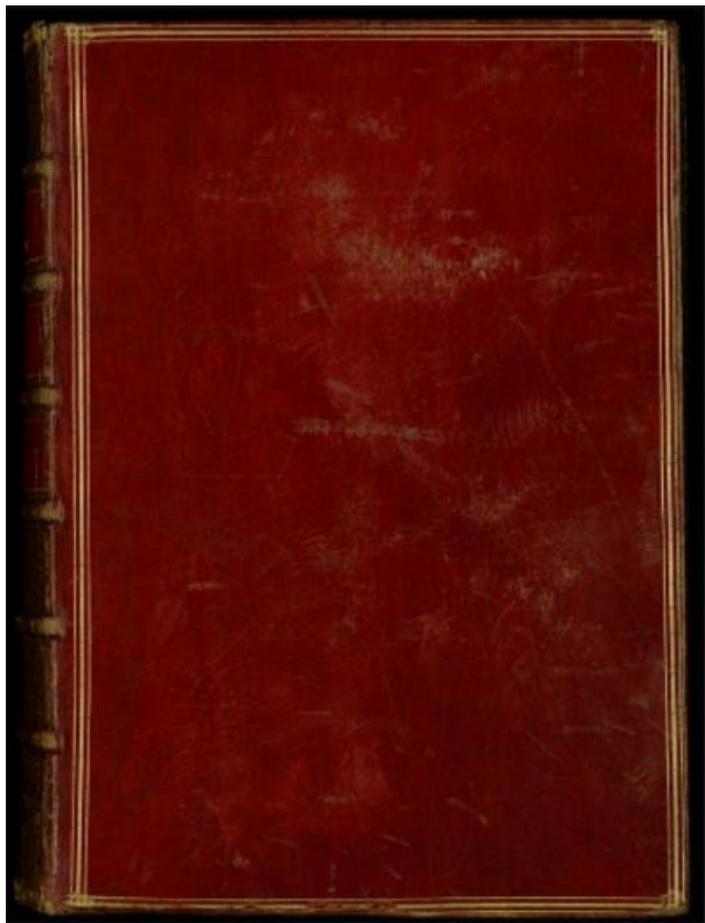
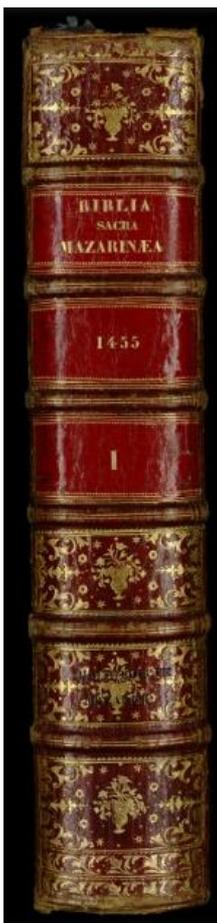


Figure 23. Fenollar, Bernat de. *Lo Procés de les olives e disputa dels jovens hidels vells. Fet per alguns trobadors avant nomenats e lo sompni de Johan Johan, Valence, Lope de la Roca, 1497* (Paris, Bibliothèque Mazarine, Inc 943), couverture.¹⁶⁰



Figures 24 et 25. Dos et plat supérieur de la *Biblia latina* (1^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1-1).

¹⁶⁰ La version numérisée de cet ouvrage est consultable sur le site Mazarinum à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/collection/137-lo-proces-de-les-olives-add-lo-somni-de-joan-jo/?n=1>.

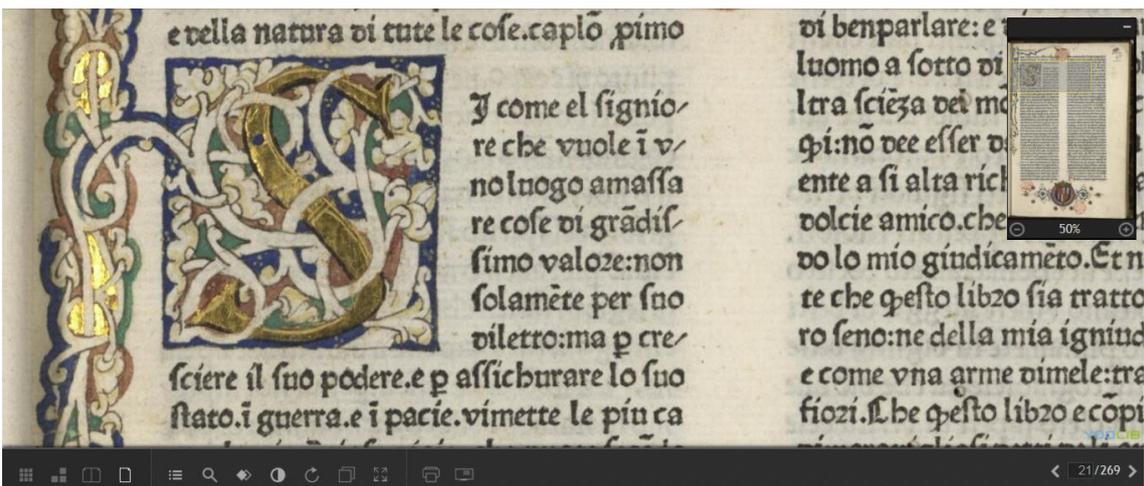
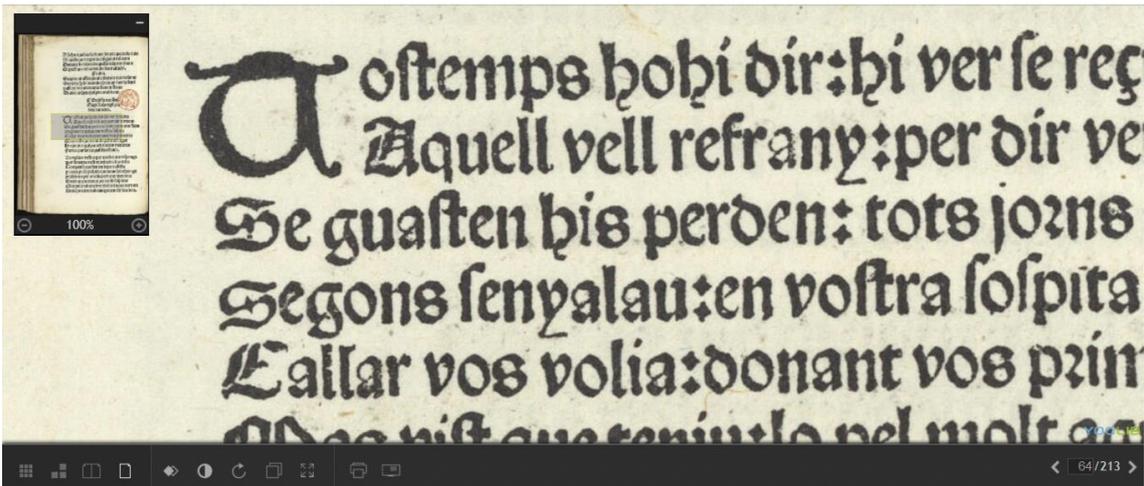


Figure 26. Latini (Brunetto), *El tesoro* ; Trad. Bono Giamboni, Treviso, Gerard de Lisa, 1474 (Paris, Bibliothèque Mazarine, Inc 95), zoom sur les pages 64, 91 et 21.¹⁶¹

¹⁶¹ La version numérisée de cet ouvrage est consultable sur le site Mazarinum à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/collection/69-el-tesoro/?n=1>.

Il est cependant certain qu'un tel soin apporté à la numérisation est le privilège des ouvrages les plus rares qui soient conservés et que les efforts nécessaires à un travail de cette qualité, qu'ils soient humains ou économiques, ne peuvent présentement pas s'appliquer aux masses considérables d'ouvrages que d'autres structures sont en train de numériser. Les incunables de Mazarinum rendent du moins compte des possibilités sans précédent qu'offrent les moyens technologiques actuels en termes de numérisation. S'offrent alors aux chercheurs et aux historiens du livre des objets de recherche nouveaux, accessibles et d'une maniabilité nouvelle.

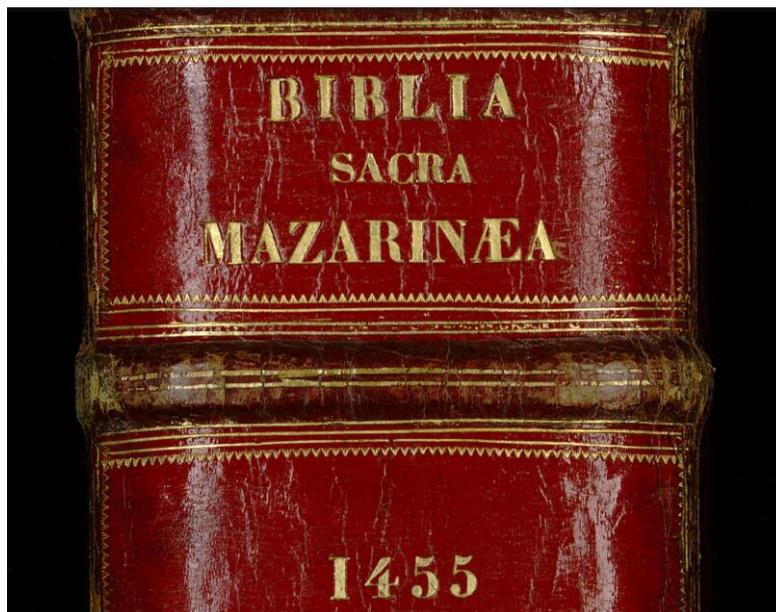


Figure 27. Zoom sur le dos de la reliure de la *Biblia latina* (1^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1-1).

L'offre numérique s'avère donc de nature muable, et l'utilité des politiques de numérisation directement liée à la qualité du processus de constitution du fichier numérique. Loin de se réduire à la seule reproduction en images d'un original papier, la numérisation demande un travail méticuleux d'*ornementation* du fichier scanné, ornements que constituent les métadonnées internes ou externes au document, qui l'enrichissent et le contextualisent, incluant sa transcription en mode texte. Vient ensuite la nécessité d'indexer et de référencer correctement et soigneusement le fichier afin qu'il ne disparaisse pas dans les flots de la masse informationnelle, condition de sa *searchabilité* numérique. Enfin, au même titre qu'un ouvrage papier ancien, le fichier numérique en tant qu'il est inséparable de ses métadonnées doit faire l'objet d'une politique de conservation destinée à lutter contre l'instabilité propre au monde numérique.

Dès lors que ce processus de numérisation est accompli avec rigueur et qu'il continue d'évoluer avec les technologies qui le conditionnent, l'offre numérique ouvre des perspectives non négligeables dans le traitement des sources historiques.

TRAVAIL DES SOURCES

Partant ainsi de l'idée qu'une numérisation correctement menée permet de constituer une image viable et crédible d'un ouvrage que l'on veut considérer comme une source historique, quelles perspectives offrent les outils de traitement numériques ? La source historique se prête en effet à un travail concret de lecture critique, *lecture* au sens large d'évaluation du support contenant ainsi que de son contenu.

RÉALITÉ MATÉRIELLE DE L'IMMATÉRIEL

En ce sens, la dématérialisation du support-livre force à s'interroger sur la *représentativité* du document numérisé – en ce qu'il représente bien, au sens étymologique de « action de replacer devant les yeux de quelqu'un »¹⁶², une réalité physique et matérielle – et sur la valeur de cette représentation.

Quid de la matérialité du livre ?

Nous avons vu qu'une numérisation de qualité pouvait, sans bien sûr remplacer le contact du livre papier, donner une image très fidèle de la matérialité du livre. Plus encore, le fichier numérique offre, paradoxalement peut-être, une maniabilité qui n'est pas comparable à celle du livre en dur.

Pensons ainsi aux précautions que nécessite la manipulation d'un livre ancien : délicatesse dans son maniement, lors de son ouverture, lors de la tourne des pages. La consultation en fonds anciens est réglementée, souvent brève, demande le port de gants et il n'est bien sûr pas possible de marquer le livre ou de l'annoter. La version électronique, reproductible à l'envie, est en cela un support incroyablement maniable et malléable : un fichier numérique se retouche, s'annote ; il est possible d'isoler les pages, de les travailler autant que besoin. La fonction « zoom » des lecteurs d'image offre, si l'image est de qualité, une souplesse d'examen sans déformation de l'information que ne permet pas la meilleure des loupes de poche. Le fichier numérique n'est aujourd'hui soumis à

¹⁶² Voir la page « représentation » du site du CNRTL consultable à l'adresse suivante : <http://www.cnrtl.fr/etymologie/repr%C3%A9sentation>

aucune contrainte de lieu : il se lit sur de multiples supports (ordinateurs, téléphones portables, liseuses) et se transporte aisément – avec l'accès généralisé à Internet et l'avènement des *clouds*¹⁶³, les clés USB et disques durs externes eux-mêmes tendent désormais à être relégués au placard. Les historiens n'ont-ils d'ailleurs pas, dès lors que ça a été possible, utilisé l'appareil photographique pour conserver une image des documents qu'ils ne pouvaient consulter qu'en temps limité ?¹⁶⁴ Admettre cette praticité du document numérique n'est pas signer la mort du livre, elle relève pour les points évoqués ci-dessus de l'évidence. La valeur du livre papier n'en est d'ailleurs pas diminuée. Un tel constat ne fait que confirmer la complémentarité qui peut exister entre le livre et son double numérique.

Les détracteurs du numérique peuvent cependant opposer à ces remarques de non-moindres évidences : le numérique n'est actuellement pas en mesure de rendre compte de la matérialité de l'original papier. Même lorsque sont numérisés la couverture et le dos du livre, un écran ne peut *reproduire* l'épaisseur de l'ouvrage ; au mieux, il la *représente*. C'est la « planéité de l'image » à l'écran qu'évoquent les auteurs de l'*Encyclopédie de la chose imprimée*¹⁶⁵. De même, le rendu des *textures* et de la *couleur* peut être critiquable malgré la qualité des appareils de numérisation, le scan ayant tendance à « aplanir » les surfaces. Sans compter bien sûr les problèmes déjà évoqués d'authentification du document. Un exemplaire numérisé dont il manquerait des pages peut ainsi être complété avec les pages d'un autre exemplaire d'une même édition ancienne sans que la supercherie soit immédiatement visible par l'historien : il lui faudrait là pouvoir étudier le papier par transparence et le raccord éventuel des pages avec la couverture.

Il est à évoquer enfin ce que nombreux historiens définissent comme une forme de « lisibilité » du livre imprimé, à savoir les innombrables marques qui rappellent que le livre est un objet soumis aux effets du temps. On pense à la couleur du papier, à l'odeur de l'encre, aux marques de lecture et à la dégradation des matériaux. Cette « lisibilité » de l'imprimé, Christian Lacombe l'appelle « signification historique »¹⁶⁶ du livre et le présente en ces termes :

« Certes, la possibilité d'accéder à des textes et à les consulter grâce à la numérisation de masse devient quasi infinie, mais nous devons garder à

¹⁶³ Le *cloud computing*, ou l'informatique en nuage, est l'exploitation de la puissance de calcul ou de stockage de serveurs informatiques distants par l'intermédiaire d'un réseau, généralement Internet. Ces serveurs informatiques sont dénommés, dans l'usage courant, des *clouds*. (Voir la page Wikipédia relative au *Cloud computing* consultable à l'adresse suivante : https://fr.wikipedia.org/wiki/Cloud_computing)

¹⁶⁴ Sur ce point, considérons une remarque d'Alain Cullière qui rappelle que les particuliers, autant que les institutions plus larges comme les cabinets de lecture ou les librairies princières, ont joué un rôle essentiel à la Renaissance dans le travail de copie qui a permis de préserver nombre d'ouvrages rares et précieux : des « copies de sauvegarde ». Un tel fait s'applique aujourd'hui avec les particuliers qui, armés de leur appareil photographique, constituent anonymement des copies de sauvegarde de nombreux documents précieusement conservés en fonds anciens. (Alain CULLIERE, « Exploitation d'une « copie de sauvegarde » : le manuscrit 687 du Musée Condé », dans *L'historien face au manuscrit...*, Fabienne HENRYOT (dir.), Louvain-la-Neuve : Presses universitaires de Louvain, 2012, p. 77-78.

¹⁶⁵ Voir l'article « Ecran » dans l'*Encyclopédie de la chose imprimée : du papier à l'écran*, Marc COMBIER et Yvette PESEZ (dir.), Paris : Éd. Retz, 2004, p.100

¹⁶⁶ Christian LACOMBE, « La numérisation ou le grand sommeil du livre imprimé »... p. 41

l'esprit que par cette forme nouvelle de diffusion, la signification historique disparaît. Au sujet de la signification historique d'un livre, je signalerais simplement qu'un vieux livre porte sa propre histoire qu'il transmet au lecteur, de la même manière qu'une personne, avant même de parler, révèle sa vie, qu'il porte sur son visage et dans son allure. Pour appuyer davantage ce propos, je renverrai le lecteur à une conférence de Sarah de Bogui¹⁶⁷ dans laquelle elle oppose le côté « bavard » du livre ancien au « mutisme » du livre contemporain. Nous pouvons aisément remplacer ici le livre contemporain dont elle parle par le livre numérisé, sans, je crois, trahir son propos. »¹⁶⁸.

Ce rapport à la matérialité signifiante et « bavarde » du livre fait écho à ce que Frédéric Barbier appelait « l'environnement du livre » qui « apporte autre chose du point de vue sémiologique »¹⁶⁹ à la caractérisation de l'imprimé et qu'un historien du livre verra comme partie prenante de ce qui relève d'un rapport ontologique à la réalité matérielle de l'objet-livre.

Bibliographie (im)matérielle, fondements d'une archéologie du livre numérisé

Ce rapport physique à l'objet considéré comme un héritage du passé et doté d'une temporalité propre est au fondement d'une discipline récente qui a pour nom la « bibliographie matérielle », et qui consiste en une étude approfondie de la matérialité des ouvrages anciens pour en extraire des informations historiques significatives, relatives à la fois à l'histoire de l'ouvrage lui-même et à l'histoire du livre plus généralement. Citons l'historien Dominique Varry : « La bibliographie matérielle est, en fait, une archéologie du livre imprimé. Elle s'intéresse à la façon dont le livre, en tant qu'objet porteur d'un texte, a pu être fabriqué et transmis. »¹⁷⁰ Le numérique ouvre dans ce domaine une voie nouvelle vers ce que l'on appellera une « archéologie du livre numérisé », en référence à l'expression d'« archéologie du livre imprimé » précédemment utilisée. Quels outils peut offrir le numérique au spécialiste en bibliographie matérielle ? Relativement à ces perspectives ouvertes par le numérique, il est difficile d'être exhaustif. Les recherches dans le domaine du traitement informatique des images et des textes sont nombreuses, se recoupent et se complètent, sans toutefois converger vers l'analyse du livre ancien en lui-

¹⁶⁷ Sarah DE BOGUI, « Les livres voyageurs de la bibliothèque jésuite du collège Jean-de-Brébeuf », p. 3, cité par Christian LACOMBE ; non disponible en mai 2016.

¹⁶⁸ Christian LACOMBE, « La numérisation ou le grand sommeil du livre imprimé »... p. 43

¹⁶⁹ Voir note 110

¹⁷⁰ Voir la page « Qu'est-ce que la bibliographie matérielle ? » du site de Dominique VARRY, consultable à l'adresse suivante : <http://dominique-varry.enssib.fr/node/31>

même. Il est toutefois utile de s'intéresser aux pistes qu'ont pu ouvrir des équipes de recherche très diverses dans le monde de la *bibliographie immatérielle*.

Le numérique est un outil qui, s'il est maîtrisé, a pour vocation de simplifier et d'accélérer les démarches humaines. Il sert en cela la « mécanisation » du travail, c'est-à-dire qu'il permet de déléguer à des machines les tâches fastidieuses sinon impossibles à réaliser pour l'homme. Gardons cependant en mémoire qu'il se doit de toujours rester un moyen et non une fin, un outil et non un acteur dans la démarche de construction du savoir historique.

Dans le cas de l'historien du livre, le numérique intervient aujourd'hui dès les débuts de cette démarche épistémique avec les catalogues des bibliothèques qui sont désormais informatisés et consultables *via* l'Internet. Bien souvent, la rencontre avec l'objet choisi pour étude continue sur écran : on cherche sur Internet des informations relatives au document, des références bibliographiques, voire l'ouvrage lui-même lorsqu'il a été numérisé. On use de bases de données, de portails et de sources d'information diverses. En amont, donc, le travail historique se passe rarement du numérique, et il est rare qu'il s'en passe par la suite. L'historien dont l'objet d'étude n'est pas situé à proximité de son lieu de travail cherchera ainsi à le faire numériser ou, dans le pire des cas, en prendra lui-même un maximum d'images photographiques s'il a l'occasion de le consulter. Ces images, qu'il pourra retoucher et sur lesquelles il pourra zoomer, deviennent elles-mêmes support de travail. Le sujet de la recherche donnera enfin lieu à un travail de rédaction qui se fera lui-même au moyen d'un traitement de texte informatique. Une telle schématisation de la démarche historique – qui ne tient d'ailleurs même pas compte de la communication entre les chercheurs qui passe une fois de plus par les outils de l'Internet – n'est utile que pour montrer combien le numérique est en tout point devenu constitutif du travail historique contemporain. Plus encore, il ouvre de nouvelles voies quant au travail possible sur un ouvrage qui aurait été numérisé. Les outils informatiques d'extraction automatique de l'information s'avèrent en effet prometteurs quant au traitement des images.

Outre l'océrisation que nous avons déjà évoquée et qui n'apporte que peu d'aide au bibliographe qui n'a pas pour vocation d'étudier le contenu textuel du livre mais son archéologie, il existe de nombreux logiciels d'extraction automatique ou semi-automatique de l'information visuelle. Ainsi, certains programmes informatiques sont capables de reconnaître la structure d'une page, par segmentation physique et étiquetage logique. D'autres sont à même de caractériser la structure de lettrines pour les indexer et permettre une recherche thématique par mots-clés. Cette annotation semi-automatique des lettrines, que l'on peut imaginer applicable aux ornements typographiques de livres anciens, ouvre de larges perspectives quant à la constitution de bases thématiques référencées, dans lesquelles il serait possible d'utiliser des moteurs de recherche qui moissonnent les documents, non en fonction de leur description textuelle, mais de leurs caractéristiques visuelles. De

là, la constitution possible de logiciels de comparaison automatique d'images. Quand on sait combien le travail de bibliographie matérielle s'attache à comparer des images (letrines, ornements, structures, caractères), les enjeux s'avèrent importants.

Les espoirs du numérique

Il s'agit donc de revenir ici sur quelques avancées technologiques en lien avec la reconnaissance automatique d'information sur des pages d'ouvrages numérisées. Il existe déjà, comme nous l'avons montré, des bases de données thématiques qui regroupent des images ciblées intéressant l'historien du livre¹⁷¹ ; pensons aux ornements typographiques, par exemple. Ces bases de données sont moissonnables par les moteurs de recherche à une seule condition : que le document ait été correctement catalogué et qu'aient été constituées des métadonnées correctes et précises destinée à identifier l'image numérisée. On cherche alors une image en fonction de ces métadonnées contextuelles : ouvrage dont elle est issue, taille ou toute autre information préétablie concernant son sujet ou son organisation. Le référencement des images demande donc l'intervention d'un personnel humain dont la subjectivité détermine le choix des informations descriptives, ce qui induit un biais informationnel non négligeable dans le cas où l'on tenterait d'identifier un ornement décontextualisé.

