

L'histoire et le numérique : vers de nouveaux modes de lecture des sources

Frédéric Clavert

Docteur en histoire contemporaine. Ingénieur de recherche (LabEx EHNE / Paris Sorbonne)

(Version auteur de : Clavert, Frédéric. « Vers de nouveaux modes de lecture des sources ». In *Le temps des humanités digitales*, édité par Olivier Le Deuff. FYP EDITIONS, 2014.)

Introduction

Les usages de l'informatique dans les sciences historiques en France sont à peu près aussi anciens que l'ordinateur lui-même. La France accueille les premiers supercalculateurs dans les années 1950. Le premier article majeur mentionnant l'usage de méthodes « mécanographiques » - le terme informatique n'existe pas encore - est publié par les *Annales* en 1959. François Furet et Adeline Daumard y évoquent les aspects méthodologiques de l'usage de la mécanographie dans la lecture d'archives notariales. Les sciences historiques entrent alors dans une période marquée par l'histoire économique et sociale quantitative fondée sur de grandes séries statistiques, période qui se prête particulièrement bien aux usages informatiques émergents. Dès lors, les historiens – du moins certains d'entre eux – n'ont cessé d'utiliser les outils informatiques.

Les années 1960 voient l'émergence d'historiens et de thèses majeurs ayant intensément utilisé l'informatique: *L'Armée française de la fin du XVIIe siècle au ministère Choiseul* (1964) d'André Corvisier¹ et *Les paysans du Languedoc* d'Emmanuel Le Roy Ladurie² en 1966 font partie des travaux de grande ampleur, pionniers à la fois en histoire économique et histoire militaire et dans l'utilisation de l'ordinateur en histoire. Ces deux thèses reposent sur l'exploitation de corpus documentaires suffisamment imposants pour entraver une exploitation strictement humaine de ces données. Le saut qualitatif qui a été alors pour partie rendu possible par le traitement de larges corpus de données quantitatives par l'informatique est celui du passage vers l'histoire du temps long, le passage du particulier aux masses, de l'histoire bataille à l'histoire des soldats. C'est, parallèlement, à cette époque (1962) qu'est inventé le terme « informatique », contraction d'« information » et d'« automatique »³.

Le rôle de l'École des Annales dans l'émergence sur le long terme des Humanités numériques est alors déterminant. Dans son ouvrage *Graphs, Maps and Trees*, Franco Moretti⁴ s'interroge sur la manière dont l'histoire de la littérature pourrait devenir moins particulière en cessant de ne s'attacher qu'aux grands auteurs pour envisager une histoire plus globalisante de la littérature. Or, la première référence qu'il cite est l'École des Annales. Toutefois, les historiens de l'École des Annales ne sont pas les seuls à utiliser l'informatique à cette époque⁵. La *New economic history* aux États Unis en use afin de « transformer l'histoire économique inspirée de théories implicites, vagues et non testées en une discipline fondée sur des théories définies avec rigueur et appuyées par des faits »⁶. C'est pour partie la démarche qu'ont aussi suivi les premiers chercheurs en *linguistics computing*⁷ à la même époque.

En France, dans les années 1970 et 1980, l'emploi de l'informatique stagne. Il n'y a pas eu réalisation de la fameuse prophétie d'Emmanuel Leroy-Ladurie : « l'historien sera programmeur ou ne sera pas »⁸, probablement pour plusieurs raisons : le reflux de l'histoire économique fondée sur de grandes séries statistiques, mais également l'évolution de

¹ André Corvisier, *L'Armée française de la fin du XVIIe siècle au ministère de Choiseul. Le soldat.*, vol. 1-2, 2 vol. (Paris: Presses Universitaires de France, 1964).

² Emmanuel Le Roy Ladurie, *Les Paysans du Languedoc* (Paris: S.E.V