Nous avons évoqué plus haut le travail de la base Passe-Partout pour contrer cette subjectivité inhérente au principe du catalogage en tant qu'il est fondé sur une description sémantique et textuelle de l'image. Citons une nouvelle fois Silvio Corsini : « Rien de plus subjectif que l'interprétation textuelle d'une image, aussi simple soit-elle ! Où l'un voit une rose, l'autre voit une pivoine... Sans parler des motifs purement décoratifs dans lesquels il est difficile de déceler un objet clairement identifiable. »¹⁷² (**figures 28 et 29**) Qui recherche un ornement figurant une rose perd ainsi peut-être la chance de l'identifier puisque n'ayant pas remarqué sa similitude avec une pivoine, et n'ayant donc pas en main le mot-clé auquel le moteur de recherche s'attache pour fouiller la base de données. La reconnaissance automatique des formes assistée par ordinateur permet au contraire de déléguer à un logiciel informatique la tâche de référencer l'image au moyen d'un langage numérique que l'on jugera plus neutre, attaché par exemple à la disposition des pixels de l'image et à la distinction automatique des différentes couches informatives constituant l'ornement.

¹⁷¹ Voir *Ressources thématiques, les nouvelles collections*, page 30.

¹⁷² Silvio CORSINI, « Passe-Partout, banque internationale d'ornements d'imprimerie »..., p. 74

Partant de ce principe, le comparateur TODAI employé par Passe-Partout peut ainsi comparer l'image numérisée avec le reste des images contenues dans la base : il propose ensuite à l'utilisateur les dix ornements qui lui sont les plus proches mathématiquement, sans s'attacher à leur parenté sémantique ou thématique.¹⁷³

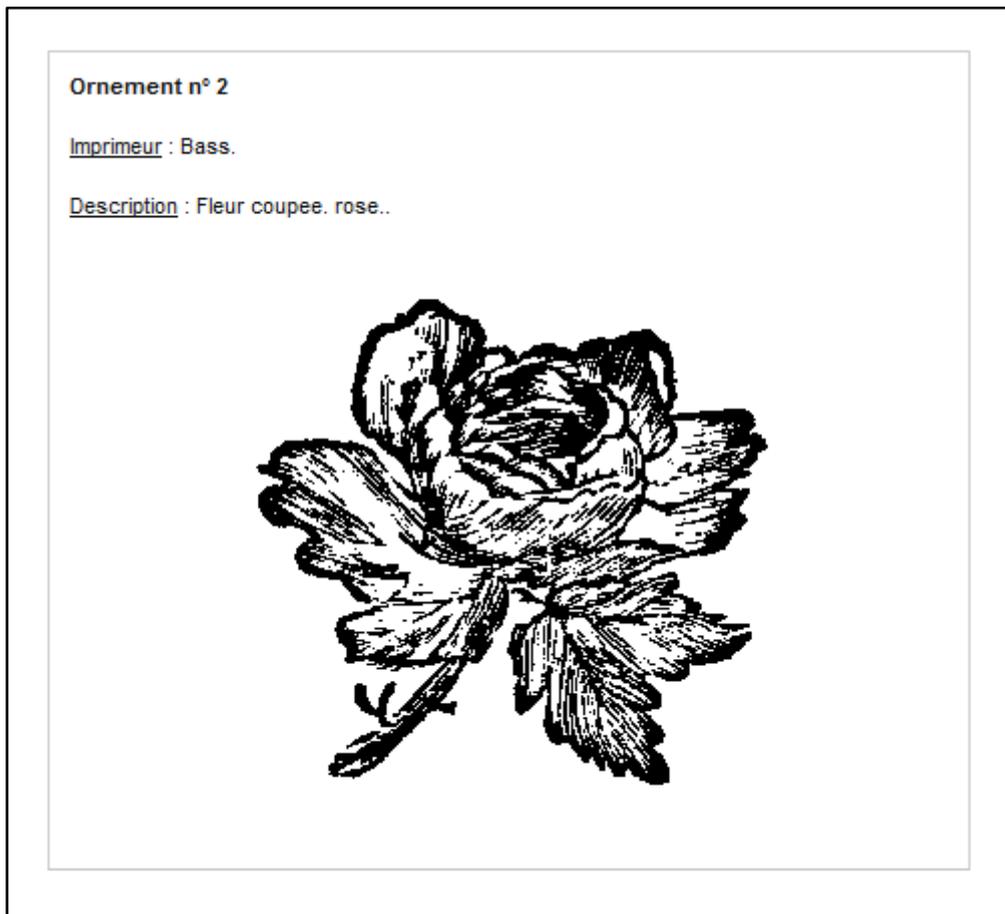


Figure 28. Exemple d'ornement issu de la base Môriâne, utilisé par l'imprimeur Bassompierre à Liège entre 1769 et 1793.

La description référence l'image comme étant une « rose », pourtant cette fleur pourrait être une pivoine : ce flou terminologique, négligeable pour qui n'est pas botaniste, peut rendre l'ornement introuvable lors d'une recherche sémantique.

¹⁷³ Voir l'« Introduction » du site Passe-Partout, consultable à l'adresse suivante : <https://bcutodai.unil.ch/BCUTodai/app/todaiGetIntro.do?uri=todaiInfo&page=todaiIntro.html>

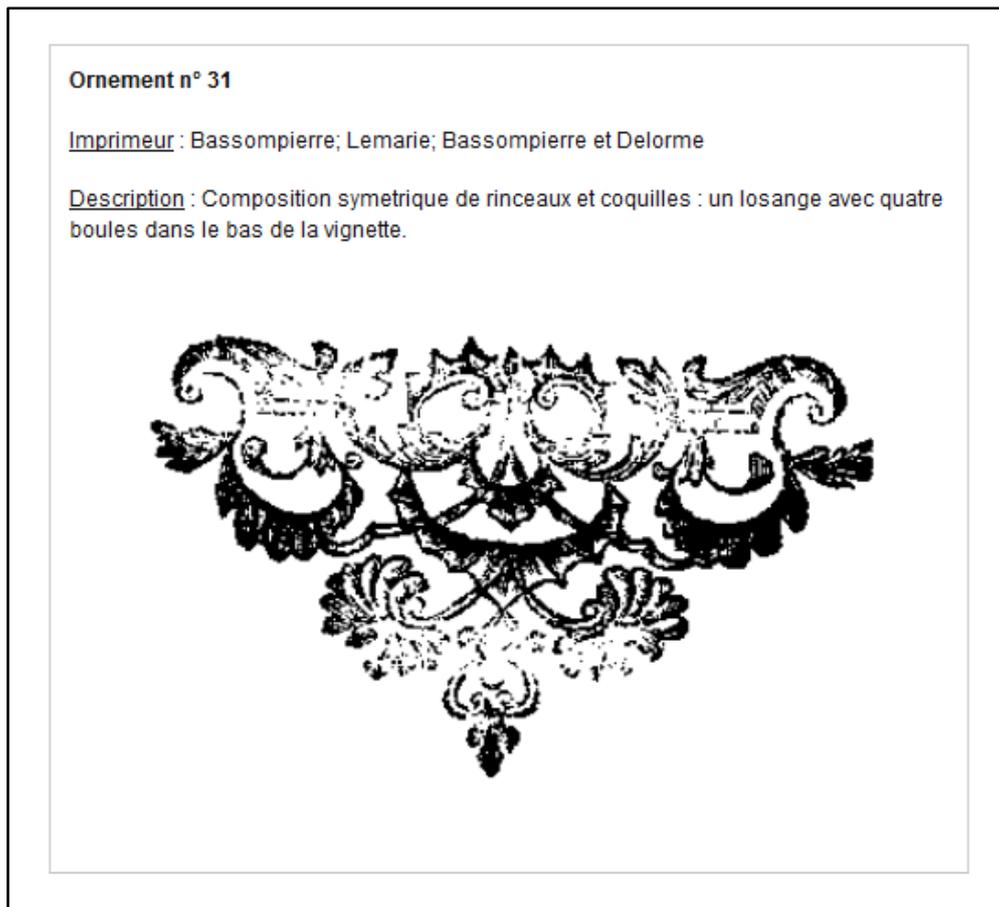


Figure 29. Exemple d'ornement issu de la base Môriâne, également utilisé et vendu par l'imprimeur Bassompierre à Liège entre 1744 et 1793.

Exemple d'ornement purement décoratif, non figuratif, pour lequel il est difficile sinon impossible d'établir une description sémantique objective.

L'auto-référencement des images numériques par le moyen de logiciels de reconnaissance optique peut donc laisser espérer une plus grande objectivité dans le traitement de l'information. Il ne s'agit plus de décrire l'image en fonction d'une discipline, d'un vocabulaire et d'un contexte particuliers, mais de formater une machine qui puisse, sans compréhension de son action, l'étiqueter en fonction d'un langage standardisé.

Considérons ainsi rapidement les travaux menés par Mickaël Coustaty relativement au traitement informatique des documents anciens à l'aide d'ontologies¹⁷⁴. Ce

¹⁷⁴ Voir la page Wikipédia relative à l'ontologie en informatique consultable à l'adresse suivante : https://fr.wikipedia.org/wiki/Ontologie_%28informatique%29. Le terme d'ontologie est ainsi « repris en informatique et en science de l'information, où une ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations, que ce soit par les métadonnées d'un espace de noms, ou les éléments d'un domaine de connaissances. L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Elle est employée pour raisonner à propos des objets du domaine concerné. Plus simplement, on peut aussi dire que l'ontologie est aux données ce que la grammaire est au langage ».

chercheur en informatique s'appuie sur des lettrines¹⁷⁵ tirées de livres datés du XVI^e siècle et cherche à faire le lien entre les régions extraites par les algorithmes de traitement d'images, c'est-à-dire les zones spatiales automatiquement découpées et extraites par les logiciels de reconnaissance de formes, et les éléments sémantiques d'autre part qui servent à donner sens à une image à l'aide de la description textuelle. Ce travail est ainsi destiné à aider les chercheurs dans leur travail d'analyse d'image en permettant de considérer un ornement à la fois comme un signifiant et comme un signifié¹⁷⁶, de « définir une sémantique explicite entre concepts visuels et langage naturel. », « d'associer des mots du langage naturel aux images »¹⁷⁷. Si nous ne nous attarderons pas sur les modalités de ce traitement ontologique des images, indiquons pour résumer ces travaux que l'ontologie recherchée intègre caractéristiques numériques des régions d'intérêt de l'image traitée, informations spatiales entre les régions et sémantique historique¹⁷⁸. Les images sont partitionnées : sont ainsi distinguées des couches (lettre, motif, fond, cadre et couche forme, couche texture) et des régions d'intérêt (région située au centre de l'image, région avec peu de trous, région d'aire maximale...), à partir desquelles les chercheurs espèrent pouvoir réduire de manière semi-automatique le fossé sémantique entre le vocabulaire utilisé par les historiens de la Renaissance et celui du traitement d'images, dans l'optique de concevoir un moteur de recherche qui ne soit plus seulement basé sur le catalogage textuel¹⁷⁹ (**figures 30 et 31**). Cela induit bien sûr de la part de l'historien-chercheur une bonne maîtrise de l'informatique et des logiciels en question.

¹⁷⁵ Les lettrines sont des lettres décorées placées en début de paragraphe dans les livres anciens.

¹⁷⁶ Ces termes, empruntés à la linguistique saussurienne, servent ici à distinguer le dessin de ce qui est dessiné, le trait purement graphique de la lecture sémantique.

¹⁷⁷ Mickaël COUSTATY, Norbert TSOPZE, Karel BERTET, Alain BOUJU et Georges LOUIS, « Traitement des documents anciens à l'aide d'ontologie », dans « Valorisation des corpus numérisés », CHEVRY PEBAYLE, Emmanuelle (dir.), *Les Cahiers du numérique*, vol. 8, 2012/3, p. 93

¹⁷⁸ Mickaël COUSTATY, Norbert TSOPZE, Karel BERTET, Alain BOUJU et Georges LOUIS, « Traitement des documents anciens à l'aide d'ontologie »..., p. 96

¹⁷⁹ *Ibid.*, *op. cit.*, p. 116

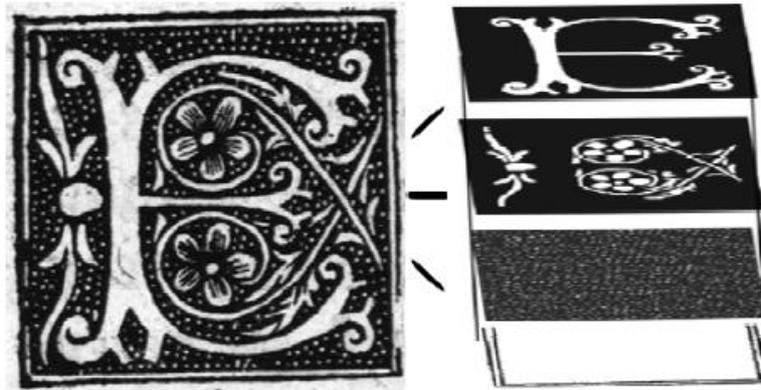


Figure 30. Exemple de décomposition en couches d'une lettrine entre lettre, motif, fond et cadre. Image empruntée à l'article « Traitement des documents anciens à l'aide d'ontologie » de Mickaël COUSTATY, Norbert TSOPZE, Karell BERTET, Alain BOUJU et Georges LOUIS¹⁸⁰.



(a) Image d'origine (b) Couche forme (c) Couche texture

Figure 31. Exemple de décomposition en couches d'une lettrine entre forme et texture. Image empruntée à l'article « Traitement des documents anciens à l'aide d'ontologie » de Mickaël COUSTATY, Norbert TSOPZE, Karell BERTET, Alain BOUJU et Georges LOUIS¹⁸¹.

Si l'on étend de tels outils aux autres catégories d'images que l'on peut extraire d'une page de livre ancien, comme les autres ornements voire la page entière, on imagine combien ces bases de données automatiquement référencées pourraient aider le chercheur qui souhaiterait comparer divers ouvrages. Elles permettraient de répondre informatiquement à des questions qui intéressent le bibliographe telles que : dans quel ouvrage de quel imprimeur tel bois a-t-il déjà été employé ? Combien de temps tel imprimeur a-t-il utilisé tel caractère dans son atelier ? Tel ornement nouveau ne vient-il pas d'un autre atelier d'imprimerie, ce qui révélerait une fausse adresse ou une datation erronée ? De telles données élargiraient de beaucoup le champ de la recherche sur la circulation des imprimés ou encore sur l'histoire des techniques relatives à l'imprimerie.

¹⁸⁰ Mickaël COUSTATY, Norbert TSOPZE, Karell BERTET, Alain BOUJU et Georges LOUIS, « Traitement des documents anciens à l'aide d'ontologie », dans « Valorisation des corpus numérisés », CHEVRY PEBAYLE, Emmanuelle (dir.), *Les Cahiers du numérique*, vol. 8, 2012/3, p. 100

¹⁸¹ *Idem*, *Op. cit.*, p. 102

Plus encore, on peut rêver de la mise au point d'un logiciel qui comparerait avec une grande précision différents exemplaires d'une même édition d'un même ouvrage pour en lister les différences et ainsi effectuer le travail fastidieux auquel œuvrent les bibliographes au moyen du *collator*.¹⁸²

Une autre technique, peut-être moins avancée et dont on imagine mal qu'elle puisse s'appliquer à grande échelle, est la *photogrammétrie*. Celle-ci consiste en l'extraction des informations volumiques d'une page, pour conserver la réalité physique d'un ouvrage, ce que ne permet bien sûr pas la numérisation 2D. Les chercheurs Rodrigo Almeida, Pierre Cubaud, Jérôme Dupire et Alexandre Topol en propose ainsi l'expérience et justifient leur démarche en ces termes : « Les prises de vue n'altèrent pas l'ouvrage par une mise à plat forcée et permettent d'obtenir un maillage 3D assez précis. Un livre n'est pas seulement un ensemble de pages planes. C'est un volume, complexe, avec parfois une riche dynamique d'interaction lorsqu'il comporte des systèmes dépliables. »¹⁸³. L'utilisation conjointe d'un laser à balayage et d'un appareil photographique numérique permet ainsi d'envisager la numérisation en trois dimensions d'ouvrages rares et anciens. Et eux de conclure : « Un grand soin est en général pris dans la numérisation des textes, mais on ne prêt pas assez attention au médium sous-jacent (le livre), à son environnement physique (la bibliothèque), ainsi qu'aux gestes et aux démarches des lecteurs. Passer de l'interface textuelle unidimensionnelle au volume, comme les technologies 3D temps réel l'autorisent dès à présent, pourrait permettre d'élargir le flux de communication entre les bibliothèques numériques et leurs usagers. »¹⁸⁴.

Il est à noter que les interfaces numériques de visualisation – visualisation que l'on dit « immersive » – peuvent également servir à construire la représentation d'une bibliothèque numérique en trois dimensions qui rendrait compte du volume des collections, des ouvrages, et dans laquelle il serait possible de se déplacer virtuellement.¹⁸⁵ Bien loin de ces bibliothèques « virtuellement réelles », une interface de visualisation 3D sert déjà sur certains sites à la consultation des ouvrages numérique, qui permet de « feuilleter » de manière interactive l'ouvrage consulté. Il s'agit cependant là d'une fonctionnalité purement décorative à valeur

¹⁸² Le *collator*, ou *machine à collationner*, est un appareil qui sert à comparer deux exemplaires *a priori* identiques d'un même ouvrage. Par un jeu de miroirs très précis, il permet de superposer l'image de deux pages afin de distinguer rapidement d'éventuelles différences typographiques mises en évidence par une zone de flou : si les pages sont identiques, les deux images superposées vont au contraire se confondre et le texte sera parfaitement lisible. Pour plus d'informations sur le *collator*, voir la page « Le matériel du bibliographe » mise en ligne sur le site de Dominique VARRY et consultable à l'adresse suivante : <http://dominique-varry.enssib.fr/node/30>

¹⁸³ Rodrigo ALMEIDA, Pierre CUBAUD, Jérôme DUPIRE et Alexandre TOPOL, « Interactions et métadonnées riches pour les bibliothèques numérisées », dans « Visualisation pour les bibliothèques numériques », FEKETE, Jean-Daniel et LECOLINET, Eric (dir.), Revue *Document numérique*, vol. 9, 2006/2, p. 86

¹⁸⁴ *Ibid.*, *op. cit.*, p. 105

¹⁸⁵ A ce sujet, se reporter à l'ouvrage précédemment cité, notamment le chapitre 2, « Interaction 3D immersive pour la lecture et la navigation », p. 86

historique nulle. Mazarinum¹⁸⁶ permet ainsi de feuilleter les incunables de sa bibliothèque numérique¹⁸⁷.

Les perspectives ouvertes par le numérique relativement au traitement des sources numérisées sont donc conséquentes. Elles concernent aussi bien la capacité des techniques numériques à rendre compte de la matérialité du support original que les potentialités de certains logiciels informatiques à référencer en tant qu'image des textes et des ornements qui deviennent ainsi interrogeables et comparables de manière automatisée, sans qu'ils perdent de leur intégrité matérielle et de leur unicité historique.

QUE RESTE-T-IL DE L'OBJET-LIVRE ?

Si l'on espère voir les images numériques conserver cette intégrité matérielle et cette unicité historique qui déterminent la légitimité du document numérique à servir de source, intégrité et unicité ne concernent cependant pas uniquement les fragments textuels ou picturaux qui peuvent intéresser l'historien. La question se pose en effet de la conservation du livre en tant qu'ensemble signifiant, notion que vient mettre à mal un processus de fragmentation propre au numérique qui porte atteinte à la mémoire même de l'original comme contenant temporel à historiciser.

La reconnaissance optique de caractères, mémoire de l'art typographique

Outre les quelques pistes que nous venons d'explorer relatives à la conservation et au traitement des fichiers numériques comme reproduction du livre en tant qu'objet, et qui ne représentent finalement qu'une voie réduite parmi les vastes sentiers ouverts par la révolution numérique, de nombreux projets mettent à mal la notion même d'objet-livre considéré dans sa dimension matérielle. Ces projets concernent le traitement du texte contenu dans les fichiers numérisés : le texte y est donc isolé du cadre figé qui le véhicule, à savoir le livre. Nous avons déjà abordé le principe de l'océrisation, au moyen duquel des logiciels d'extraction automatique de l'informatique décodent les caractères du texte numérisé et en proposent une traduction nativement numérique (**figure 32**).

¹⁸⁶ Le site est consultable à l'adresse suivante : <http://mazarinum.bibliotheque-mazarine.fr/>

¹⁸⁷ Cliquer sur la commande « Feuilletage » dans la barre des tâches de l'interface de visualisation du site de Mazarinum.

Les limites de ces logiciels quant au traitement des imprimés anciens relèvent de l'histoire de l'imprimerie et des techniques qu'elle mobilise. En ce sens, une analyse de ces limites servira la mémoire de l'art typographique.

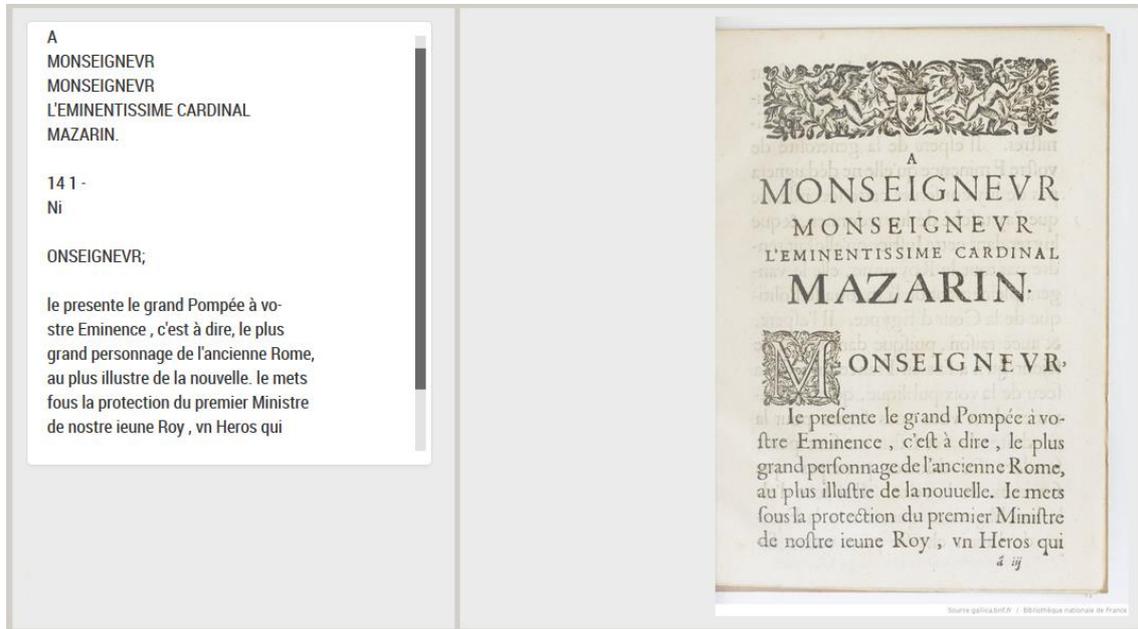


Figure 32. Première page d'une dédicace de Pierre Corneille au Cardinal Mazarin, accompagnée de sa version texte sur Gallica. Référence : Pierre Corneille, *La mort de Pompée*, à Paris, A. de Sommaville et A. Courbé, 1644 (Paris, BnF).¹⁸⁸

Voyons ainsi quelques cas sur lesquels les logiciels d'océrisation butteraient et qui dévoilent les ficelles des techniques typographiques. Nous ne tiendrons pas compte des problèmes liés au vieillissement ou à la dégradation du papier que nous avons déjà évoqués, tels que les taches, moisissures, plis, déchirures et autres traces du temps et de l'usage, et considérerons deux points : les accidents typographiques et le problème des caractères spéciaux.

L'impression d'une page à l'époque des presses dites *typographiques* avait lieu en plusieurs temps : des caractères mobiles étaient assemblés puis fermement maintenus pour former un texte en relief appelé *forme imprimante* sur laquelle appliquer le papier après encrage. Les typographes, par erreur, pouvaient ainsi placer les caractères mobiles dans le mauvais sens ou remplacer un caractère par un autre. Plus rare, un caractère extérieur à la forme imprimante passait parfois à l'impression : on imagine qu'il s'agissait là d'un caractère mal agencé qui aurait sauté sous l'effet de la pression exercée sur le cadre qui enserre les caractères, ou bien un caractère tout simplement oublié par inadvertance sur la forme avant

¹⁸⁸ La version numérisée de cet ouvrage est consultable sur le site Gallica à l'adresse suivante : <http://gallica.bnf.fr/ark:/12148/bpt6k6578719b.r=La%20mort%20de%20Pomp%C3%A9e%201644>.

l'impression.¹⁸⁹ (figure 33). Enfin, un caractère mobile placé à l'envers, c'est-à-dire « pied-en-haut », pouvait remplacer un caractère manquant à l'imprimerie et donne à l'impression deux petits rectangles noirs (figure 34), les « pieds » du caractère, séparés par une *gouttière* (voir figure 35, numéro 8). Il s'agit là de spécificités typographiques relativement rares, mais face auxquelles les logiciels sont pris au dépourvu et qu'ils tentent, au regard de leur forme, de traduire par des lettres. La lecture du texte transcrit en devient difficile.

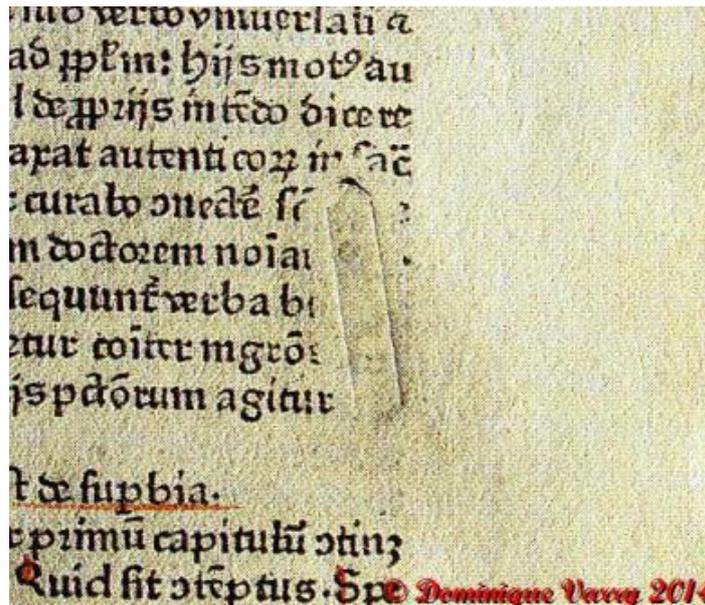


Figure 33. Caractère mobile tombé sur la forme imprimante. Photo prise par Dominique VARRY du recto du second feuillet d'un exemplaire *in-quarto* du *De lepra morali Tractatus* de Johannes Nider (1479) conservé à la Bibliothèque municipale de Versailles sous la cote Inc M132.¹⁹⁰

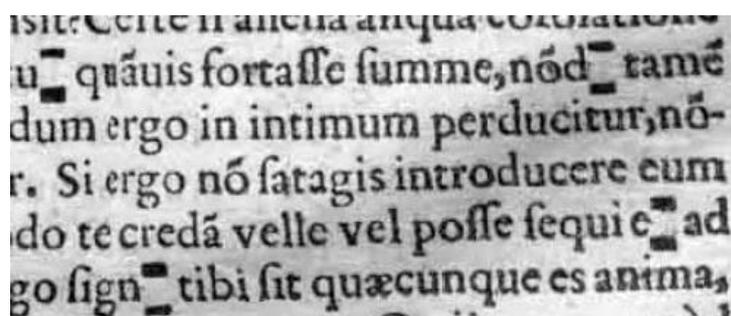


Figure 34. Exemple d'un caractère mobile retourné pied-en-haut pour indiquer que le caractère « u » manque à la presse. Photo empruntée à Rémi Jimenes¹⁹¹ et tirée des *Epreuves de S. Bernard*, Opera, Paris, Ch. Guillard, 1551 (Lyon, BM).

¹⁸⁹ A ce sujet, voir l'article « UFOs – OVNI, Unidentified Fallen Objects, Objets non identifiés tombés sur la forme » du blog de Dominique Varry, consultable à l'adresse suivante : <http://dominique-varry.enssib.fr/node/210>.

¹⁹⁰ Voir la page consacrée du blog de Dominique VARRY à l'adresse suivante : <http://dominique-varry.enssib.fr/node/210>.

¹⁹¹ « Pratique d'atelier et corrections typographiques à Paris au XVI^e siècle : les œuvres de saint Bernard imprimées par Charlotte Guillard (1551) », dans *Passeurs de texte, imprimeurs et libraires à l'âge de l'humanisme*,

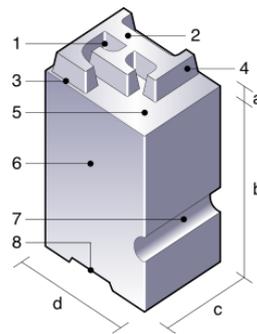


Figure 35. Schéma du caractère mobile en plomb typographique de la lettre F grande capitale. Image issue de la section « Caractère mobile en métal » de l'article « Caractère (typographie) » de Wikipédia¹⁹².

Plus répandus, les défauts d'impression déjà évoqués posent le même problème : dégradations des caractères mobiles et inégalités d'encre et de pression sont causes d'une irrégularité graphique qui peut induire le logiciel en erreur. Parce qu'elle est mécanique, l'imprimerie n'offre donc pas la régularité et l'uniformité que l'on connaît avec les traitements de texte de ces dernières décennies : ces caractéristiques sont source de « bruit » et entravent le bon fonctionnement des processus d'océrisation.

Une équipe de chercheurs de la ville de Tours, qui travaille sur les imprimés de la Renaissance dans le cadre du projet baptisé les Bibliothèques Virtuelles Humanistes, a ainsi buté sur le problème de la transcription des caractères particuliers utilisés dans les ateliers d'impression de l'époque. Leurs travaux ont donné lieu à de très intéressantes publications relatives à la gestion informatisée des écritures anciennes. Il s'agit là, explique ainsi Rémi Jimenes, d'un domaine de recherche important à l'heure où « les historiens des textes se trouvent ainsi confrontés plus souvent qu'auparavant à la réalité graphique des sources originales et envisagent d'en rendre compte dans leurs travaux de transcription. »¹⁹³.

Prenons ainsi pour exemple les travaux de Jacques André et de ce même Rémi Jimenes qui décrivent, dans « Transcription et codage des imprimés de la Renaissance »¹⁹⁴, les limites des logiciels d'océrisation et les problèmes que cela pose pour une « paléotypographie » du texte numérisé et transcrit. Le terme de

Christine BENEVENT, Annie CHARON, Isabelle DIU et Magali VENE (éd.), Paris : Ecole nationale des chartes, 2009, p. 205-228

¹⁹² La page est consultable à l'adresse suivante : [https://fr.wikipedia.org/wiki/Caract%C3%A8re_\(typographie\)](https://fr.wikipedia.org/wiki/Caract%C3%A8re_(typographie)).

¹⁹³ Rémi JIMENES, « Introduction », dans « Gestion informatisée des écritures anciennes », BENEVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 8

¹⁹⁴ Jacques ANDRE et Rémi JIMENES, « Transcription et codage des imprimés de la Renaissance : réflexions pour un inventaire de caractères anciens », dans « Gestion informatisée des écritures anciennes », BENEVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 113-139

« paléographie », utilisé par les deux auteurs, est repris à Hendrik Vervliet qui désigne l'étude, à des fins d'identification bibliographique, des fontes employées par les imprimeurs des siècles passés¹⁹⁵.

Une forme d'« inertie des pratiques manuscrites »¹⁹⁶ existe encore à la Renaissance qui pousse les imprimeurs à travailler sur le rendu visuel des caractères d'imprimerie afin qu'ils imitent l'écriture manuscrite. Aussi les abréviations, contractions, chevauchements, caractères phonétiques, ligatures et signes spécifiques de ponctuation sont-ils légion. Ces subtilités typographiques viennent gonfler la gamme des caractères existants que les logiciels d'océrisation doivent être à même de transcrire pour que le texte soit lisible et qu'il puisse donner lieu à une analyse paléotypographique approfondie. Aussi, si la typographie « constitue de ce fait un système d'écriture *fermé*, nécessairement *normalisé* et *fini* »¹⁹⁷, les alphabets actuels qui servent de référence à ces logiciels ne sont pas adaptés et restent incomplets.

Jacques André et Rémi Jimenes militent donc pour que le codage¹⁹⁸ le plus complet et le plus utilisé qui soit, Unicode¹⁹⁹, soit enrichi de ces signes typographiques propres à la Renaissance. Un projet de médiévistes, le projet MUFI, cherche ainsi à lister ces caractères manquants pour qu'Unicode les intègre à son codage. La démarche n'entre par ailleurs pas en contradiction avec la vocation d'Unicode qui est de constituer un catalogue exhaustif des caractères graphiques, sans restriction temporelle ou linguistique.

Pour que la transcription soit plus exacte, ils introduisent la notion de « typème »²⁰⁰ – contraction de *type* et de *graphème* – qu'ils jugent avoir plus de pertinence : l'appellation admise, qui différencie actuellement le « caractère » de ces multiples variantes « glyphiques », tend selon eux à nier la réalité des caractères propres à la Renaissance. Directement lié à la réalité technique de la typographie d'imprimerie, le typème qualifie ainsi « l'unité minimale effective entrant dans la composition d'un texte affiché ou imprimé »²⁰¹ : le « *ct* » (« *ct* ligaturé »), considéré par Unicode comme une variante « glyphique » de « *c t* » et qui sera donc traduite comme telle, est pourtant un « typème » au sens où il existait, chez les imprimeurs, un caractère fondu spécifique – un « type » – destiné

¹⁹⁵ H. D. L. VERVLIEET, *The Palaeotypography of the French Renaissance : Selected Papers on Sixteenth-Century Typefaces*, Brill, Leiden, 2 vol., 2008, cité en note de bas de page par Jacques ANDRE et Rémi Jimenes dans « Transcription et codage... », p. 115

¹⁹⁶ Jacques ANDRE et Rémi JIMENES, « Transcription et codage... », p. 114

¹⁹⁷ *Ibid.*

¹⁹⁸ Le codage sert à la « transmission de caractères entre ordinateurs et périphériques » ; il ne s'agit pas d'un logiciel, mais d'un codage qui associe à chaque caractère référencé un numéro (*code point*), un nom et des propriétés descriptives. (*Ibid.*, *op. cit.*, p. 120)

¹⁹⁹ Publié en janvier 1991, l'Unicode est un système de codage informatique sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés.

Pour plus de précisions, voir « Unicode » dans le *Dictionnaire du NEF* de Marie Lebert, consultable à l'adresse suivante : <http://www.etudes-francaises.net/dico/final.htm>, ou la page Wikipédia dédiée : <https://fr.wikipedia.org/wiki/Unicode>

²⁰⁰ Voir la quatrième partie de « Transcription et codage... » de Jacques ANDRE et Rémi JIMENES : « Le typème, chaînon manquant entre caractère et glyphe », p. 123

²⁰¹ *Ibid.*

à l'impression du « ct », qui devrait donc être transcrit ainsi à l'océrisation. Selon Jacques André et Rémi Jimenes, une transcription typémique rigoureuse permettrait également de pallier certains accidents typographiques évoqués précédemment. Les caractères illisibles ou non-reconnus parce qu'abîmés ou mal encrés pourraient être remplacés par un glyphe unique qui s'appliquerait dès lors que la reconnaissance automatique est approximative. De même, un glyphe « pied-en-haut » pourrait figurer un rectangle noirci qui marquerait l'utilisation accidentelle ou volontaire d'un « type » retourné ; un code spécifique servirait de même à traduire l'utilisation d'un caractère renversé, c'est-à-dire positionné dans le mauvais sens.²⁰² Un tel projet, s'il était pris en compte par Unicode, permettrait certainement de mieux océriser, au plus près du texte original, les textes imprimés de la Renaissance qui intéressent ces chercheurs.

La prise en compte par les codages de texte des originalités typographiques de la Renaissance ouvre ainsi une voie nouvelle vers la transcription automatique de textes anciens dont la numérisation dépasserait de loin la seule constitution d'un fac-similé. Les ouvrages numérisés en acquerraient une légitimité nouvelle à faire office de sources historiques fiables. Nous laisserons les deux auteurs conclure : « Seule l'existence de ce niveau de transcription « typémique » permet de coller au plus près de la réalité concrète du texte imprimé et de mettre en œuvre des recherches paléotypographiques susceptibles d'intéresser linguistes et historiens.»²⁰³.

Certains de ces historiens, cependant, sont en droit d'être sceptiques. Car avec l'océrisation et la transcription dans un langage nativement numérique des ouvrages scannés, c'est l'objet-livre en lui-même qui est dépassé. Il ne s'agit pas seulement de constituer un mode texte à partir du mode image pour des questions de rapidité de lecture et de compréhension du texte, il s'agit de rendre compte en tout point de l'original à travers sa version numérique. Une transcription précise a donc pour but de rendre autonome le texte de l'image dont il est tiré. C'est ainsi du moins que l'envisagent ces différents chercheurs, en témoignent les propos de Rémi Jimenes :

« Le document écrit est un *tout*. Son contenu intellectuel est manifesté par une réalité matérielle qui contribue à lui conférer sa valeur historique – c'est-à-dire sa capacité à représenter « un moment déterminé de l'évolution dans un domaine quelconque de l'activité humaine » (Aloys Riegl). Un même document est dès lors susceptible d'être soumis à des analyses sémantiques, stylistiques, linguistiques, paléographiques ou archéologiques, et chaque spécialiste, muni de ses propres grilles d'analyse, y découvrira des ressources différentes. Naturellement, l'opération de transcription ne permet pas de respecter l'intégrité de la source originale, dans la mesure où elle implique la

²⁰² *Ibid.*, *op. cit.*, p. 132-133

²⁰³ *Ibid.*, *op. cit.*, p. 131

perte des informations relatives à la disposition précise des éléments, à la texture du support, à la couleur des encres, etc. Mais une transcription peut néanmoins rendre compte avec fidélité de certaines informations graphiques, comme la présence dans le texte d'une ligature, d'un signe abrégatif, d'une variante allographétique... Une transcription idéale pourrait donc servir de base à des analyses non seulement littéraires, lexicales, syntaxiques, mais aussi à des études orthographiques ou paléographiques. Elle serait à ce titre *définitive*. Le texte ainsi établi constituerait une base solide à laquelle chacun pourrait apporter, moyennant quelques modestes transformations, les enrichissements nécessaires à ses propres travaux (balisage sémantique, encodage de structures grammaticales, lemmatisation, regroupement des entités allographétiques, etc.). »²⁰⁴

La recherche du fragment, la découverte du Tout-venant

Avec le mode texte, ce ne sont plus seulement les métadonnées ou la table des matières retapée manuellement qui sont référencées, c'est le contenu de l'ouvrage dans son ensemble qui est indexé et donc moissonnable par les moteurs de recherche. Il devient ainsi possible, en un temps incroyablement bref, de fouiller une quantité astronomique de fichiers. Cela représente une manière nouvelle de faire de la recherche qui vient réellement bouleverser les pratiques historiennes. Le « lecteur » peut en effet chercher, dans une œuvre ou dans un corpus, un terme ou une phrase qui l'intéresse, et être renvoyé automatiquement aux pages où figurent les termes de sa recherche. Si cela ouvre pour certains des opportunités sans précédent de travail sur les sources et sur l'histoire de la langue, d'autres historiens expriment un bien plus profond scepticisme.

C'est que cette nouvelle manière d'appréhender le texte comme agrégat de termes linguistiques encourage le lecteur, historien compris, à ne s'intéresser qu'aux fragments de texte qu'un moteur de recherche aura sélectionnés pour lui, suite à une recherche lexicométrique. Le texte se voit ainsi désolidarisé de son contenant, d'une part, à savoir l'objet-livre, et du texte dans son ensemble, d'autre part. Il n'est plus un Tout signifiant, organisé suivant un schéma démonstratif soigneusement établi par son auteur ; il est *fragment*. Ce n'est pas seulement le travail bibliographique qui en est transformé – on peut en effet y voir un moyen efficace d'élargir les corpus d'étude –, c'est le travail sur les sources lui-même qui évolue et qui pourrait pâtir de cette décontextualisation de l'élément sémantique. La considération de l'information textuelle comme imbriquée dans un ensemble

²⁰⁴ Rémi JIMENES, « Introduction », dans « Gestion informatisée des écritures anciennes », BENEVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 8

signifiant plus large oblige au contraire le lecteur à évaluer cette information, à ne pas l'utiliser à sa guise sous couvert d'une autorité auctoriale citée à mauvais escient. La recherche fragmentaire au contraire incite à multiplier les références et les citations sans tenir compte de l'origine et donc de la fidélité du propos.

L'autre biais induit par cette recherche informatisée consisterait en un « lissage » du paysage patrimonial, c'est-à-dire qu'une recherche fragmentaire tend à gommer la spécificité des documents et la légitimité des textes considérés. L'information est arrachée à l'ensemble textuel dans lequel elle fait sens qui est lui-même dissocié de son support matériel, support matériel qui peut être, dans le monde du livre, hautement signifiant. L'information considérée n'a pourtant pas la même valeur si elle est tirée, par exemple, d'une monographie scientifique, d'un journal ou d'un écrit intime.

L'information est, par nature, hautement contextuelle : orientée, politisée, rattachée à un temps donné et à son contexte de production. Le numérique et l'écran tendent à l'inverse à aplanir le paysage informatif et à niveler la valeur générale des sources.

Le « Tout-venant » informatique sous sa forme fragmentaire nécessite donc une approche critique de l'information-source et un travail constant de recontextualisation historique. Ce n'est qu'ainsi que l'historien pourra user correctement de l'Internet et de ses bases de données et profiter, par là-même, des potentialités énormes qu'ils peuvent offrir à qui est à même d'interroger la valeur de l'information.

La mort de l'objet-livre, essai de philologie numérique

Une telle considération nous amène à interroger le rapport qui peut unir ou opposer le *texte* et le *livre*, considéré comme objet. Cette dualité texte-livre en recoupe d'autres, que sont notamment l'opposition entre *donnée* et *document* ou l'antinomie *numérique-analogique*. Ces notions invitent à porter un regard de sémiologue sur le monde de la philologie numérique où se rencontrent science et technique.

La *philologie* se définit comme l'étude du langage à partir de documents écrits, à la rencontre entre critiques littéraire, historique et linguistique.²⁰⁵ Le numérique, s'il bouleverse les modalités d'appropriation de l'écrit, c'est-à-dire la lecture, n'en présente pas moins des documents écrits, bien qu'il ait tendance, nous l'avons vu précédemment, à désolidariser le contenu de son contenant. Cette

²⁰⁵ Voir la page Wikipédia « Philologie », consultable à l'adresse suivante : <https://fr.wikipedia.org/wiki/Philologie>

désolidarisation texte-support, avec tous les problèmes qu'elle pose en termes d'intégrité et d'authenticité, force l'historien à repenser le rapport qui peut exister entre la *science* – du latin « *scire* » : savoir, connaître – et la *technique* qui conditionne la « possibilité de la rétention des souvenirs »²⁰⁶. Car la *donnée textuelle*, qui est *savoir*, s'autonomise avec le numérique du *document-livre*, qui est production technique. François Rastier, dans *La sémiotique des textes, du document à l'œuvre*, le rappelle : « la notion philologique de *document* est liée à une inscription sur un support doué d'une stabilité temporelle, en général un objet mobilier. [...] Restitué à chaque affichage, le document numérique n'a pas la continuité ni la stabilité matérielle des documents traditionnels. »²⁰⁷. La révolution numérique appelle donc à réévaluer les frontières traditionnelles qui distinguent linguistique, philologie numérique et herméneutique matérielle pour considérer dans leur interdépendance les notions de *document*, de *texte* et d'*œuvre*²⁰⁸. Il s'agit par là-même de ne pas tomber dans la facilité des *lectures industrielles*²⁰⁹, travail de déchiffrement effectué automatiquement par des programmes de recherche d'information. Il faut au contraire réapprendre à ne pas, par un culte de la *donnée*, mettre à mort le *document* comme signifiant historique. Ce que l'on appelle le « Web sémantique » conduirait ainsi à terme à substituer au « Web des documents » le « Web des données »²¹⁰ où l'utilisateur ne considérerait plus qu'une toile immense mettant en relation des mots clés – les fameuses ontologies – désolidarisés des œuvres desquelles ils sont issus. François Rastier met ainsi en garde le lecteur numérique devant ce qu'il nomme un phénomène de *surdocumentation*, « processus qui consiste à remplacer un texte par ses métadonnées documentaires et sa lecture par l'accès à ces métadonnées »²¹¹. Ce culte des données auquel participent pleinement les politiques de numérisation en mode texte, s'il n'est nullement utile de le diaboliser, ne doit donc pas supplanter l'éternel travail historique d'évaluation et de critique du texte comme partie prenante d'un *support* et d'une *œuvre* hautement contextuels. Le *texte-livresque*²¹² et l'objet-livre ne doivent pas disparaître devant le rêve de l'immédiate accessibilité de l'information et jamais l'historien ne doit faire l'économie des démarches philologiques préalables à toute étude de source. Les problématiques de

²⁰⁶ Valentine FREY et Matteo TRELEANI, « Introduction » à *Vers un nouvel archiviste numérique...*, p. 11

²⁰⁷ François RASTIER, « La sémiotique des textes, du document à l'œuvre », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.), p. 21

²⁰⁸ *Ibid.*, *op. cit.*, p. 25

²⁰⁹ *Ibid.*, *op. cit.*, p. 32

²¹⁰ *Ibid.*, *op. cit.*, p. 53

²¹¹ *Ibid.*, *op. cit.*, p. 55

²¹² Ce terme, emprunté à Ivan Illich qui en inaugure l'usage au sujet des manuscrits du XII^e siècle, sert à qualifier un objet de l'écrit « doté d'une configuration formelle précise, il se concrétise dans un dispositif médiatique défini et se comprend au sein d'une discipline intellectuelle déterminée ». Voir Yves JEANNERET, « Le procès de numérisation de la culture, un défi pour la pensée du texte. », *Protée*, vol. 32, n°2, 2004, p. 12 qui cite Ivan ILLICH, *Du lisible au visible : la naissance du texte*, Paris, Le Cerf, 1991, 150 p.

l'indexation ne peuvent l'emporter sur celles du déchiffrement, de la lecture et de l'interprétation²¹³.

La dématérialisation de l'information interroge donc la nature même de l'information textuelle. « Là où les documents médiatiques étaient toujours fixés sur des supports matériels, dans le cas des médias technologiques tout comme dans les médias perceptifs, les documents numériques sont des codes qui peuvent être répliqués à l'infini. Autrement dit, le numérique impose un décodage du contenu pour permettre sa consultation. »²¹⁴, nous rappellent Valentine Frey et Matteo Treleani. Ce décodage passe par des machines et des outils numériques qui soulèvent de nouveaux enjeux relatifs au traitement des données. La mécanisation des tâches de l'historien induite par le numérique semble en réalité faire trembler les fondations de la discipline elle-même.

²¹³ François RASTIER, « La sémiotique des textes, du document à l'œuvre », dans *Vers un nouvel archiviste numérique*, Valentine FREY et Matteo TRELEANI (coord.), p. 55

²¹⁴ Valentine FREY et Matteo TRELEANI, « Introduction » à *Vers un nouvel archiviste numérique...*, p. 10

TRAVAILLER AVEC CES SOURCES

L’outil informatique offre des moyens nouveaux de travail sur les sources historiques. Ces nouvelles méthodes de traitement de l’information interrogent les fondements mêmes de la discipline et questionnent le rapport qu’entretient le chercheur avec les données désormais accessibles en masse sur l’Internet. Considérant l’héritage scriptural des sociétés contemporaines, et non leurs productions présentes, notre domaine d’étude concerne quasi exclusivement les bases de données numérisées et ne s’étend pas au Web dans son ensemble, qui ouvre de bien plus vastes champs d’exploration.

FAUT-IL SE JOUER DES DONNÉES ?

L’utilisation des données que l’historien peut désormais extraire des bases en ligne divise la communauté historique. Nombreux sont en effet aujourd’hui les logiciels informatiques qui permettent d’accompagner le travail de l’historien, sinon de l’effectuer à sa place, quant au traitement quantitatif des données. Certains voient ainsi dans l’informatisation de l’analyse historique une forme de décrédibilisation de l’activité historique qui donne à croire que le numérique est à l’origine d’une véritable crise disciplinaire.

Le traitement des données

Ces outils informatiques de traitement des données s’appliquent aux sciences humaines et sociales dans leur ensemble, dès lors qu’elles concernent la transmission écrite de l’information, passé comme présente. Nous restreindrons cette étude aux sources passées de l’écrit, bien que les méthodes envisagées puissent s’appliquer à l’information contemporaine d’une manière quasi analogue. Plus encore, elles semblent s’appliquer *de facto* à l’écrit nativement numérique mais peuvent être étendues au monde de l’écrit numérisé – en mode texte, bien entendu. Nombre de ces outils concernent la *statistique*, la *lexicométrie* et la *cartographie*.

Parmi les nombreux outils à disposition de l’historien qui souhaiterait traiter d’importantes masses de données, et indépendamment de sa capacité à constituer

des corpus pertinents, nombre d'entre eux fonctionnent sur un principe que l'on appellera de « lemmatisation » qui consiste à indexer les mots importants d'un texte et à les regrouper au sein de « familles » signifiantes que les moteurs de recherche sont ensuite à même de rechercher, de comptabiliser et de présenter à l'utilisateur numérique au gré de ses requêtes. Cette indexation automatisée est partie intégrante des processus d'océrisation, désormais maintes fois évoqués, qui tendent à transformer le texte en son propre index lexical. Partant de cette auto-indexation des contenus numérisés et océrisés, les textes deviennent traitables par des logiciels informatiques dits lexicométriques, relatifs, donc, à la lexicométrie : soit la science qui étudie statistiquement l'usage des mots²¹⁵. À partir de ces comptabilisations lexicales peut être représentée l'information, sous forme de tableaux, de graphiques voire d'imageries 3D.

Considérons par exemple les travaux de Yannick Rochat et de Frédéric Kaplan sur « L'analyse des réseaux de personnages dans les *Confessions* de Jean-Jacques Rousseau »²¹⁶. Partant de l'ouvrage au format texte, qui permet une lecture informatisée, les deux auteurs effectuent une analyse de réseaux en littérature, qui « étudie l'organisation des entités narratives au moyen de la théorie des graphes »²¹⁷. C'est-à-dire qu'ils construisent, à partir de la cooccurrence du nom des personnages, un schéma graphique destiné à rendre compte de la représentativité des différents personnages et de leur mise en relation dans l'intrigue de l'œuvre. Il suffit pour cela d'utiliser l'index, lui-même résultat d'un processus informatique automatisé, et d'user du concept de *centralité*, qui signifie ici que chaque personnage est considéré comme le nœud d'un tissu relationnel qu'il centralise à plus ou moins grande échelle. Ils obtiennent ainsi un « réseau des cooccurrences des personnages » (**figure 36**) à partir duquel ils sont en mesure de construire des graphiques et de formuler des hypothèses quant au rôle joué par tel ou tel personnage dans l'action en fonction de sa position dans le réseau.

²¹⁵ Voir la page Wikipédia « Lexicométrie », consultable à l'adresse suivante : <https://fr.wiktionary.org/wiki/lexicom%C3%A9trie>

²¹⁶ Yannick ROCHAT et Frédéric KAPLAN, « Analyse des réseaux de personnages dans les *Confessions* de Jean-Jacques Rousseau », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 109-133

²¹⁷ Claude BERGE, *Théorie des graphes et ses applications*, Paris : Dunod, 1958, 267 p., cité par Yannick ROCHAT et Frédéric KAPLAN dans « Analyse des réseaux... », p. 111

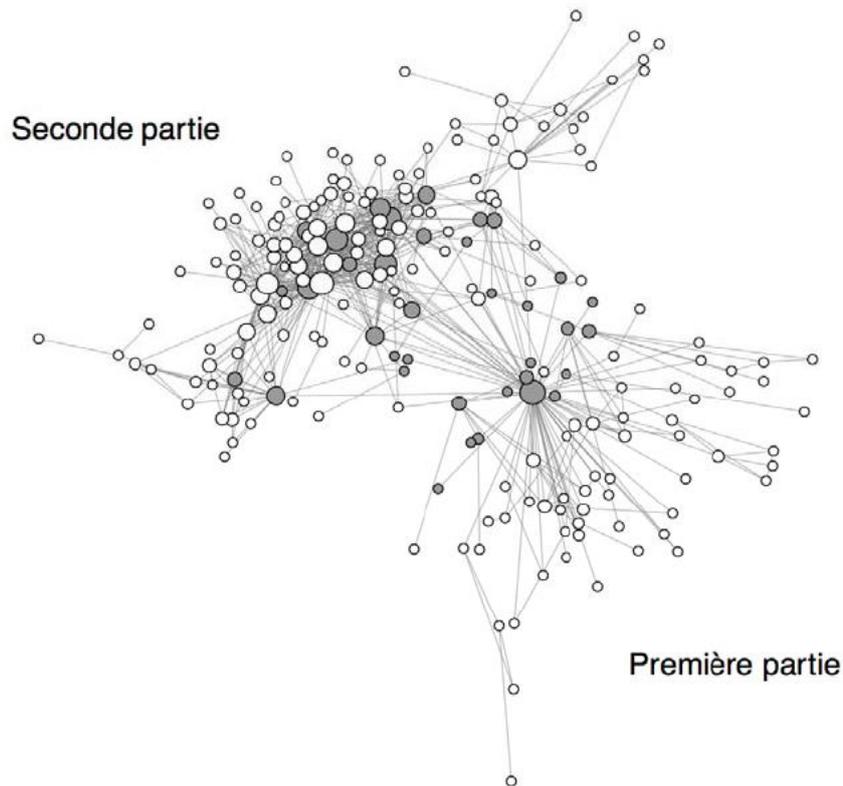


Figure 36. Réseau des cooccurrences des personnages des *Confessions* de Jean-Jacques Rousseau. Image empruntée à Yannick ROCHAT et Frédéric KAPLAN²¹⁸.

La méthode n'est cependant pas uniquement quantitative puisqu'elle tient compte d'une connaissance *a priori* de l'intrigue qui oriente les hypothèses évaluées. Les chercheurs émettent ainsi eux-mêmes un doute sur son application possible à plus grande échelle : « Une question plus difficile est de savoir si une approche strictement quantitative, n'utilisant pas les connaissances spécifiques sur l'œuvre mais seulement l'analyse structurelle des réseaux de personnages est concevable. [...] Les résultats que nous proposons dans cet article ne nous permettent pas de conclure sur la pertinence intrinsèque d'une approche uniquement quantitative, et donc applicable à des grands corpus d'œuvres. »²¹⁹.

Un autre outil de traitement de l'information, très peu renseigné en définitive, est évoqué par Jo Guldi et David Armitage dans « Le retour de la longue durée » ; il s'agit des *geoparsers*. Ils consistent en des « algorithmes qui, en soumettant chaque mot à un index géographique, génèrent une carte des lieux cités dans un corpus donné. »²²⁰ Cet outil de traitement, spatial également, relie donc la littérature au sens large et la cartographie.

²¹⁸ Yannick ROCHAT et Frédéric KAPLAN, « Analyse des réseaux de personnages dans les Confessions de Jean-Jacques Rousseau », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 120 (Figure 1).

²¹⁹ *Idem*, *op. cit.*, p. 129-130

²²⁰ Jo GULDI et David ARMITAGE, « Le retour de la longue durée, une perspective anglo-américaine », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 291

Considérons finalement deux autres outils évoqués par les mêmes auteurs que sont Wordle²²¹ et Paper Machines²²² : « En partant de la fréquence d'occurrence des mots, un *wordle* permet de visualiser les thèmes d'un texte et de les hiérarchiser. Les mots les plus fréquents sont affichés au centre de la page et leur police est augmentée. »²²³ Il est d'ailleurs étonnant que les deux auteurs citent un tel programme qui se présente lui-même comme un « jouet »²²⁴ et qui relève bien plus de la pédagogie que de la recherche historique. (**figure 37**) Paper Machines enfin « est un plug-in conçu pour le logiciel Zotero²²⁵. Il permet d'appliquer aisément wordles, ngrams et autres outils de visualisation aux corpus réunis manuellement par les historiens sur leur ordinateur. »²²⁶



Figure 37. Exemple de Wordle réalisé sur l'application en ligne *wordle.net* à partir de notre partie sur « Le vrac » (p. 23-25). Notons que les commandes « *Remove common French words* » et « *Remove common Names Words* » ne fonctionnent pas correctement et que l'usage d'une telle application en est d'autant plus limitée.

De tels exemples illustrent les usages possibles pour l'historien des outils de collecte et de traitement de l'information contenue en masse sur la Toile. Il force pourtant à émettre des doutes sur l'autosuffisance de ces méthodes qui, si elles

²²¹ Le site est consultable à l'adresse suivante : <http://www.wordle.net/>

²²² Le site est consultable à l'adresse suivante : <http://papermachines.org/>

²²³ Jo GULDI et David ARMITAGE, « Le retour de la longue durée... », p. 312

²²⁴ Voir sur la page d'accueil du site : « *Wordle is is a toy for generating "word clouds" from text that you provide.* »

²²⁵ Sur ce sujet voir l'article de Frédéric Clavert posté sur le Blog Zotero francophone, consultable à l'adresse suivante : <http://zotero.hypotheses.org/476>.

²²⁶ Jo GULDI et David ARMITAGE, « Le retour de la longue durée... », p. 312

peuvent illustrer une hypothèse de recherche, ne remplacent aucunement le travail préparatoire à toute recherche historique. Qu'illustre le réseau de personnages de Rochat et Kaplan si ces derniers ne l'expliquent ni ne l'interprètent ? Le non-sens du *tout-quantitatif* est en tout cas un des arguments majeurs des détracteurs de cette mouvance mi-scientifique mi-littéraire, que les plus critiques associent à une forme grave de « quantophrénie ».

« Culturomique » et « quantophrénie »

La *quantophrénie* sert à qualifier l'excès d'usage mathématique en sociologie, sorte de maladie de la mesure qui affecterait les amateurs de données chiffrées en sciences humaines et sociales.²²⁷ Cette accusation, portée par les opposants à l'usage des outils informatiques de traitement de l'information, s'adresse en particulier aux historiens déterminés à faire tomber la frontière qui sépare sciences dures et sciences « molles » et qui souhaiteraient faire de l'usage du numérique un terrain nouveau pour la discipline historique. Si elle stigmatise donc les utilisateurs des outils de traitement automatique de l'information évoqués précédemment, la diabolisation de la « quantophrénie » est surtout née en réaction à une application informatique : Ngram Viewer, et à un article scientifique : « *Quantitative Analysis of Culture Using Millions of Digitized Books* »²²⁸, publié en 2011 par l'équipe ayant travaillé sur l'application.

Ngram Viewer²²⁹ est une application informatique associée à Google Livres depuis 2010, qui sert à l'analyse statistique de la langue. Conçue comme un moteur de recherche standard, où l'on rentre un ou plusieurs mots dont on souhaite évaluer l'usage à travers le temps dans les sources imprimées contenues dans Google Livres, l'application construit ensuite, quasi instantanément mais sans information aucune sur le fonds moissonné, un graphique représentant la fréquence d'apparition du ou des termes dans le corpus consulté, en fonction du temps²³⁰ (figures 38 et 39). De fonctionnement très simple – trop, diront beaucoup, pour qu'il soit un outil historien –, les critères de recherche s'y résument à la période étudiée et à la langue du corpus consulté. Le nom du bouton de commande de la recherche, « *Search lots of books* »²³¹, est révélateur du flou dans lequel l'utilisateur est tenu relativement à la constitution de ce corpus. On imagine

²²⁷ Voir la page du Wiktionnaire « quantophrénie » consultable à l'adresse suivante : <https://fr.wiktionary.org/wiki/quantophr%C3%A9nie>

²²⁸ Jean-Baptiste MICHEL, Yuan KUI SHEN *et al.*, « *Quantitative Analysis of Culture Using Millions of Digitized Books* », *Science*, vol. 331, n° 6014, 14 janvier 2011, p. 176-182

²²⁹ L'application est consultable à l'adresse suivante : <https://books.google.com/ngrams>.

²³⁰ Voir la page Wikipédia « Ngram Viewer » consultable à l'adresse suivante : https://fr.wikipedia.org/wiki/Ngram_Viewer

²³¹ « Rechercher dans beaucoup de livres »

cependant qu'il se limite au fonds même de Google Livres, aux ouvrages numérisés en mode texte et que certains fonds linguistiques sont particulièrement limités. On peut de plus affirmer, sans risque de se tromper, que la fiabilité de l'outil est dépendante de la période étudiée, les ouvrages anciens étant, on le sait, non seulement rares mais également très mal océrisés.

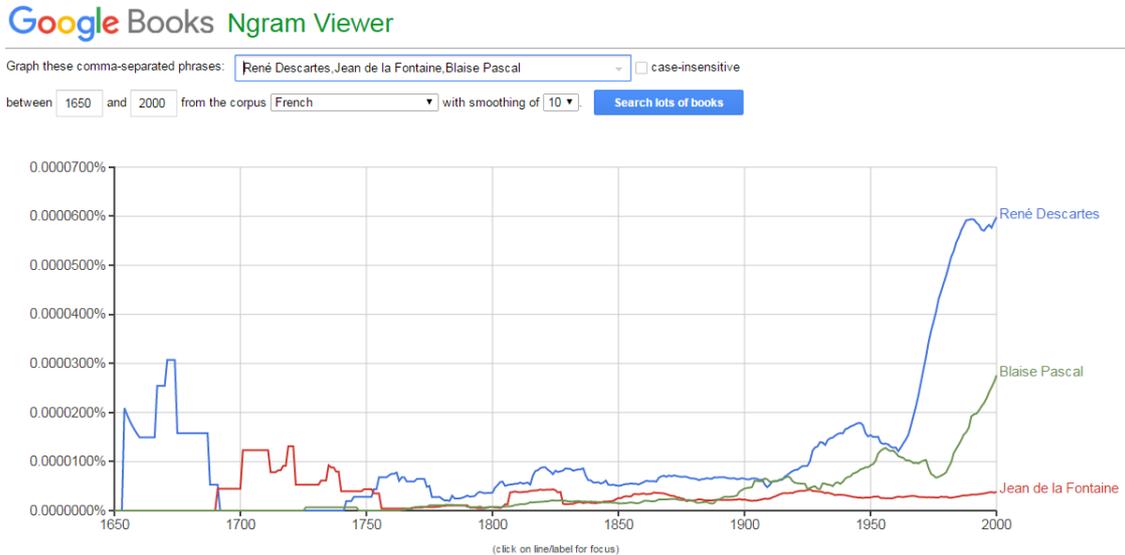


Figure 38. Exemple d'utilisation de l'application Ngram Viewer, sur la cooccurrence des termes « René Descartes », « Jean de la Fontaine » et « Blaise Pascal » depuis 1650 dans les ouvrages français numérisés par Google Livres.

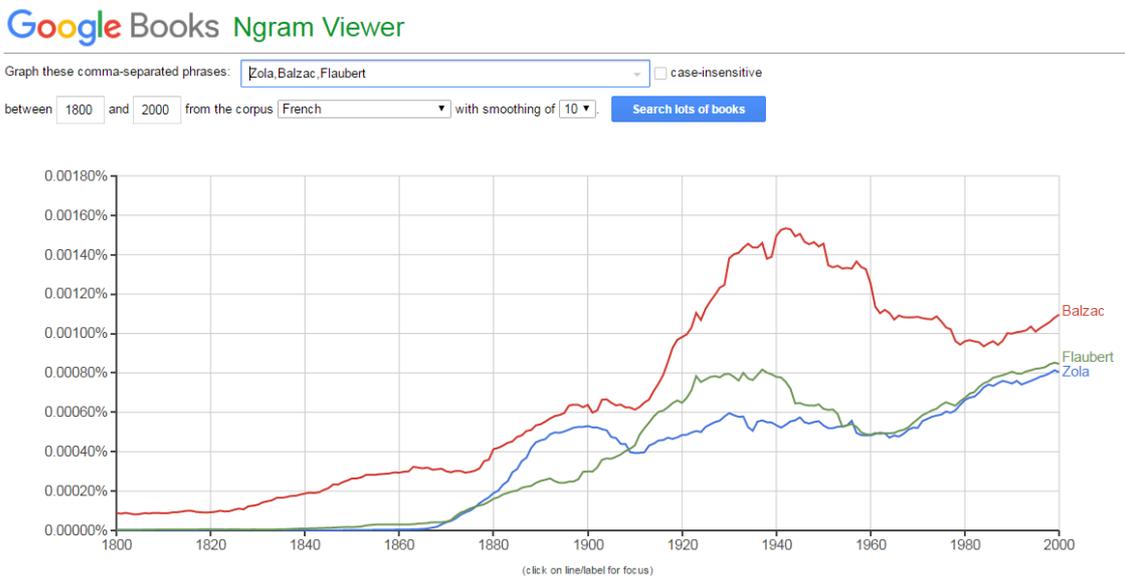


Figure 39. Exemple d'utilisation de l'application Ngram Viewer, sur la cooccurrence des termes « Zola », « Balzac » et « Flaubert » depuis 1800 dans les ouvrages français numérisés par Google Livres.

L'année suivant la mise en ligne de l'application, le 14 janvier 2011, l'équipe ayant travaillé sur sa réalisation publie un article dans la revue *Science*, intitulé « Analyse culturelle quantitative à partir de millions de livres numérisés » : les auteurs y décrivent l'usage de Ngram Viewer et son application possible à l'analyse de l'évolution des pratiques culturelles par l'étude chiffrée des usages linguistiques – par *usage linguistique*, entendons ici simple cooccurrence d'un terme ou d'un groupe de mots dans un corpus fermé. À la croisée des courants de la sociologie dite « quantitative » et de l'histoire baptisée « de longue durée », les auteurs décrivent l'application comme un outil révolutionnaire en sciences humaines et sociales.

Résumons brièvement l'article : partant d'un corpus de textes numérisés représentant quelques 4 % de la production imprimée globale²³², les auteurs se sont concentrés sur les phénomènes linguistiques et culturels reflétés par la langue anglaise entre 1800 et 2000 pour en déduire quantitativement des « tendances culturelles ». Les champs étudiés concernent ainsi aussi bien la lexicographie²³³ que la grammaire, la mémoire collective, la technologie, le succès, la censure ou l'épidémiologie²³⁴. Citons les auteurs concernés : « *Culturomics extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities* »²³⁵. Nous passerons sur l'étude des données présentées dans l'article pour n'en évoquer que quelques conclusions : la langue anglaise a vu son lexique s'enrichir de près de 70 % dans la seconde moitié du XX^e siècle ; les sociétés cessent progressivement de s'intéresser à leur passé ; plus nous nous rapprochons des temps présents, plus le succès est rapide mais de courte durée ; on peut repérer les effets de la censure politique par la disparition sur une période donnée et dans un corpus linguistique ciblé d'un nom ou d'un terme cooccurrent dans les productions d'autres pays. Les auteurs concluent finalement à l'avènement d'une discipline nouvelle en sciences humaines et sociales, la « *culturomics* » – que nous traduirons par « culturomique »²³⁶ – qui consiste en l'étude des tendances culturelles au regard de données chiffrées collectées en grande nombre et informatiquement étudiées. « *Culturomics is the application of high-throughput data collection and analysis to*

²³² Les auteurs parlent de 5 195 769 ouvrages considérés, sur les « plus de quinze millions d'ouvrages [qui] ont déjà été numérisés (soit environ 12 % de l'ensemble des livres jamais publiés). Pour l'analyse, nous en avons sélectionné plus de cinq millions en fonction de la qualité de leur océrisation et de leurs métadonnées. Nous avons exclu les périodiques » (p. 176), soit plus de 500 milliards de mots (361 millions sont anglais, 45 milliards français, 45 milliards espagnols, 37 milliards allemands, 13 milliards chinois, 35 milliards russes, 2 milliards hébreux). Les ouvrages les plus anciens datent du XVI^e siècle (*ibid.*).

²³³ La lexicographie est l'étude scientifique et analytique des faits de lexique.

²³⁴ L'épidémiologie est l'étude des facteurs influant sur la santé et les maladies de populations.

²³⁵ « La culturomique étend les frontières de l'enquête quantitative rigoureuse à un vaste ensemble de phénomènes nouveaux couvrant les sciences sociales et les humanités » (Jean-Baptiste MICHEL, Yuan KUI SHEN *et al.*, « *Quantitative Analysis of Culture...*, p. 176)

²³⁶ Traduction empruntée à Pierre Mounier dans l'Avant-propos à *Read/Write Book 2. Une introduction aux humanités numériques*, OpenEdition Press, 2012, 262 p. 15

the study of human culture. »²³⁷. « *These [...] [trends] will furnish a great cache of bones from which to reconstruct the skeleton of a new science.* »²³⁸

Si les auteurs sont pleins d'optimisme et l'article plutôt enthousiasmant, les critiques qui ont fait suite à sa publication ont cependant été nombreuses et virulentes. Il en ressort un scepticisme profond vis-à-vis de cette « culture du chiffre » que certains baptiseront « quantophrénie » – nous pensons à Michel Wieviorka – ; les « quantosceptiques » y voient comme une offense faite à la rigueur démonstrative de la discipline historique. Citons abondamment *L'impératif numérique*²³⁹ afin de donner un aperçu des réactions historiques qui ont fait suite à ce petit manifeste de la culturomique :

« Ce type d'outil autorise de procéder à ce que des historiens ont appelé le *text mining* et donc de plonger dans de gigantesques archives et autres bibliothèques numériques comme le fait dans une mine un chercheur d'or ou de tout autre minerai, amateur ou professionnel. On peut désormais aussi chercher de minuscules aiguilles dans de gigantesques bottes de foin [...] Les outils de *text mining* existent depuis longtemps [...] Google en propose une version appauvrie « clé en main » et surtout sans la théorie qui l'accompagne : c'est la technologie sans la science. [...] La recherche fondée sur le *text mining* est scientifiquement élémentaire et ne saurait pour l'instant trouver sa place qu'à la marge dans la vie intellectuelle. Il entre même une part de naïveté dans l'arrogance de certains de ses pionniers, dont l'enthousiasme repose trop souvent sur l'illusion de l'objectivité qu'apporterait la quantification – une logique qui pourrait, pour utiliser le mot du sociologue Pitirim Sorokim²⁴⁰, tourner à la « quantophrénie ». Les données utilisées sont elles-mêmes une production humaine et non un fait de nature, elles ont été construites d'une certaine façon, variant culturellement dans le temps et l'espace, dans un contexte historique donné, à certaines fins, etc. : avec les Big Data, va-t-on se dispenser de réfléchir aux questions d'échantillonnage, de représentativité, d'homogénéité du corpus, de conditions dans lesquelles les données sont produites, agrégées ou combinées ? »²⁴¹.

De telles accusations ne sont évidemment pas vides de sens et rendent compte d'une réactualisation contemporaine du débat qui oppose adeptes et détracteurs du quantitatif.

²³⁷ « La culturomique consiste en la collection des données à haut débit et leur analyse pour l'étude de la culture humaine » (Jean-Baptiste MICHEL, Yuan KUI SHEN *et al.*, « *Quantitative Analysis of Culture...*, p. 181)

²³⁸ « Ces [...] [tendances] fourniront une réserve d'os avec lesquels reconstruire le squelette d'une science nouvelle. » (*Ibid.*, *op. cit.*, p. 182)

²³⁹ Michel WIEVIORKA, *L'impératif numérique, ou la nouvelle ère des sciences humaines et sociales ?*, Paris : CNRS Editions, 2013, 64 p.

²⁴⁰ Pitirim Aleksandrovich SOROKIN, *Fads and Foibles in Modern Sociology and Related Sciences*, Chicago: H. Regnery Co., 1956, 357 p.

²⁴¹ Michel WIEVIORKA, *L'impératif numérique...*, p. 25-26

L'affrontement entre « quantitatif » et « qualitatif »

Car si la culturomique, de par son caractère « quantophrénique », est l'abcès de fixation des grands débats historiographiques contemporains, c'est qu'elle ravive un conflit qui ne date pas d'hier entre « quantitatif » et « qualitatif ». Cette antinomie recoupe d'autres grands courants intellectuels tels que le *prédicatif*, opposé à l'*explicatif*, ou le « *close reading* » opposé au « *distant reading* »²⁴².

Les adeptes du « qualitatif » répondront au culte de la donnée chiffrée chère aux partisans du quantitatif par la rigueur de la démonstration théorique. Les premiers défendent donc la puissance analytique de l'intellection ; les seconds la force déductive des mathématiques. Cet affrontement cristallise une vieille querelle quoique très abstraite entre les sciences de l'esprit, les SHS²⁴³, et les sciences de la nature considérées au sens large ; c'est la distinction aujourd'hui encore admise dans l'enseignement académique entre Lettres et Sciences, entre lettres et chiffres. Une telle différenciation est d'autant plus abstraite, et critiquable, qu'elle tend à nier l'interpénétration et l'interdépendance de ces deux domaines dont il est pourtant difficile de douter : l'Histoire use autant du Nombre que les Mathématiques du Verbe. Si nous n'évoquons pas dans le détail les modalités de ce débat, il nous amène cependant à considérer deux approches que l'on voudrait distinctes de la donnée et du corpus historiques : le *close reading* et le *distant reading*.

Ces deux approches de la source historique consistent pour la première à considérer le texte au plus près de sa matérialité, la seconde comme le fragment d'un ensemble textuel plus vaste. Elles ne sont donc pas sans évoquer la distinction fragment-Tout déjà introduite précédemment. Même si cela peut sembler réducteur, le *close reading* se rapporte à l'analyse qualitative du texte dans son caractère d'exception et dans la signifiante historique du détail ; elle peut se rapprocher en cela de l'approche « littéraire » de la source. Le *distant reading* au contraire considère que la masse informative fait sens et qu'elle peut donner lieu à des analyses mathématiques quantifiées à partir desquelles la donnée en tant qu'objet fait sens. Ce serait là donc là une approche plus « scientifique » de la source. En réalité, l'affrontement quantitatif-qualitatif interroge aujourd'hui la place à accorder aux données chiffrées en SHS, au regard des outils informatiques disponibles. Une question peut résumer ce débat, en le caricaturant : peut-on finalement utiliser Ngram Viewer en Histoire ?

²⁴² Expressions que l'on peut traduire ici par « lecture proche » ou « fine » et « lecture distante » ou « distancée ». Certains choisissent de la traduire par « lecture de près » et « lecture de loin ». Le concept de *distant reading* est introduit par Franco MORETTI dans son article « *Conjectures on World Literature* », concept qui donnera ensuite son nom à une œuvre rassemblant dix de ses articles : *Distant Reading*, London ; New York : Verso, 2013, 244 p.

²⁴³ Nous emploierons cette abréviation admise pour parler des Sciences Humaines et Sociales.

Le *text mining*, évoqué ci-dessus par Michel Wieviorka, qui consiste à déléguer à une machine le soin de chercher « de minuscules aiguilles dans de gigantesques bottes de foin »²⁴⁴ met en cause les méthodes historiques de constitution et d'analyse de corpus. « Va-t-on se dispenser de réfléchir aux questions d'échantillonnage, de représentativité, d'homogénéité du corpus, de conditions dans lesquelles les données sont produites, agrégées ou combinées ? »²⁴⁵, s'interroge-t-il.

Cette problématique déjà pointée n'a de sens que si l'historien n'a pas la main sur l'outil informatique dont il se sert. En cela la critique de Michel Wieviorka est pertinente, car si lors de la consultation des bases de données le chercheur a droit de regard sur le matériau qu'il consulte et sélectionne, Ngram Viewer n'offre pas cette opportunité. Pour cette raison, un graphique produit par l'application n'a en effet aucune légitimité à faire source : il n'a pas valeur de preuve, ne clame pas d'évidence historique. Un graphique ne remplace pas l'historien puisqu'il n'est pas producteur de sens : il n'est qu'une illustration de son travail, et peut tout au mieux orienter *a priori* ses perspectives de recherche, ou les confirmer *a posteriori*. De cela, il semble cependant que les concepteurs de Ngram Viewer aient conscience : « *Of course, many voices – already lost to time – lie forever beyond our reach. Culturomics results are a new type of evidence in the humanities. As with fossils of ancient creatures, the challenge of culturomics lies in the interpretation of this evidence.* »²⁴⁶. Le terme de « *evidence* » qu'ils utilisent se traduit en effet à la fois par « preuve », « donnée » ou « témoignage ». S'opposer trop catégoriquement à l'interrogation informatisée des bases de données reviendrait par ailleurs à oublier qu'elles ne sont pas figées. Philippe Rygiel, qui reprend la métaphore de la botte de foin, semble ainsi répondre à Michel Wieviorka : « Les historiens, de longue date habitués à chercher les aiguilles dissimulées dans les bottes de foin, sont particulièrement sensibles à cela. Ils semblent moins intéressés par le fait que ces mêmes transformations autorisent à remodeler les bottes de foin et à en proposer des descriptions inédites. »²⁴⁷ Si Ngram Viewer fonctionne selon un principe pour l'instant très critiquable car trop opaque, cela ne signifie pas que l'analyse sémantique informatisée que l'application propose ne sera pas l'objet d'un nouveau programme qui en permettrait un usage plus raisonné. Pensons au plug-in de l'application Zotero *Paper Machines* déjà abordé qui devrait permettre de restreindre l'usage de Ngram Viewer à un corpus de textes manuellement constitué par son utilisateur.

²⁴⁴ Michel WIEVIORKA, *L'impératif numérique...*, p. 25, expression déjà citée en note 197.

²⁴⁵ *Ibid.*, *op. cit.*, p. 26, idem.

²⁴⁶ « Bien sûr, de nombreuses voix – déjà perdues dans la nuit des temps – se trouvent pour toujours hors de portée. // Les résultats de la culturomique représentent un nouveau genre de donnée dans les humanités. Comme avec les fossiles d'antiques créatures, le défi de la culturomique consiste à interpréter cette donnée. » (Jean-Baptiste MICHEL, Yuan KUI SHEN *et al.*, « *Quantitative Analysis of Culture...*, p. 181)

²⁴⁷ Philippe RYGIEL, « Écriture de l'histoire et réseaux numériques », dans *Regards croisés sur l'Internet*, Éric GUICHARD (dir.), Villeurbanne : Presses de l'Enssib, 2011, p. 119

Relativement au recul et à l'attitude critique nécessaires à toute approche scientifique, quel historien doute encore de la nécessité d'*évaluer* et d'*interpréter* un témoignage historique ? Il y a peut-être, dans la diabolisation de Ngram Viewer, une surinterprétation de l'utilité que l'application se donne. Ngram Viewer ne fait pas œuvre d'historien, elle est tout au plus un outil statistique destiné à lui fournir des données linguistiques sur de grandes tendances culturelles à analyser. Un simple graphique a-t-il déjà fait l'histoire ?

Concluons avec ce propos d'Éric Guichard : « Apparaît ici un phénomène sous-estimé par les tenants du qualitatif : l'excès quantitatif stimule la conceptualisation. Il invite à repenser l'organisation des savoirs, à la refonder et parfois de fond en comble. »²⁴⁸

Une crise historiographique ?

Vouloir substituer le quantitatif au qualitatif revient à substituer à la lecture humaine une lecture que l'on dira « computationnelle », c'est-à-dire réalisée par un ordinateur. On sait pourtant que la lecture est œuvre d'interprétation et qu'elle est un processus étroitement lié à l'intelligence du cerveau humain. On sait *a contrario* que les travaux sur l'intelligence artificielle rendent pour l'instant compte d'une déficience profonde des machines à effectuer des choix non programmés qui relèveraient justement d'un travail d'interprétation contextuelle. Il ne fait nul doute, donc, que la lecture computationnelle n'ait présentement pas grand sens et qu'elle ne peut fonctionner sans supervision humaine. À la « lecture » humaine, il faudrait donc plutôt opposer le « traitement » computationnel, terme moins ambigu, puisque le travail historique ne se résume justement pas au seul traitement des données. La substitution d'un savoir-faire à un autre n'en est pas moins à l'origine d'un changement de régime de littérature²⁴⁹.

La capacité des ordinateurs et programmes informatiques à brasser des masses considérables de données pour en effectuer des traitements chiffrés tend en revanche à encourager les *analyses sérielles* au détriment des approches dites monographiques. L'analyse historique se pense alors sur des périodes temporelles plus larges puisqu'elle peut profiter du numérique pour traiter une quantité bien plus importante de documents. Il ne s'agit plus seulement de « culturomique », centrée sur les chiffres, mais plus largement d'un changement de paradigme

²⁴⁸ Éric GUICHARD, « Introduction » aux *Regards croisés sur l'Internet*, Villeurbanne : Presses de l'Enssib, 2011, p. 13

²⁴⁹ Nous reprenons là une idée que formule Éric GUICHARD dans son « Introduction » aux *Regards croisés sur l'Internet*, Villeurbanne : Presses de l'Enssib, 2011, p. 15

historique : à l'histoire ciblée sur des événements d'une durée brève, le numérique permet de substituer une « histoire de la longue durée ». Adoptant un point de vue anglo-saxon sur la question, c'est du moins ce qu'affirment David Armitage et Jo Guldi avec leur « retour de la longue durée »²⁵⁰. Leur manifeste postule l'existence d'une « crise de la discipline »²⁵¹ historique, un tournant dans l'approche scientifique des sources en lien direct avec les perspectives d'étude ouvertes par le numérique. Se plaçant dans la lignée des *Annales*²⁵², les deux historiens plaident pour un retour à l'histoire de la longue durée, « combinée à la mobilisation des capacités techniques actuelles de traitement de données à grande échelle, que l'on nomme désormais *big data*. »²⁵³. Ce manifeste historiographique a incité de nombreux historiens, notamment européens, à s'exprimer à leur tour, généralement pour critiquer une approche qu'ils jugent réductrice des pratiques historiographiques contemporaines. C'est le signe, du moins, que l'appel au débat n'est pas resté lettre morte et qu'une réflexion s'ouvre aujourd'hui sur la place qui est celle des outils techniques contemporains dans notre manière de faire l'Histoire.

Reprenons rapidement la thèse de David Armitage et Jo Guldi : pour des raisons politiques et techniques²⁵⁴, la longue durée, née avec Fernand Braudel et les *Annales*, fait son grand retour, « avec de nouvelles ambitions, une vigueur renouvelée et la promesse d'une influence encore plus grande »²⁵⁵. Cette vigueur naît de la multiplication des sources et de leurs outils de traitement mis à la disposition des chercheurs par la révolution numérique. « Les bases de données numériques ont rendu accessibles des quantités astronomiques de textes et d'images. Quant aux outils d'analyse, ils ne posent plus problème ; ils sont même particulièrement utiles pour se saisir de ces questions, car ils sont temporels (pensons aux *ngrams* de Google) et spatiaux (c'est le cas des *geoparsers*²⁵⁶). »²⁵⁷. Ils permettent d'étendre les études historiques sur des périodes longues et des territoires étendus, facilitant en effet le traitement de corpus très vastes. À cette histoire de la longue durée, les auteurs opposent la démarche historiographique

²⁵⁰ Les deux auteurs ont à ce propos publié un livre : *The History Manifesto (The History Manifesto)*, Cambridge : Cambridge University Press, 2014, 165 p.) ainsi qu'un article : *Le retour de la longue durée*, traduit et commenté par les *Annales* en 2015 (« Le retour de la longue durée, une perspective anglo-américaine », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 289-318)

²⁵¹ « Editorial », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 285

²⁵² Les *Annales*, fondées en 1929 par Marc Bloch et Lucien Febvre, sont une revue historique consacrée à la recherche en histoire et à son intrication avec les autres sciences sociales. Dans le numéro consacré à ce retour de la longue durée, dans lequel elles cherchent à confronter les opinions et à nuancer le propos de David Armitage et Jo Guldi, les *Annales* refusent cette paternité : « les *Annales* ne partagent pas leur conception de la longue durée » (« Editorial », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 286)

²⁵³ « Editorial », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 286

²⁵⁴ Jo GULDI et David ARMITAGE, « Le retour de la longue durée, une perspective anglo-américaine », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Editions de l'EHESS, 2/2015, p. 290

²⁵⁵ *Ibid.*, *op. cit.*, p. 291

²⁵⁶ Nous reviendrons sur les *ngrams* dans « Culturomique et quantophrénie » en page 97 et sur les *geoparsers* dans « Le traitement des données » en page 93.

²⁵⁷ *Ibid.*

actuellement admise, une manière de faire de la « micro-histoire » plus centrée sur l'événement – personnage, affrontement, zone géographique réduite – et sur des échelles temporelles biologiques comprises entre cinq et cinquante années²⁵⁸. Bien plus théorique, elle se rattache, selon Armitage et Guldi, à l'étude de sources obscures et difficiles à comprendre²⁵⁹ et permet à l'historien d'« afficher son raffinement théorique » et de « prouver sa maîtrise des sources et son engagement sur le terrain », loin des réalités politico-économiques, « comme un astronome dans son observatoire »²⁶⁰.

Depuis les années 90, qui correspondent notamment à l'expansion des technologies informatiques, de rares historiens participeraient au contraire à une redécouverte de l'histoire longue et de ses implications socio-politiques : « *Big is back* »²⁶¹, annoncent les deux historiens ; « associés à de nouveaux logiciels d'analyse, [les] outils numériques sont une invitation à tester des hypothèses historiques sur des périodes de plusieurs siècles. La nature de ces outils et la richesse des textes disponibles permettent une histoire qui soit à la fois de longue durée et fondée sur un travail archivistique. »²⁶². Un tel bouleversement méthodologique invite à réfléchir sur les changements d'échelle, entre micro et macro, entre *short history* et « *big* » *history*.

Enfin, l'idée récurrente de ce manifeste pour un retour de la longue durée concerne la place même dans la société de l'historien, que les deux auteurs aimeraient voir regagner le devant de la scène politique.

Ce texte, qui a fait couler beaucoup d'encre et qui est à l'origine de vives réactions, reprend pourtant des idées qui ne sont pas nouvelles : il est indéniable que le numérique occupe une place croissante dans le travail de l'historien, et ce presque à chaque étape de la réflexion. Sans restreindre la gamme des outils disponibles au seul Ngram Viewer, qui est peut-être le plus critiquable, il apparaît vraisemblable que l'historien manipule de plus en plus régulièrement les technologies informatiques. Cela ne surprend personne : les travaux historiques contemporains ne concernent pas toujours des problématiques socio-politiques cruciales qui viendraient redorer le blason de l'historien comme acteur de la vie politique. Le plus gênant, dans ce plaidoyer, consiste peut-être en son caractère généralisant : une volonté d'étiqueter l'historien et son rapport à l'histoire associée à une volonté certaine de secouer le petit monde de l'historiographie amènent les deux auteurs à omettre l'évidence selon laquelle, indépendamment des courants historiographiques existants, les travaux historiens ne se ressemblent pas.

²⁵⁸ *Ibid.*, *op. cit.*, p. 301

²⁵⁹ *Ibid.*, *op. cit.*, p. 300

²⁶⁰ *Ibid.*, *op. cit.*, p. 309

²⁶¹ *Ibid.*, *op. cit.*, p. 310

²⁶² *Ibid.*, *op. cit.*, p. 312

Il s'agit pourtant de penser une complémentarité des approches historiographiques, exactement comme il s'agit de penser ensemble qualitatif et quantitatif, approche monographique et analyse sérielle, lecture humaine et traitement computationnel, *close* et *distant reading*. Le travail de l'historien est rétrospectif : il passe par l'auto-analyse et l'autocritique, par une réflexion constante sur les problèmes de corpus, de représentativité des sources, d'objectivité et de non-généralisation. Une approche des faits et des sources historiques n'en exclut cependant pas une autre : un retour de la longue durée ne condamne pas la micro-histoire à l'oubli, il encourage au contraire à penser, au regard des moyens numériques actuels, une *approche plurielle* de l'Histoire. C'est ce que concluent les auteurs de l'Éditorial des *Annales* à ce sujet : « Nous sommes convaincus que le moment historiographique actuel appelle à l'expérimentation des approches en même temps qu'à la reformulation des liens entre l'histoire et les autres sciences sociales. »²⁶³.

DE LA SCIENCE HISTORIQUE À LA SCIENCE INFORMATIQUE : FONDEMENTS D'UNE NOUVELLE DISCIPLINE ?

Cette approche plurielle de l'histoire passe en effet par un décroisement des disciplines admises. Les liens que les *Annales* souhaitent voir se reformuler entre l'histoire et les autres sciences sociales semblent même devoir s'étendre au-delà de cette distinction entre SHS et sciences dites exactes. Car dans l'expérimentation des approches que ces auteurs préconisent, on trouve l'approche quantitative et informatique qui pénètre en profondeur la science historique et inaugure l'avènement d'une transdiscipline des plus fécondes.

L'avènement du « cyberhistorien » ?

Les *Big Data*, « mégadonnées » en français, désignent des ensembles si conséquents de données numériques que leur gestion se doit d'être informatisée²⁶⁴ : le terme s'applique donc tout particulièrement aux bases de données détaillées précédemment. La définition que donne ici Wikipédia des *Big Data* est d'autant

²⁶³ « Editorial », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Éditions de l'EHESS, 2/2015, p. 286

²⁶⁴ Voir la page Wikipédia « *Big Data* » consultable à l'adresse suivante : https://fr.wikipedia.org/wiki/Big_data

plus intéressante qu'elle évoque l'usage des outils informatiques comme une *nécessité* dans le traitement des données informatiques constituant de ces mines d'information disponibles sur le Net. Or il est bien évident que la base de données ne peut fonctionner indépendamment des moteurs de recherche, internes ou externes, qui permettent d'en brasser le contenu quasi instantanément. Le monde des livres numérisés apparaît donc comme inextricablement lié à l'usage des technologies de l'information et le Tout numérique induit une indubitable *médiation technique*. Cette nécessaire médiation technique – l'usage de machines – est au fondement d'un nouvel environnement de travail pour le chercheur en sciences humaines et sociales et inaugure de nouvelles pratiques historiennes. L'ère du numérique verra-t-elle la naissance du « cyberhistorien »²⁶⁵ ?

Par « cyberhistorien », n'entendons pas un historien-robot ou robotisé mais plutôt un historien-informaticien ou « informatisé », à même d'appréhender des sources numériques avec en main la maîtrise des outils nouveaux de traitement de l'information. Un tel terme induit donc la connaissance pratique des multiples questions précédemment abordées, de la base de données aux logiciels les plus basiques qui servent à leur utilisation et leur exploitation. Le *cyberhistorien* se caractérise ainsi par une maîtrise globale de l'informatique de bureau, notamment l'usage de l'Internet, ainsi que son application aux problématiques historiennes – pensons à la constitution de corpus, au travail bibliographique, à l'accès aux sources et à leur bonne exploitation, à la visualisation des documents numérisés et à leur traitement graphique, *et cætera*.

Au regard des évolutions techniques actuelles, la société contemporaine est le creuset qui voit naître une véritable *culture du numérique* qui bouleverse l'ensemble des disciplines constituant les vastes « sciences humaines et sociales », de plus en plus médiatisées par la technique. Sans verser dans le scénario de science-fiction, les chercheurs en SHS doivent donc admettre une généralisation des pratiques numériques à l'intérieur même des disciplines qui s'intéressent, de près ou de loin, à l'objet-livre. L'historien du livre, au plus près de la matérialité papier, doit également intégrer de nouvelles pratiques dont le support n'est plus le livre mais l'écran et ses médias : la souris et le clavier. En cela, il y a bien effritement de la frontière qui séparait les « Sciences des Lettres » des « Sciences des Chiffres ».

C'est bien en tant que lecteur-chercheur que l'historien, s'il n'est pas encore « cyborg »²⁶⁶, se doit d'acquérir un socle de compétences solides en informatique. Si l'histoire s'informatise à grande vitesse à l'heure où la numérisation acquiert un caractère politique certain, elle ne signe cependant pas l'arrêt de mort de

²⁶⁵ Ce terme est emprunté à Philippe RYGIEL qui titre « Les curiosités des cyberhistoriens » dans son article « Écriture de l'histoire et réseaux numériques » pour des *Regards croisés sur l'Internet*, Éric GUICHARD (dir.), Villeurbanne : Presses de l'Enssib, 2011, p. 109

²⁶⁶ Référence à l'article de Nicolas DELALANDE et Julien VINCENT, « Portrait de l'historien-ne en cyborg », dans le dossier : « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 5-29

l'historien indépendant et lettré : elle force seulement à l'usage du numérique comme un *outil* nouveau d'appréhension et de conservation du passé.

Nouvelles contraintes disciplinaires

Et comme pour tout outil, son maniement requiert l'assimilation d'une *pratique*. L'usage des technologies n'est pas inné et demande au contraire un réel travail d'appropriation et de partage d'expérience. Il s'agit de former les historiens d'hier et de demain aux outils informatiques qui tendent à bouleverser la nature du travail attendu. L'historien-informaticien n'est pas programmeur professionnel, mais il est concerné par les évolutions de la discipline informatique et se démarque par le maniement efficace des outils qu'il a à sa disposition dans le traitement des données textuelles. Ces nouvelles *contraintes disciplinaires* demandent formation et apprentissage. Des outils méthodologiques de découverte de l'informatique appliquée aux SHS existent déjà, et constituent le socle d'un savoir qui fonde une herméneutique nouvelle indispensable à la bonne constitution d'une science historique numérique : ces outils ne changeront pas tout seuls l'Histoire, il s'agit au contraire de voir évoluer l'approche et l'attitude des historiens eux-mêmes devant cette offre nouvelle. Comme le remarque si pertinemment Philippe Rygiel, « les transformations d'une discipline et d'une institution ne dépendent pas uniquement de la disponibilité d'outils nouveaux »²⁶⁷. En revanche, ils en sont la condition de possibilité.

Nous citerons pour exemple l'œuvre de Claire Lemercier et de Claire Zalc, *Méthodes quantitatives pour l'historien*, un ouvrage fondateur dans le domaine et qui s'offre comme une réflexion sur le traitement quantitatif de la donnée en ce qu'il permet de « multiplier les angles d'approche »²⁶⁸ pour l'étude d'un corpus historique. Notons que les auteures présentent leur projet avec beaucoup de pondération, affirmant ainsi que « la quantification n'est pas pour nous une fin, plutôt un outil parmi d'autres. » et qu'il s'agit de « contribuer à former des lecteurs critiques pour les recherches quantifiées, sans fétichisme ni phobie du chiffre. »²⁶⁹. Elles mettent par ailleurs en garde contre les excès du quantitatif qui mèneraient au traitement de corpus disproportionnés et à une véritable « course au gigantisme »²⁷⁰, citant Gérard Béaur qui s'exprime à ce sujet en 1996 alors que la longue durée chiffrée à la mode dans les années 1970 connaît un moment de crise :

²⁶⁷ Philippe RYGIEL, « Écriture de l'histoire et réseaux numériques », dans *Regards croisés sur l'Internet*, Éric GUICHARD (dir.), Villeurbanne : Presses de l'Enssib, 2011, p. 124

²⁶⁸ Claire LEMERCIER et Claire ZALC, *Méthodes quantitatives pour l'historien*, Paris : La Découverte, 2008, p. 7

²⁶⁹ *Ibid.*, *op. cit.*, p. 3

²⁷⁰ *Ibid.*, *op. cit.*, p. 12

« la croyance qu'il suffisait de rassembler des données par brouettées pour écrire l'histoire et que les conclusions allaient sortir tout armées de l'ordinateur, telle Athéna émergeant de la tête de Zeus. »²⁷¹. À cela, il faut opposer un travail réfléchi d'échantillonnage et de constitution de corpus, qui passe par le respect de la diversité des sources quantifiables et par une attention constante portée à la taille et au choix de l'échantillon ainsi qu'au sens et à l'interprétation des chiffres récoltés.

Dans leur sillage, Jacques Cellier et Martine Cocard complètent la réflexion sur les outils quantitatifs de l'historien par un manuel pratique : *Le traitement des données en Histoire et Sciences Sociales*, qui cherche à approfondir ces méthodes et outils mathématiques de gestion de l'information « sous une forme comestible par les historiens »²⁷². L'ouvrage présente ainsi le vocabulaire propre au monde de l'informatique, l'usage pratique de différents logiciels ainsi qu'une aide à l'interprétation des résultats, le tout illustré par une abondante quantité d'images et de captures d'écran. Les deux auteurs s'intéressent ainsi aussi bien à la constitution d'une base de données et à sa mise en ligne qu'à l'analyse de données multidimensionnelles et autres méthodes exploratoires, à l'analyse longitudinale ou encore à l'Analyse des Réseaux Sociaux. Cette réelle « boîte à outils » de l'historien, terme qu'ils emploient eux-mêmes en introduction²⁷³, n'a pas l'ambition d'être exhaustive mais illustre les efforts qui sont faits par de nombreux chercheurs pour rapprocher méthodologiquement Histoire et techniques informatiques. La démarche quantitative prend alors tout son sens :

« Le recours à des techniques qui permettent de reculer le moment de la subjectivité et d'évaluer la solidité des résultats est particulièrement utile pour lutter contre des usages paresseux ou abusifs de la citation. »²⁷⁴

« La quantification a précisément l'avantage de contraindre à expliciter choix et procédures. [...] Elle incite à multiplier les essais – donc les erreurs – et à varier les échelles d'analyse. Loin de brider l'intuition ou l'inventivité de l'historien, elle peut au contraire les stimuler. »²⁷⁵

De même, certains historiens éclairés partagent leurs expériences digitales et tentent d'aller à l'encontre de la discrétion des chercheurs relativement à leurs pratiques numériques²⁷⁶. On sait qu'une majorité d'historiens qualifiés restent sceptiques devant les apports de l'informatique à leur propre discipline et que leur

²⁷¹ *Ibid.*

²⁷² Martine COCAUD et Jacques CELLIER, *Le traitement des données en histoire et sciences sociales : méthodes et outils*, Rennes : Presses universitaires de Rennes, 2012, p. 9

²⁷³ *Ibid.*, *op. cit.*, p. 11

²⁷⁴ Claire LEMERCIER et Claire ZALC, *Méthodes quantitatives pour l'historien...*, p. 50

²⁷⁵ *Ibid.*, *op. cit.*, p. 104

²⁷⁶ Nous pensons, pour n'en citer que deux, à l'article *Faire de l'histoire à l'ère numérique : retours d'expériences*, de Franziska Heimburger et Emilien Ruiz, et à *L'enquête historique à l'ère numérique*, de Philippe Rygiel.

rapport à l'ordinateur reste des plus rares et des plus utilitaristes – disons des plus fonctionnels.

Sans que l'informatique ne supplante le travail de réflexion à la source de tout discours historique, il y a ainsi beaucoup à gagner, semble-t-il, à un rapprochement interdisciplinaire plus poussé.

Les « humanités numériques », aux fondements d'une nouvelle épistémologie ?

Il faut cependant noter que le lien qui tend à unir Histoire et Informatique n'est pas nouveau. Depuis les années 1970, avec l'engagement de nombreux ingénieurs et techniciens au sein du Centre de recherches historiques (CHR) de l'EHESS, l'incursion des pratiques numériques en sciences humaines et sociales donne lieu à une réflexion sur le fondement même de ces disciplines, alors que naissait le projet Gutenberg déjà évoqué. Retenons ainsi du monde de la presse spécialisée quelques grandes revues fondatrices relatives à l'Histoire et à l'Informatique.

Au printemps 1979 naît la revue *Le Médiéviste et l'Ordinateur* qui a compté quarante-cinq numéros avant de prendre fin en 2003²⁷⁷, heureux témoignage de l'ancienneté de la réflexion historique sur l'usage des technologies informatiques. En 1986, une revue à la thématique plus large voit le jour, *Histoire et Mesure*, qui interroge la pratique historique et la pratique de la mesure chiffrée en Histoire sous toutes ses formes. Elle questionne donc, depuis trente ans, les problématiques du chiffre dans la discipline historique. En 1987 naît la revue *History and Computing*, destinée à rendre compte des avancées technologiques dans le domaine historique. Devenue l'*International Journal of Humanities and Arts Computing*, cette revue est aujourd'hui pionnière dans la presse spécialisée relative aux humanités numériques : pluridisciplinaire et ouverte à toutes les approches, elle témoigne d'un vaste courant de réflexion sur l'impact des technologies numériques dans la recherche en sciences humaines et sociales. Deux ans plus tard sort le premier numéro de *Mémoire vive*, bulletin de l'Association française pour l'histoire et l'informatique. Ce périodique biannuel, dans la lignée des trois autres, s'intéresse à l'impact de l'informatique dans les domaines et de la pédagogie et de la recherche.

²⁷⁷ L'ensemble des numéros est désormais disponible sur le Web grâce à l'Ecole des Chartes. Les vingt premiers numéros sur le site Ménéstrel, à l'adresse suivante : <http://www.menestrel.fr/spip.php?rubrique667> ; les vingt-et-un suivants sur le site de l'IRHT : <http://lemo.irht.cnrs.fr/>.

Ces différentes publications rendent compte de l'émergence précoce d'une réflexion relative à l'irruption de l'informatique en sciences sociales et préfigurent la naissance d'une discipline de plus en plus reconnue mais qui fait largement débat, notamment dans le milieu académique et de l'enseignement supérieur : les Humanités Digitales. Les *humanités digitales*, aussi appelées « humanités numériques », sont un domaine de recherche, d'enseignement et d'ingénierie au croisement de l'informatique et des arts, lettres, sciences humaines et sociales. Elles se caractérisent par des méthodes et des pratiques liées à l'utilisation des outils numériques, en ligne et hors ligne, ainsi que par la volonté de prendre en compte les nouveaux contenus numériques, au même titre que des objets d'étude plus traditionnels²⁷⁸. En ce sens, les humanités numériques modifient en profondeur certaines pratiques de recherche au point de redéfinir, semble-t-il, les fondements de certaines disciplines. Les *humanités numériques* seraient-elles au fondement d'une nouvelle épistémologie ?

Un tel propos revient à accorder au numérique une place prépondérante dans les modalités de construction et d'appropriation du savoir de nos contemporains. Par là même, il ouvre un chemin nouveau vers la connaissance. C'est admettre que dans le rapport des lecteurs au contenu numérique ou numérisé, quelque chose a changé qui dépasse le seul outillage. L'outil technique et le savoir scientifique, intimement liés, ne souffrent d'aucun principe de hiérarchisation : ils s'interpénètrent et se déterminent l'un l'autre. La recherche de la connaissance conditionne les avancées techniques, tout autant que ces dernières orientent les démarches cognitives et épistémiques. En cela, science et technique se conditionnent mutuellement ; ils sont chacun la condition de possibilité de l'autre. Comme outil technique, le numérique influence donc la démarche de construction de la connaissance : il est en cela facteur de bouleversement épistémique, au point d'avoir signé l'avènement d'une épistémologie d'un ordre nouveau. Cette épistémologie prend racine dans un terreau interdisciplinaire en constitution : les humanités numériques.

Pour un historien spécialiste du livre comme Roger Chartier, l'entrée dans l'ère du numérique apparaît comme une rupture bien plus nette dans l'histoire de l'écrit qu'une autre « révolution » telle que celle de l'imprimerie, qui ne tient pas compte du continuum qui a pu exister entre le livre manuscrit et le livre imprimé, depuis l'invention du *codex*. Si la typographie a supplanté l'écriture manuscrite, le livre en lui-même n'a en effet pas radicalement changé ; en témoigne l'incunable du premier demi-siècle post-Gutenberg qui allait jusqu'à imiter graphiquement l'esthétique du manuscrit. Une première révolution, marquée par le passage de l'écrit sur *volumen* à l'écrit sur *codex*, tient donc à l'invention de la *page*, support matériel figé ayant bouleversé en profondeur les pratiques d'écriture et, surtout, de

²⁷⁸ Voir la page Wikipédia « Humanités numériques » consultable à l'adresse suivante : https://fr.wikipedia.org/wiki/Humanit%C3%A9s_num%C3%A9riques

lecture. On date cette première révolution, le passage du rouleau au *codex*, d'entre le II^e et le IV^e siècle. Cette invention apporte alors à la lecture une forme nouvelle de liberté, liée à la maniabilité de la page, à la pagination, aux index et aux tables des matières, « autorisant une écriture simultanée et permettant la comparaison et la citation. »²⁷⁹ La généralisation des écrans aujourd'hui apparaît ainsi comme une seconde révolution qui, de la même manière, marque en profondeur les modalités d'appropriation du texte et le rapport texte-lecteur. Attention cependant à ne pas réduire la pensée de Roger Chartier qui, s'il tend à réévaluer l'impact qu'a pu avoir l'invention de l'imprimerie en Occident, n'en cherche pas moins à éviter tout schématisme théorique : il constate ainsi l'importance de l'apparition du livre unitaire, datée des XIV^e-XV^e siècles, qui marque également une rupture dans l'histoire de l'écrit²⁸⁰. Aussi, « la révolution du texte électronique est une révolution des structures du support matériel de l'écrit comme des manières de lire. [...] l'écran apparaît comme le point d'aboutissement du mouvement qui a séparé le texte du corps. »²⁸¹

Parce qu'il est la source d'une mobilité nouvelle du texte, le passage au numérique représente, non pas une forme de retour au *volumen* – ce serait là une comparaison réductrice – mais une redécouverte certaine de la labilité du texte désolidarisé du cadre fermé de la page papier. Remarquons ainsi que la comparaison, la citation, l'écriture simultanée et la maniabilité de l'écrit atteignent avec le numérique un bien plus haut degré d'accomplissement encore qu'avec la page papier. La lecture et le rapport au texte en général sont d'autant plus libres que ce dernier n'est plus lié à un support matériel stable ; au contraire, c'est justement cette souplesse d'utilisation qui vient compenser le caractère encore peu « lisible » de l'écran, sans compter l'amélioration considérable du confort de lecture que promettent aujourd'hui les liseuses sur écran rétroéclairé.

Cette mouvance et cette instabilité du texte numérique sont ainsi à la source d'un courant épistémologique nouveau qui engage de nouvelles pratiques de lecture et l'utilisation d'outils nouveaux à même de pallier une insuffisance du texte numérique à faire source. Il est à dire, donc, que le numérique engendre, plus que l'imprimerie, une mutation sans précédent du lire et de l'écrire. Il touche, via la lecture, au fondement même de la construction de la connaissance et impacte le travail des chercheurs qui, dans toutes les disciplines, doivent apprendre à utiliser et à évaluer les données que brasse le World Wide Web.

En cela, le numérique a le mérite d'être au fondement d'une science nouvelle que l'on pourra nommer « épistémologie digitale »²⁸² et qui concerne le rapport

²⁷⁹ Gaëlle BÉQUET, *Trois bibliothèques européennes face à Google...*, p. 44

²⁸⁰ A ce sujet, voir *Qu'est-ce qu'un livre ?* de Roger Chartier.

²⁸¹ Roger CHARTIER, *Le livre en révolutions...*, p. 13

²⁸² L'expression, empruntée à Éric Guichard, est employée dans « Epistémologies digitales des sciences humaines et sociales », Éric GUICHARD et Thierry POIBEAU (dir.), *Revue Sciences/Lettres*, n° 2, 2014

qu'entretiennent avec le Savoir scientifique les acteurs multiples des humanités numériques. Une *épistémologie digitale* qui fonde une philosophie des sciences numériques ou numérisées et qui questionne la valeur épistémique du numérique. De là les multiples courants de pensée qui depuis *Le Médiéviste et l'Ordinateur* interrogent le rapport que peuvent entretenir les sciences humaines et l'informatique comme science du nombre. Née de la rencontre entre Savoir et Technique, cette épistémologie nouvelle – numérique, ou digitale – force à réévaluer la valeur du discours historique et de l'opération historiographique. L'ancienneté de la réflexion sur ce sujet, éclairée par un bref coup d'œil sur l'histoire de la presse, révèle par ailleurs une heureuse propension des historiens à réfléchir sur leurs propres pratiques et rend compte d'un très valorisant regard rétrospectif porté par les chercheurs sur leur propre discipline.

« Transdiscipline » et « Data Sciences »

La frontière que l'on imagine séparer les disciplines « littéraires » et les disciplines « scientifiques » n'est donc pas si opaque qu'elle peut en avoir l'air. Les revues « informatico-historiennes » ou « historico-informatiques » évoquées précédemment rendent compte de l'existence certaine d'un débat destiné à réévaluer la concomitance et la perméabilité de ces deux disciplines. Aussi, si de nombreux chercheurs militent aujourd'hui pour le décloisonnement des savoirs, notamment dans l'enseignement, et clament haut et fort que l'avenir est à l'*interdisciplinarité*, il est à remarquer que la problématique n'est pas toute nouvelle.

Qui critique la trop grande autonomie de l'historien et sa détermination à s'exclure du monde de la recherche moderne laisse pour compte les historiens qui, de longue date, tentent de rendre effective une communication interdisciplinaire qui, dans les faits, existe à plus ou moins grande échelle. Les TIC, les fameuses Technologies de l'Information et de la Communication, ont progressivement infiltré la diversité des champs de la recherche académique et s'affirment aujourd'hui comme des maillons incontournables de la chaîne de la construction du savoir. Est-ce à dire que les TIC ont déconstruit le métier de l'historien ? Évidemment non. Celui-ci conserve une *autonomie* certaine dans la pratique de sa discipline, bien qu'il lui faille certainement se familiariser avec les outils informatiques afin de conserver cette relative autonomie. Nous l'avons vu, les pratiques historiennes ne se résument pas à des pratiques informatiques : il y a, à tous les niveaux du processus de construction de l'objectivité historique, un travail d'évaluation, de critique et de sélection essentiel à la production d'un discours valable scientifiquement. Mais cette validité scientifique est aujourd'hui créditée par des pratiques numériques de

traitement des données – quelle que soit leur quantité – qui forcent le lecteur à manipuler machines et logiciels qui constituent le fondement de ce qu'on peut appeler une « transdiscipline »²⁸³.

Cette *transdiscipline* n'est pas seulement interdisciplinarité : elle *transcende* les cloisonnements qui peuvent exister entre les différents domaines du savoir pour un usage croisé des théories et des outils, méthodologiques et pratiques, propres à chaque discipline. Elle gomme progressivement les frontières qui autonomisent les Lettres des Sciences et annonce la réconciliation des chiffres et des lettres. Cette transdiscipline, avec tout ce qu'elle peut avoir de fantasmée, reste pertinente pour qualifier les bouleversements épistémologiques qui ébranlent l'herméneutique et la philologie et nous force à réévaluer le rapport que peuvent entretenir les hommes avec le monde de l'écrit.

Elle ne naît donc pas des *Data Sciences*, ces « sciences des données »²⁸⁴ qui se résument finalement à l'exploitation informatisée de bases de données numériques, et dont on a suffisamment observé les limites et les faiblesses. Les *Data sciences* ne constituent pas à proprement parler une discipline à part entière, du moins pas selon cette conception réductrice qui tend à subordonner la science à la technique. Disons plutôt que la transdisciplinarité en puissance des sciences au sens large englobera, parmi d'autres, un domaine de recherche dédié aux données numériques.

L'idée selon laquelle l'ordinateur remplacera l'historien est inepte, tant que l'intelligence artificielle n'aura pas signé l'avènement de l'ère des machines douées de facultés de raisonnement et d'esprit critique. Car ces deux traits de l'esprit humain, *raisonnement* et *esprit critique*, sont au fondement de l'analyse historique et de la production d'un discours raisonné. Toutes les sciences, exactes ou non, humaines ou non, analysent des données ; ces données, à l'heure actuelle, sont majoritairement médiatisées par le numérique ; toutes les sciences sont donc finalement des *Data sciences*, ou bien aucune. Il faut seulement que les sciences parviennent à s'adapter à la massification des données disponibles, à approfondir et à assimiler les méthodes nécessaires à leur traitement. C'est là qu'apparaît la « transdiscipline » à la croisée des mondes de l'intellect, c'est là que se réconcilient des domaines que l'on a longtemps voulu opposer : les sciences molles se durcissent d'un afflux de données chiffrées qui les dressent à la hauteur supposée des sciences exactes.

²⁸³ Pierre MOUNIER (dir.), *Read/Write Book 2. Une introduction aux humanités numériques*, OpenEdition Press, 2012, 262 p. 22

²⁸⁴ Pensons notamment à Jean-Christophe Plantin qui dans « Les chercheurs en SHS rêvent-ils de code informatique ? », quatrième chapitre du *Temps des humanités digitales*, Olivier LE DEUFF (dir.), examine le profil du *data scientist*, « qui doit lier des compétences hétérogènes afin d'extraire de grandes masses de données, les interpréter et en imaginer les applications » (p. 65).

CONCLUSION

L'entrée fracassante du numérique dans les sociétés contemporaines est cause de profonds bouleversements au sein des pratiques disciplinaires propres aux sciences humaines et sociales. La discipline historique en particulier connaît, avec la massification des ressources documentaires disponibles en ligne, un véritable changement de paradigme épistémique : les modalités d'appréhension, de constitution et d'assimilation des connaissances en sont en effet profondément transformées.

Si l'Internet est devenu un moyen courant de communication entre les chercheurs, et les traitements de texte un outil admis de constitution de l'information scripturale, c'est la numérisation plus encore qui induit de nouvelles pratiques de recherche et force les historiens à porter un regard nouveau sur l'avenir de leur discipline. En effet, la mise en ligne et à disposition d'un large public de la version numérisée d'ouvrages anciens tend progressivement à dévaluer l'importance des originaux papier conservés en fonds anciens, qui s'y endorment. Pourtant, ces originaux papier restent le support de l'étude historique et la source du travail scientifique d'un nombre non négligeable de chercheurs en histoire de l'écrit et de l'image.

À la question de savoir ce qu'il reste de nos bibliothèques physiques, il est donc difficile de répondre : il est certain que deux mondes, le monde bibliothéconomique et le monde scientifique, ont aujourd'hui à composer pour s'entendre sur les modalités de numérisation et de conservation des ressources anciennes. Une telle concertation est garante du respect de l'intégrité du patrimoine livresque mondial et un garde-fou contre la marchandisation du domaine public et son appropriation par l'entreprise qu'est Google. Loin de toute considération économique ou politique, il est en effet certain qu'une situation de monopole établie sur la propriété des sociétés a de quoi inquiéter les détracteurs les plus frileux d'une bibliothèque numérique mondiale. Cette mondialisation des ressources induit par ailleurs une offre en « vrac » constitutive du processus de numérisation de masse et qui va à l'encontre du travail de classement qu'ont toujours mené les institutions bibliothéconomiques. Le chercheur doit ainsi apprendre à faire face à une démultiplication des sources potentielles d'intérêt ainsi qu'à de récurrentes défaillances du travail de catalogage et d'indexation des œuvres numérisées. Il s'agit alors d'acquérir la maîtrise de nouveaux outils de recherche et d'accès à l'information en ligne, pour l'appréhension des fichiers numériques et des métadonnées qui les identifient et les complètent.

L'assimilation de telles habitudes numériques s'avère une nécessité, induite par la nature même de l'offre numérique, fondamentalement instable et objet d'une obsolescence qui force les opérateurs de la numérisation à travailler la « trouvabilité » des fichiers sur le Net : cette *searchabilité* du fichier passe par du méta-texte et de la publicité. Cependant, l'appréhension de l'offre numérique n'est pas tout, il est à l'heure du numérique de nouveaux outils de traitement de l'information qui depuis plusieurs décennies investissent le champ de la recherche historique.

Ces outils de traitement de l'information médiatisent une forme nouvelle de travail des sources. Parce que la numérisation induit dématérialisation du support scriptural physique, elle force à penser une technique neuve de traitement du livre : une forme de bibliographie de l'immatériel, une archéologie du livre numérisé qui succède à l'archéologie du livre imprimé telle qu'elle se pratique depuis la première moitié du XX^e siècle. Les bases de données multiples, spécialisées ou non, deviennent le lieu de pratiques nouvelles d'authentification de l'imprimé ancien et d'analyses paléotypographiques. Ces dernières sont rendues d'autant plus aisées que des logiciels de reconnaissance optique des formes permettent d'automatiser le processus de déchiffrement des textes, déchiffrement à partir duquel des logiciels de traitement des données peuvent seconder, faciliter ou compléter le travail de l'historien. Le numérique, en cela, est à l'origine d'une expansion des approches historiques quantitatives, que certains diront quantophréniques. Loin de signer la mort de l'approche historique qualitative et monographique, le quantitatif si décrié semble au contraire être en mesure de forcer les historiens à élargir leur champ disciplinaire : il induit de nouvelles contraintes disciplinaires qui façonnent un cyberhistorien de demain, non plus seulement tourné vers les sciences humaines et sociales mais concerné par les pratiques des sciences exactes. Un tel rapprochement disciplinaire laisse à croire qu'émerge de nos jours une transdiscipline qui tend à combler l'actuel fossé qui sépare les chiffres des lettres. Cette transdiscipline, que l'on tente aujourd'hui de penser sous l'expression « humanités numériques », force un rapprochement entre les sciences humaines et sociales et les sciences informatiques, rapprochement que nous sommes en droit d'espérer fructueux.

Cependant, l'approche du numérique comme outil mis au service de l'historien du livre est vaste. Nous l'avons vu, le numérique s'insère désormais à tous les stades de la recherche historique et donne lieu aujourd'hui encore à des programmes de recherche relativement nombreux qui tendent à élargir la gamme des outils aujourd'hui disponibles dans la boîte de l'historien. Il serait donc tout à fait réducteur de croire que ce bref état des lieux ait pu embrasser l'ensemble des bouleversements qu'induit le rapprochement contemporain entre pratique historique et pratique numérique. Qu'il s'agisse des projets actuels de numérisation

ou des outils informatiques à la disposition des scientifiques, les changements sont en cours : loin d'être achevée, la révolution numérique n'en est présentement qu'à ses débuts. Au service de l'historien du livre, le numérique est donc, en tous points, des plus prometteurs.

BIBLIOGRAPHIE

NB : Tous les liens Internet renvoyant vers des ressources numériques ont été vérifiés au 31 mai 2016.

LE LIVRE, LES BIBLIOTHÈQUES ET LE NUMÉRIQUE

Monographies

Actes du colloque *Vers une nouvelle érudition : la numérisation au service de l'histoire du livre*, codirigé par VARRY, Dominique, BAUDRY, Guylaine et CHARON, Annie, à l'occasion des 12^e entretiens du Centre Jacques Cartier, Lyon-Villeurbanne, Enssib, 6-7 décembre 1999, disponible sur le Web : www.enssib.fr/bibliotheque-numerique/document-1512

BARBIER, Frédéric, *Histoire des bibliothèques : d'Alexandrie aux bibliothèques virtuelles*, Paris : A. Colin, 2013, 301 p.

BARBIER, Frédéric, *Histoire du livre en Occident*, Paris : A. Colin, 2012, 351 p.

BENHAMOU, Françoise, *Le livre à l'heure numérique : papier, écrans, vers un nouveau vagabondage*, Paris : Ed. du Seuil, 2014, 215 p.

BÉQUET, Gaëlle, *Trois bibliothèques européennes face à Google : aux origines de la Bibliothèque numérique (1990-2010)*, Paris : École des Chartes, 2014, 473 p.

BON, François, *Après le livre*, Paris : Éd. du Seuil, 2011, 274 p.

CALDERAN, Lisette, HIDOINE, Bernard et MILLET, Jacques (coord.), *Pérenniser le document numérique*, séminaire Inria, 2-6 octobre 2006, Amboise : ADBS éd., 2006, 206 p.

CATANÈSE, Christophe, DUPLAND, Laurence, HAMON, Hélène, *et al.*, *Numérisation de fonds anciens : les réalisations de la Bibliothèque nationale de France et de la Bibliothèque Municipale de Lyon Part-Dieu*, travail de formation à la recherche pour le diplôme de conservateur des bibliothèques, dir. RIONDET, Odile, , Villeurbanne : Bibliothèque numérique de l'Enssib, 1999 : <http://www.enssib.fr/bibliotheque-numerique/notices/61882-numerisation-de-fonds-anciens-les-realizations-de-la-bibliotheque-nationale-de-france-et-de-la-bibliotheque-municipale-de-lyon-part-dieu>

CHARTIER, Roger, *Le livre en révolutions : entretiens avec Jean Lebrun*, Paris : Textuel, 1997, 159 p.

CHEVRY, Emmanuelle, *Stratégies numériques : numérisation et exploitation du patrimoine écrit et iconographique*, Paris : Hermès Sciences Pub. : Lavoisier, 2011, 268 p.

CLAERR, Thierry et WESTEEL, Isabelle (dir.), *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la librairie, 2013, 407 p.

COMBIER, Marc et PESEZ, Yvette (dir.), *Encyclopédie de la chose imprimée : du papier à l'écran*, Paris : Éd. Retz, 2004, 544 p.

FREY, Valentine et TRELEANI, Matteo (coord.), *Vers un nouvel archiviste numérique*, Paris : L'Harmattan, 2013, 217 p.

GILMONT, Jean-François et VARRY, Dominique, *Du parchemin à l'ère électronique : une histoire du livre et de la lecture*, Liège : Céfal, 2014, 154 p.

ILLICH, Ivan, *Du lisible au visible : la naissance du texte*, Paris, Les Éditions du Cerf, 1991, 150 p.

JACQUESSON, Alain, *Google livres et le futur des bibliothèques numériques : historique du projet, techniques documentaires, alternatives et controverses*, Paris : Éd. du Cercle de la librairie, 2010, 223 p.

JACQUESSON, Alain et RIVIER, Alexis, *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux*, 2nde éd., Paris : Éd. du Cercle de la librairie, 2005, 573 p.

JEANNENEY, Jean-Noël, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, 3^e éd., Paris : Éd. Mille et une nuits, 2010, 219 p.

LE RAY, Éric et LAFRANCE, Jean-Paul (dir.), *La bataille de l'imprimé, à l'ère du papier électronique*, Montréal, Presses de l'Université de Montréal, 2008, 257 p.

MCKITTERICK, David, *Old Books, New Technologies: The Representation, Conservation and Transformation of Books since 1700*, New-York: Cambridge University Press, 2013, 286 p.

MOATTI, Alexandre, *Au pays de Numérix*, Paris : PUF, 2015, 166 p.

MOATTI, Alexandre, TESNIERE, Valérie et LESQUINS, Noémie, Rapport BNUE « Bibliothèque numérique européenne », remis le 31 janvier 2006 au Ministre de la Culture et de la Communication, 2006, disponible sur le Web : <http://www.bibnum.eu/article-4226019.html>

POLASTRON, Lucien Xavier, *La Grande Numérisation : y a-t-il une pensée après le papier ?*, Paris : Denoël, 2006, 198 p.

RACINE, Bruno, *Google et le nouveau monde*, Paris : Plon, 2010, 149 p.

SANDOZ, Jean-David, *Du bon usage des Lumières : le livre numérique libre*, Condeixa-a-Nova : la Ligne d'ombre, 2015, 184 p.

SARZANA, Jean et PIERROT, Alain, *Impressions numériques : quels futurs pour le livre ?*, Paris : Éd. du Cerf, 2011, 230 p.

SERVAIS, Paul et MIRGUET, Françoise (dir.), *L'archive dans quinze ans : vers de nouveaux fondements*, Louvain-la-Neuve : Academia : l'Harmattan, 2015, 270 p.

SOCCAVO, Lorenzo, *De la bibliothèque à la bibliosphère : les impacts du livre numérique dans le monde des bibliothèques*, Andrézieux-Bouthéon : Éd. Morey, 2011, 115 p.

SOCCAVO, Lorenzo, *Gutenberg 2.0 : Le futur du livre. Six siècles après Gutenberg une nouvelle révolution va changer votre façon de lire...*, Paris : M21 Éditions, 2008, 222 p.

SOUAL, Laurent, *Le livre numérique en bibliothèque : état des lieux et perspectives*, Paris : Éd. Du Cercle de la librairie, 2015, 119 p.

TESSIER, Marc, RACINE, Bruno, JEANNENEY, Jean-Noël et al., *La révolution du livre numérique : état des lieux, débats, enjeux*, Paris : Odile Jacob, 2011, 208 p.

TESSIER, Marc, *Rapport sur la numérisation du patrimoine écrit. Remis par Marc Tessier au ministre de la Culture et de la Communication le 12 janvier 2010*, Paris : Ministère de la Culture et de la Communication, 2010, 64 p. : http://www.syndicat-librairie.fr/fr/rapport_tessier_sur_la_numerisation_du_patrimoine_ecrit_2010

VANDENDORPE, Christian, *Du papyrus à l'hypertexte : essai sur les mutations du texte et de la lecture*, Paris : Éd. La Découverte, 1999, 271 p.

VARRY, Dominique (dir.), *50 ans d'histoire du livre : 1958-2008*, Villeurbanne : Presses de l'Enssib, 2014, 223 p.

Articles et revues

ANDRO, Mathieu, CHAIGNE, Marion et SMITH, Franck, « Valoriser une bibliothèque numérique par des choix de référencement et de diffusion : l'expérience de la bibliothèque Sainte-Geneviève », dans « Valorisation des corpus numérisés »,

CHEVRY PEBAYLE, Emmanuelle (dir.), *Les Cahiers du numérique*, vol. 8, 2012/3, p. 75-90 : <http://www.cairn.info/bibliotheque-nomade2.univ-lyon2.fr/revue-les-cahiers-du-numerique-2012-3.htm>

BENHAMOU, Françoise, « Le livre et son double, réflexion sur le livre numérique », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 90-102, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

BERMÈS, Emmanuelle et MARTIN, Frédéric, « Le concept de collection numérique », *Bulletin des bibliothèques de France*, t.55, n°3, 2010, p. 13-17, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2010-03-0013-002>

BERMÈS, Emmanuelle, FRÉON, Marie-Élise et MARTIN, Frédéric, « Tous les chemins mènent au numérique », *Bulletin des bibliothèques de France*, t.53, n° 6, 2008, p. 34-39, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2008-06-0034-006>

BLAIS, Jean-Marc, « Le numérique au service du patrimoine », *Continuité*, n° 99, 2003-2004, p. 44-46, disponible sur le Web : <http://erudit.org/culture/continuite1050475/continuite1055478/15635ac.pdf>

CHARTIER, Robert, « L'avenir numérique du livre », *Le Monde*, mardi 27 octobre 2009, p. 20

CHARTIER, Roger, « L'écrit et l'écran, une révolution en marche », *Le Monde*, 13 octobre 2007, p. 24

COMPAGNON, Antoine, « Lire numérique », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 103-106, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

COUZINET, Viviane, « Ressources numériques dans l'espace européen : impact sur l'évaluation de la recherche scientifique », *Les Cahiers du numérique*, Vol. 11, 2015/1, p. 141-160

DARNTON, Robert, « La chandelle de Jefferson, une bibliothèque numérique nationale », traduction de Pierre-Emmanuel Dautat, dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 112-119, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

DARNTON, Robert, « Pour une solution internationale », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 130-132, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

DARNTON, Robert, « Accès public, contrôle privé. La bibliothèque universelle, de Voltaire à Google. », *Le Monde diplomatique*, 1^{er} mars 2009, p. 24-25

EL HADI WIDAD, Mustafa et FAVIER, Laurence, « L'archivage numérique des savoirs : perspectives européennes », *Les Cahiers du numérique*, Vol. 11, 2015/1, 166 p.

GAUVAIN, Claire-Lise, « Outils numériques et fonds anciens », *Bulletin des bibliothèques de France*, n° 1, 2009, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2009-01-0086-002>

GUICHARD, Éric, « Numérique (livre) », dans GHORRA-GOBIN, Cynthia (dir.), *Dictionnaire critique de la mondialisation*, Paris : Armand Colin, 2012, p. 462-469

HERRENSCHMIDT, Clarisse, « De l'écriture numérique aux « nouvelles » bibliothèques », dans GHORRA-GOBIN, Cynthia (dir.), *Dictionnaire critique de la mondialisation*, Paris : Armand Colin, 2012, p. 462

JABLONKA, Ivan, « Le livre, son passé, son avenir. Entretien avec Roger Chartier », dans « Le livre demain », *La vie des idées*, sept-oct 2008, consultable à l'adresse suivante : <http://www.laviedesidees.fr/Le-livre-son-passe-son-avenir.html>

JEANNENEY, Jean-Noël, « Allégresse et vigilance », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 120-125, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

JEANNERET, Yves, « Le procès de numérisation de la culture, un défi pour la pensée du texte. », *Protée*, vol. 32, n°2, 2004, p. 9-18, disponible sur Érudit : <http://www.erudit.org/revue/pr/2004/v32/n2/011168ar.html>

LACOMBE, Christian, « La numérisation ou le grand sommeil du livre imprimé », *Documentation et bibliothèques*, Montréal, Asted, vol. 55, n°1, 2009, p. 39-44, disponible sur Érudit : <https://www.erudit.org/revue/documentation/2009/v55/n1/1029050ar.pdf>

LEBERT, Marie, « Le livre numérique : chronologie 1971-2007 », *Dossier du Net des études françaises*, 2007, disponible à l'adresse suivante : <http://www.etudes-francaises.net/dossiers/chronologie.htm#01>

MI, Jia et NESTA, Frederick, « Marketing library services to the Net generation », *Library Management*, t. 27, n° 6-7, 2006, p. 411-422

RACINE, Bruno, « Une utopie séduisante », dans « Le livre, le numérique », *Le Débat*, n°170, 2012/3, p. 126-129, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2012-3.htm>

RENARD, Isabelle, « L'original est mort, vive la trace numérique ! », *Journal du Net (JDN)*, 27 février 2003, consultable à l'adresse suivante : http://www.journaldunet.com/solutions/0302/030227_juridique.shtml

VAUCELLE, Alain et TASSADIT, Yacine, « Les humanités numériques et le projet de Bibliothèque Numérique Franco-Berbère », *Iles D Imesli*, n°5, 2013 : <http://revue.ummtto.dz/index.php/idi/article/download/289/209>

WEBER, Anne, « La numérisation des sources de l'histoire de l'art », *Bulletin des bibliothèques de France*, n° 6, 2008, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2008-06-0054-011>

WESTEEL, Isabelle, « Le patrimoine passe au numérique », *Bulletin des bibliothèques de France*, t.54, n° 1, 2009, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2009-01-0028-003>

L'HISTORIEN, SES OUTILS ET LE NUMÉRIQUE

MONOGRAPHIES

BARBEY, Nicolas, GUILLEMAIN, Jean, PEOC'H, Géraldine et RACT, Patrice, *La renaissance du livre ancien : bilan du projet DEBORA et perspectives d'avenir*, Mémoire de recherche pour la validation du diplôme de conservateur de bibliothèques, Villeurbanne : Bibliothèque numérique de l'Esssib, 2002, 71 p., disponible sur le Web : <http://www.enssib.fr/bibliotheque-numerique/documents/995-la-renaissance-du-livre-ancien.pdf>

BERGE, Claude, *Théorie des graphes et ses applications*, Paris : Dunod, 1958, 267 p.

CHAUMIER, Jacques, en collaboration avec GICQUEL, Florence, *Les Techniques documentaires au fil de l'histoire : 1950-2000*, Paris : ADBS, 2003, 179 p.

COCAUD, Martine et CELLIER, Jacques, *Le traitement des données en histoire et sciences sociales : méthodes et outils*, Rennes : Presses universitaires de Rennes, 2012, 553 p.

COUSTATY, Mickaël, « Contribution à l'analyse complexe de documents anciens : application aux lettrines », sous la direction de Jean-Marc OGIER, Thèse de doctorat en Informatique et Applications, Université de La Rochelle, 2011, 240 p.

DE CERTEAU, Michel, *L'écriture de l'histoire*, Paris : Gallimard, 1975, 358 p.

DOUEIHI, Milad, *La Grande conversion numérique*, Paris : Éd. du Seuil, 2008, 271 p.

GENET, Jean-Philippe et ZORZI, Andrea, *Les historiens et l'informatique : un métier à réinventer*, Rome : Collection de l'École Française de Rome, 2011, 350 p.

GUICHARD, Éric, « Introduction », dans GUICHARD, Éric (dir.), *Regards croisés sur l'Internet*, Villeurbanne : Presses de l'Enssib, 2011, p. 7-23

GULDI, Jo et ARMITAGE, David, *The History Manifesto*, Cambridge : Cambridge University Press, 2014, 165 p.

HENRYOT, Fabienne, *L'historien face au manuscrit : du parchemin à la bibliothèque numérique*, Louvain-la-Neuve : Presses universitaires de Louvain, 2012, 366 p.

LE DEUFF, Olivier (dir.), *Le temps des humanités digitales : la mutation des sciences humaines et sociales*, Limoges : FYP éditions, 2014, 175 p.

LEMERCIER, Claire et ZALC, Claire, *Méthodes quantitatives pour l'historien*, Paris : La Découverte, 2008, 120 p.

MINUTI, Rolando, *Internet et le métier d'historien : réflexions sur les incertitudes d'une mutation*, Paris : Presses universitaires de France, 2002, 146 p.

MORETTI, Franco, *Distant Reading*, London ; New York : Verso, 2013, 244 p.

MOUNIER, Pierre (dir.), *Read/Write Book 2. Une introduction aux humanités numériques*, OpenEdition Press, 2012, 262 p., aussi disponible en ligne : <http://press.openedition.org/226>

RYGIEL, Philippe, « Écriture de l'histoire et réseaux numériques », dans : GUICHARD, Éric (dir.), *Regards croisés sur l'Internet*, Villeurbanne : Presses de l'Enssib, 2011, p. 101-124

SOULATGES, Magali, FORTUNY, Claudette et CORSINI, Silvio, « Quels nouveaux développements pour les bases de données d'ornements d'imprimerie ? », Synthèse de la journée d'étude internationale organisée le 29 mai 2012 par l'IRCL, juillet 2012, disponible sur le Web : <http://www.enssib.fr/quels-nouveaux-developpements-pour-les-bases-de-donnees-dornements-dimprimerie>

VINCK, Dominique et NATALE, Enrico, « La transformation des sciences historiques. La part du numérique », dans : GORGA, Anne et LERESCHE, Jean-Philippe (éd.), *Disciplines académiques en transformation. Entre innovation et résistances*, Paris : Édition des archives contemporaines, 2015, p. 89-104

VINCK, Dominique, « Les *digital humanities* comme promesses pour et par les sciences humaines », dans : AUDETAT, Marc (dir.), *Sciences et technologies émergentes : Pourquoi tant de promesses ?*, Paris : Hermann, 2015, p. 131-145

WIEVIORKA, Michel, *L'impératif numérique, ou la nouvelle ère des sciences humaines et sociales ?*, Paris : CNRS Éditions, 2013, 64 p.

ZELNIK, Patrick, TOUBON, Jacques et CERUTTI, Guillaume, *Création et Internet*, rapport remis en janvier 2010 au Ministre de la Culture et de la Communication, 2010, disponible sur le Web : <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/104000006.pdf>

Articles et revues

Annales, « Éditorial », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Éditions de l'EHESS, 2/2015, p. 285-287, disponible sur Cairn : <http://www.cairn.info/revue-Annales-2015-2-page-285.htm>

ALLIER, Bénédicte et EMPTOZ, Hubert, « Le traitement des images au service du document patrimonial, vers une pluridisciplinarité », dans « Numérisation et patrimoine », COÜASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert (dir.), *Document numérique*, vol. 7, 2003/3-4, p. 143-157, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

ALMEIDA, Rodrigo, CUBAUD, Pierre, DUPIRE, Jérôme et TOPOL, Alexandre, « Interactions et métadonnées riches pour les bibliothèques numérisées », dans « Visualisation pour les bibliothèques numériques », FEKETE, Jean-Daniel et LECOLINET, Éric (dir.), *Revue Document numérique*, vol. 9, 2006/2, p. 83-109, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2006-2.htm>

ANDRÉ, Jacques, « Numérisation et codage des caractères de livres anciens », dans « Numérisation et patrimoine », COÜASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert (dir.), *Document numérique*, vol. 7, 2003/3-4, p. 127-142, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

ANDRÉ, Jacques et JIMENES, Rémi, « Transcription et codage des imprimés de la Renaissance : réflexions pour un inventaire de caractères anciens », dans « Gestion informatisée des écritures anciennes », BÉNÉVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 113-139, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2013-3.htm>

ARMITAGE, David et GULDI, Jo, traduction de BAUDRY, Jérôme, « Le retour de la longue durée, une perspective anglo-américaine », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Éditions de l'EHESS, 2/2015, p. 289-318, disponible sur Cairn : <http://www.cairn.info/revue-Annales-2015-2-page-289.htm#no5>

ARMITAGE, David et GULDI, Jo, traduction de HEUDRE, Antoine, « Pour une histoire ambitieuse », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70e année), Éditions de l'EHESS, 2/2015, p. 367-378, disponible sur Cairn : <http://www.cairn.info/revue-Annales-2015-2-page-367.htm>

BAILLOT, Anne, « Reconstruire ce qui manque – ou le déconstruire ? Approches numériques des sources historiques. », *Entre sources, données et réseaux* (programme interdisciplinaire MOREHIST), Toulouse, Mars 2015 : <https://halshs.archives-ouvertes.fr/halshs-01133507/document>

BERTRAND, Paul, MAIREY, Aude, GUYOTJEANNIN, Olivier, *et al.*, « L'historien médiéviste et la pratique des textes : les enjeux du tournant numérique », dans « Être historien du Moyen Âge au XXI^e siècle », *Actes des congrès de la Société des historiens médiévistes de l'enseignement supérieur public*, vol. 38, n°1, 2007, p. 273-301, disponible sur Persée : http://www.persee.fr/doc/shmes_1261-9078_2008_act_38_1_1958

BLOUIN, X., Francis, « Deux sphères conceptuelles distinctes : le classement des archives et la recherche historique », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 96-99, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

BONIN, Emmanuel et DALLO, Alain, « Hyperbase et Lexico 3, outils lexicométriques pour l'historien », *Histoire & mesure*, XVIII - 3/4, 2003, mis en ligne le 03 avril 2007 : <http://histoiremesure.revues.org/840>

BOULLIER, Dominique, « Vie et mort des sciences sociales avec le big data », dans « Le tournant numérique... et après ? », *Socio*, 4, mars 2015, p. 19-38, disponible sur le Web : <http://socio.revues.org/1259>

BOURGEOIS, Frank, EMPTOZ, Hubert et TRINH, Éric, « Compression et accessibilité aux images de documents numérisés : application au projet Debora », dans « Numérisation et patrimoine », COÛASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert (dir.), *Document numérique*, vol. 7, 2003/3-4, p. 103-125, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

BRIAN, Éric, « L'horizon nouveau de l'historiographie expérimentale », dans « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 41-56, disponible sur Cairn : <http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5.htm>

CHATEAURAYNAUD, Francis, « Moteurs de (la) recherche et pragmatique de l'enquête. Les sciences sociales face au Web collectionniste », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre*

temps, n° 82, 2006/2, p. 109-118, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

CLIVAZ, Claire et VINCK, Dominique, « Introduction : des humanités délivrées pour une littérature plurielle », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 9-16, disponible sur Cairn : <http://www.cairn.info/revue-les-cahiers-du-numerique-2014-3-page-9.htm#no1>

COMBE, Sonia, « « L'historien face à l'ordre informatique : classification et histoire » », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 1-3, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

COMBE, Sonia, « De Gabriel Naudé à Rameau, les nouvelles conditions de la production historique », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 4-7, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

CORSINI, Silvio, « Passe-Partout, banque internationale d'ornements d'imprimerie », *Bulletin des bibliothèques de France*, n°5, 2001, p. 73-79, disponible à l'adresse suivante : <http://bbf.enssib.fr/consulter/bbf-2001-05-0073-010>

COÛASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert, « Numérisation et patrimoine : introduction », dans « Numérisation et patrimoine », *Document numérique*, vol. 7, 2003/3-4, p. 7-12, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

COUSTATY, Mickaël, TSOPZE, Norbert, BERTET, Karell, BOUJU, Alain et LOUIS, Georges, « Traitement des documents anciens à l'aide d'ontologie », dans « Valorisation des corpus numérisés », CHEVRY PEBAYLE, Emmanuelle (dir.), *Les Cahiers du numérique*, vol. 8, 2012/3, p. 91-118 : <http://www.cairn.info/bibliotheque-nomade2.univ-lyon2.fr/revue-les-cahiers-du-numerique-2012-3.htm>

COUSTATY, Mickaël, OGIER, Jean-Marc, PARETI, Rudolf et VINCENT Nicole, « Extraction d'informations d'images de documents anciens », Congrès des jeunes chercheurs en vision par ordinateur, Trégastel, France, 2009, consultable sur Archives ouvertes : <https://hal.archives-ouvertes.fr/hal-00435382/document>

COUSTATY Mickaël, SIDERE, Nicolas, OGIER, Jean-Marc, HEROUX, Pierre, RAMEL Jean-Yves, et al., « Content-Based Old Documents Indexing », OGIER, Jean-Marc, WENYIN, Liu, LLADOS, Josep, *Eight International Workshop on Graphics Recognition, GREC 2009*, La Rochelle, 2009, p. 217-223, disponible sur Archives ouvertes : <https://hal.archives-ouvertes.fr/hal-00382086/PDF/grec02.pdf>

DIMINESCU, Dana et WIEVIORKA, Michel, « Le défi numérique pour les sciences sociales », dans « Le tournant numérique... et après ? », *Socio*, 4, mars 2015, p. 9-18, disponible sur le Web : <http://socio.revues.org/1254>

DELALANDE, Nicolas et VINCENT, Julien, « Portrait de l'historien-ne en cyborg », dans « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 5-29, disponible sur Cairn : <http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5.htm>

DREYFUS-ARMAND, Geneviève, « De la moissonneuse-batteuse au micro-ordinateur relié au monde entier », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 80-81, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

ROIXHE, Daniel, STIENNON, Stéphane et VANWELKENHUYZEN, Nadine, « Le projet Môriâne », dans les Actes du colloque *Vers une nouvelle érudition : la numérisation au service de l'histoire du livre*, codirigé par VARRY, Dominique, BAUDRY, Guylaine et CHARON, Annie, à l'occasion des 12e entretiens du Centre Jacques Cartier, Lyon-Villeurbanne, Enssib, 6-7 décembre 1999, disponible sur le Web : <http://www.enssib.fr/bibliotheque-numerique/notices/1510-le-projet-moriane>

DUFOURNAUD, Nicole, « Des humanités aux données. Méthodes en histoire à l'ère du numérique, témoignage d'une e-historienne », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 73-88, disponible sur Cairn : <http://www.cairn.info/revue-les-cahiers-du-numerique-2014-3-page-73.htm>

DUFOURNAUD, Nicole et FEKETE, Jean-Daniel, « Comparaison d'outils pour la visualisation de sources historiques codées en XML/TEI », dans « Visualisation pour les bibliothèques numériques », FEKETE, Jean-Daniel et LECOLINET, Éric (dir.), *Document numérique*, vol. 9, 2006/2, p. 37-56, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2006-2.htm>

EL HADI WIDAD, Mustafa et FAVIER, Laurence, « Bibliothèques numériques et humanités digitales. La continuité d'une problématique sur l'organisation des connaissances en lettres et sciences humaines », *Les Cahiers du numérique*, Vol. 11, 2015/1, p. 83-102

FEKETE, Jean-Daniel et LECOLINET, Éric, « Visualisation pour les bibliothèques numériques : introduction », dans « Visualisation pour les bibliothèques numériques », dans FEKETE, Jean-Daniel et LECOLINET, Éric (dir.), *Document numérique*, vol. 9, 2006/2, p. 7-11, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2006-2.htm>

FERRO, Marc, « L'historien face à l'ordre informatique. Conclusion », dans : « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 121-122, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

FOULONNEAU, Muriel, « Recherche et numérisation du patrimoine en Europe », dans « Numérisation et patrimoine », COÛASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert (dir.), *Document numérique*, vol. 7, 2003/3-4, p. 179-189, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

GAUDET, Françoise, « Rameau et l'OPAC, vers une googlisation des pratiques ? », dans « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 89-91, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

GEFEN, Alexandre, « Les enjeux épistémologiques des humanités numériques », dans « Le tournant numérique... et après ? », *Socio*, 4, mars 2015, p. 61-74, disponible sur le Web : <http://socio.revues.org/1296>

GRANDJEAN, Martin, « La connaissance est un réseau : perspective sur l'organisation archivistique et encyclopédique », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 37-54, disponible sur Cairn : <http://www.cairn.info/revue-les-cahiers-du-numerique-2014-3-page-37.htm>

GRAU, Brigitte, LALEAU, Régine et RAMEL, Jean-Yves, « Introduction », dans « Systèmes d'indexation et recherche d'information : modèles pour la conception et l'évaluation », GRAU, Brigitte, LALEAU, Régine et RAMEL, Jean-Yves (dir.), *Document numérique*, vol. 14, 2011/2, p. 7-10, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2011-2.htm>

GUICHARD, Éric, « L'internet et les épistémologies des sciences humaines et sociales », *Revue Sciences/Lettres*, n° 2, 2014 : <http://rsl.revues.org/389>

HEIMBURGER, Franziska et RUIZ, Émilien, « Faire de l'histoire à l'ère numérique : retours d'expériences. », dans « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 70-89, disponible sur Cairn : <http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5.htm>

HUNT, Lynn, traduction de HEUDRE, Antoine, « Faut-il réinitialiser l'histoire ? », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70^e année), Éditions de l'EHESS, 2/2015, p. 319-325, disponible sur Cairn : <http://www.cairn.info/revue-Annales-2015-2-page-319.htm>

JIMENES, Rémi, « Introduction », dans « Gestion informatisée des écritures anciennes », BÉNÉVENT, Christine, JIMENES, Rémi et SARAH, Guillaume (dir.), *Document numérique*, vol. 16, 2013/3, p. 7-13, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2013-3.htm>

JIMENES, Rémi, « La base de Typographie de la Renaissance (BaTyR) », *Bulletin des bibliothèques de France*, n° 5, 2013, p.18-22, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2013-05-0018-004.pdf>

JIMENES, Rémi, « Pratique d'atelier et corrections typographiques à Paris au XVI^e siècle : les œuvres de saint Bernard imprimées par Charlotte Guillard (1551) », BÉNÉVENT Christine, CHARON Annie, DIU Isabelle et VENE Magali (éd.), *Passeurs de texte, imprimeurs et libraires à l'âge de l'humanisme*, Paris : École nationale des chartes, 2009, p. 205-228

JOURNET, Nicholas, MULLOT, Rémy, EGLIN, Véronique et RAMEL, Jean-Yves, « Analyse d'images de documents anciens : une approche texture », *Traitement du signal*, Presses universitaires de Grenoble, vol. 24, n° 6, 2008, p. 461-479, disponible sur Archives ouvertes : <https://hal.archives-ouvertes.fr/hal-00355243/document>

JOUTARD, Philippe, « Révolution numérique et rapport au passé », dans « La culture du passé », *Le Débat*, n° 177, 2013/5, p.145-152, disponible sur Cairn : <https://www.cairn.info/revue-le-debat-2013-5.htm>

LAMASSÉ, Stéphane et RYGIEL, Philippe, « Nouvelles frontières de l'historien », *Revue Sciences/Lettres*, n° 2, 2014 : <http://rsl.revues.org/411>

LEMERCIER, Claire, « Une histoire sans sciences sociales ? », dans « La longue durée en débat », *Annales. Histoire, Sciences Sociales* (70^e année), Éditions de l'EHESS, 2/2015, p. 345-357, disponible sur Cairn : www.cairn.info/revue-annales-2015-2-page-345.htm

LIKFORMAN-SULEM, Laurence, « Apport du traitement des images à la numérisation des documents manuscrits anciens », dans « Numérisation et patrimoine », COÛASNON, Bertrand, DALBERA, Jean-Pierre et EMPTOZ, Hubert (dir.), *Document numérique*, vol. 7, 2003/3-4, p. 13-26, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2003-3.htm>

MAGNANI, Éliane et RUSSO, Daniel, « Histoire, philosophie, iconographie, humanités numériques : une interdisciplinarité difficile », *Bulletin du centre d'études médiévales d'Auxerre* | BUCEMA, n° 16, 2012 : <http://cem.revues.org/12285>

MICHEL, Jean-Baptiste, KUI SHEN, Yuan, et al., « *Quantitative Analysis of Culture Using Millions of Digitized Books* », *Science*, vol. 331, n° 6014, 14 janvier 2011, p. 176-182

MOATTI, Claudia, « L'e-story ou le nouveau mythe hollywoodien », dans « La longue durée en débat », *Annales, Histoire, Sciences Sociales* (70^e année), Éditions de l'EHESS, 2/2015, p. 237-332, disponible sur Cairn : <http://www.cairn.info/revue-annales-2015-2-page-327.htm>

MONTREUIL, Florent, NICOLAS, Stéphane, HEUTTE, Laurent, GROSICKI, Emmanuèle, « Intégration d'informations textuelles de haut niveau en analyse de structures de documents manuscrits non contraints », dans « Systèmes d'indexation et recherche d'information : modèles pour la conception et l'évaluation », GRAU, Brigitte, LALEAU, Régine et RAMEL, Jean-Yves (dir.), *Document numérique*, vol. 14, 2011/2, p. 77-101, disponible sur Cairn : <https://www.cairn.info/revue-document-numerique-2011-2.htm>

NUNBERG, Geoff, « *Google's Book Search : A Disaster for Scholars* », *The Chronicle of Higher Education*, 31 août 2009, disponible sur le Web : <http://chronicle.com/article/Googles-Book-Search-A/48245/>

OLIVIER, Mathieu, « Ménéstrel : portail francophone de référence en études médiévales, et partenaire de l'IFHA », *Revue de l'IFHA*, 1/2009, p. 113-116, mis en ligne le 7 février 2013, consulté le 8 mai 2016, disponible ici : <http://ifha.revues.org/340>

POTIN, Yann, « Institutions et pratiques d'archives face à la numérisation : expériences et malentendus », dans « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 57-69, disponible sur Cairn : <http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5.htm>

ROCHAT, Yannick et KAPLAN, Frédéric, « Analyse des réseaux de personnages dans les *Confessions* de Jean-Jacques Rousseau », dans « Les humanités délivrées », *Les Cahiers du numérique*, Vol. 10, 2014/3, p. 109-133, disponible sur Cairn : <http://www.cairn.info/revue-les-cahiers-du-numerique-2014-3-page-109.htm>

RYGIEL, Philippe, « L'enquête historique à l'ère du numérique », dans « Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ? », *Revue d'histoire moderne et contemporaine*, supplément n°58-4bis, 2011/5, p. 30-40, disponible sur Cairn : <http://www.cairn.info/revue-d-histoire-moderne-et-contemporaine-2011-5.htm>

RYGIEL, Philippe, « L'ordinateur, le réseau et l'écriture de l'histoire », dans : « L'historien face à l'ordre informatique : classification et histoire », *Matériaux pour l'histoire de notre temps*, n° 82, 2006/2, p. 75-79, disponible sur Cairn : <http://www.cairn.info/revue-materiaux-pour-l-histoire-de-notre-temps-2006-2.htm>

RYGIEL, Philippe, « Des archives numériques sans historiens ? Un point de vue », dans « Historiens et usages publics du passé », *Matériaux pour l'histoire de notre temps*, n° 79, 2005, p. 11-13, disponible sur Persée : http://www.persee.fr/web/revues/home/prescript/article/mat_0769-3206_2005_num_79_1_1033

RYGIEL, Philippe, « Les sources de l'historien à l'heure d'Internet », *Hypothèses*, 2004/1 (7), p. 341-354, disponible sur Cairn : <http://www.cairn.info/revue-hypotheses-2004-1-page-341.htm>

SOULATGES, Magali, « Les bases de données d'ornements typographiques du XVIII^e siècle », *Bulletin des bibliothèques de France*, n° 6, 2012, p. 63-64, disponible sur le Web : <http://bbf.enssib.fr/consulter/bbf-2012-06-0063-005>

ZHANG, Dengsheng. ISLAM, Md. Monirul, et LU, Guojun, « A review on automatic image annotation techniques. », *Pattern Recognition*, vol. 45, 2012, p. 346-362

TABLE DES ILLUSTRATIONS

Figure 1. Sur le site « Printers' Devices » de l'Université de Barcelone, la fiche technique de l'imprimeur Charles Angot avec l'image de sa marque d'imprimerie.	33
Figure 2. Exemple de bandeau sur la base Môriâne.	35
Figure 3. Exemple de vignette sur la base Môriâne.....	35
Figure 4. Exemple de pagination ornementée dans la base Fleuron et l'ouvrage dans lequel elle apparaît.	36
Figure 5. Exemple de lettrine dans la base Fleuron et les ouvrages dans laquelle elle a été référencée.	37
Figure 6. Exemple d'encadrement dans la base Fleuron et l'ouvrage dans lequel il apparaît.....	37
Figure 7. Exemple de recherche sur la base Fleuron.	38
Figure 8. Marques de lecture sur la treizième page de la <i>Biblia latina</i> (1 ^{er} volume) publiée à Mayence vers 1455 par J. Gutenberg et J. Fust et conservé par la Bibliothèque Mazarine de Paris.	49
Figure 9. Image d'un filigrane et de lignes de chaînes visibles par transparence au travers du papier d'une page de garde.....	50
Figure 10. Zoom sur l'utilisation de deux types distincts, une lettre E grande capitale et une apostrophe, pour figurer un « E accent aigu ».....	60
Figure 11. Exemple de lettres volontairement renversées.	60
Figure 12. Zoom sur des marques d'épaulement laissées par un encrage trop important des caractères mobiles.	60
Figure 13. Onzième page du texte des <i>Méditations métaphysiques</i> de René Descartes, 3 ^e édition, édité chez M. Bobin et N. Le Gras à Paris en 1673, conservé à la Bibliothèque nationale de France sous la cote R-3503.....	61
Figure 14. Exemple d'une page abimée, en mode image et en mode texte..	63
Figure 15. Exemple d'une page rognée à la numérisation, d'une page inclinée et d'une autre courbée.	64
Figure 16. Zoom sur une page abimée de livre ancien.	65
Figure 17. Page de titre du <i>Geometria à Renato Descartes (...)</i> , Lugdunum Batavorum, J. Maire, 1649 (Paris, BnF).	66
Figure 18. Page de titre de <i>La Géométrie de René Descartes</i> , Paris, chez Charles Angot, 1664 (Vienne, Bibliothèque nationale d'Autriche).....	67
Figure 19. Page de titre du <i>Geometria à Renato Des Cartes (...)</i> , Amstelodami, ex typographia Blaviana, 1683 (Paris, BnF).	67
Figure 20. La page 7 de <i>La Géométrie de René Descartes</i> , Paris, chez Charles Angot, 1664 (Vienne, Bibliothèque nationale d'Autriche), version en ligne et même image une fois téléchargée.....	69
Figure 21. La même page en version nativement numérique constituée dans le cadre du projet Gutenberg.	70
Figure 22. Pages 136 et 137 de la <i>Biblia latina</i> (1 ^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1-1), présentées en double page.	71
Figure 23. Fenollar, Bernat de. <i>Lo Procés de les olives e disputa dels jovens hidels vells. Fet per alguns trobadors avant nomenats e lo sompni de Johan Johan</i> ,	

Valence, Lope de la Roca, 1497 (Paris, Bibliothèque Mazarine, Inc 943), couverture.	72
Figures 24 et 25. Dos et plat supérieur de la <i>Biblia latina</i> (1 ^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1- 1).	72
Figure 26. Latini (Brunetto), <i>El tesoro</i> ; Trad. Bono Giamboni, Treviso, Gerard de Lisa, 1474 (Paris, Bibliothèque Mazarine, Inc 95), zoom sur les pages 64, 91 et 21.	73
Figure 27. Zoom sur le dos de la reliure de la <i>Biblia latina</i> (1 ^{er} volume), Mayence, J. Gutenberg et J. Fust, vers 1455 (Paris, Bibliothèque Mazarine, Inc 1- 1).	74
Figure 28. Exemple d'ornement issu de la base Môriâne, utilisé par l'imprimeur Bassompierre à Liège entre 1769 et 1793.	80
Figure 29. Exemple d'ornement issu de la base Môriâne, également utilisé et vendu par l'imprimeur Bassompierre à Liège entre 1744 et 1793.	81
Figure 30. Exemple de décomposition en couches d'une lettrine entre lettre, motif, fond et cadre.	83
Figure 31. Exemple de décomposition en couches d'une lettrine entre forme et texture.	83
Figure 32. Première page d'une dédicace de Pierre Corneille au Cardinal Mazarin, accompagnée de sa version texte sur Gallica.	86
Figure 33. Caractère mobile tombé sur la forme imprimante.	87
Figure 34. Exemple d'un caractère mobile retourné pied-en-haut pour indiquer que le caractère « ù » manque à la presse.	87
Figure 35. Schéma du caractère mobile en plomb typographique de la lettre F grande capitale.	88
Figure 36. Réseau des cooccurrences des personnages des <i>Confessions</i> de Jean-Jacques Rousseau.	97
Figure 37. Exemple de Wordle réalisé sur l'application en ligne <i>wordle.net</i> à partir de notre partie sur « Le vrac » (p. 23-25).	98
Figure 38. Exemple d'utilisation de l'application Ngram Viewer, sur la cooccurrence des termes « René Descartes », « Jean de la Fontaine » et « Blaise Pascal » depuis 1650 dans les ouvrages français numérisés par Google Livres...100	
Figure 39. Exemple d'utilisation de l'application Ngram Viewer, sur la cooccurrence des termes « Zola », « Balzac » et « Flaubert » depuis 1800 dans les ouvrages français numérisés par Google Livres.	100

TABLE DES MATIÈRES

SIGLES ET ABRÉVIATIONS	9
INTRODUCTION.....	11
L’OFFRE NUMERIQUE	15
LA BIBLIOTHEQUE UNIVERSELLE S’APPELLERA-T-ELLE	
GOOGLE LIVRES ?.....	15
<i>Le mythe de la bibliothèque universelle</i>	<i>15</i>
<i>Une offre pléthorique.....</i>	<i>17</i>
<i>La suprématie googléienne.....</i>	<i>22</i>
QUE RESTE-T-IL DE NOS BIBLIOTHEQUES ?	24
<i>Le « vrac »</i>	<i>25</i>
<i>Existe-t-il une « bibliothèque numérique » ? Pour quels publics ?</i>	<i>27</i>
<i>Ressources thématiques, les nouvelles collections ?.....</i>	<i>30</i>
Projets « disciplinaires »	30
Projets « genrés », l’exemple des ornements d’imprimerie	32
Projets linguistiques	40
<i>Une politique de numérisation ?.....</i>	<i>40</i>
LA NATURE DE L’OFFRE	43
A LA SOURCE DU TRAVAIL DE L’HISTORIEN, UN FICHIER	
NUMERIQUE ?	43
<i>La valeur de la source : sa stabilité ? Lutte contre l’instabilité</i> <i>ontologique du fichier numérique</i>	<i>43</i>
<i>La searchability du fichier, condition de son existence ?</i>	<i>45</i>
<i>En amont : l’indexation et les métadonnées</i>	<i>48</i>
<i>En aval : conservation et pérennisation.....</i>	<i>52</i>
LE LIVRE NUMERISE.....	54
<i>Quid de l’original ?</i>	<i>54</i>
<i>Le processus de numérisation.....</i>	<i>57</i>
<i>Mode texte / Mode image</i>	<i>58</i>
<i>Quel rendu ?.....</i>	<i>63</i>
TRAVAIL DES SOURCES	75
REALITE MATERIELLE DE L’IMMATERIEL	
<i>Quid de la matérialité du livre ?</i>	<i>75</i>
<i>Bibliographie (im)matérielle, fondements d’une archéologie du livre</i> <i>numérisé</i>	<i>77</i>
<i>Les espoirs du numérique.....</i>	<i>79</i>
QUE RESTE-T-IL DE L’OBJET-LIVRE ?	85
<i>La reconnaissance optique de caractères, mémoire de l’art</i> <i>typographique</i>	<i>85</i>
<i>La recherche du fragment, la découverte du Tout-venant.....</i>	<i>91</i>

<i>La mort de l'objet-livre, essai de philologie numérique</i>	92
TRAVAILLER AVEC CES SOURCES	95
FAUT-IL SE JOUER DES DONNEES ?	95
<i>Le traitement des données</i>	95
« <i>Culturomique</i> » et « <i>quantophrénie</i> »	99
<i>L'affrontement entre « quantitatif » et « qualitatif »</i>	103
<i>Une crise historiographique ?</i>	105
DE LA SCIENCE HISTORIQUE A LA SCIENCE INFORMATIQUE :	
FONDEMENTS D'UNE NOUVELLE DISCIPLINE ?	108
<i>L'avènement du « cyberhistorien » ?</i>	108
<i>Nouvelles contraintes disciplinaires</i>	110
<i>Les « humanités numériques », aux fondements d'une nouvelle</i>	
<i>épistémologie ?</i>	112
« <i>Transdiscipline</i> » et « <i>Data Sciences</i> »	115
CONCLUSION	117
BIBLIOGRAPHIE	121
Le livre, les bibliothèques et le numérique	121
<i>Monographies</i>	121
<i>Articles et revues</i>	123
L'historien, ses outils et le numérique	126
<i>Monographies</i>	126
<i>Articles et revues</i>	128
TABLE DES ILLUSTRATIONS	137
TABLE DES MATIÈRES	139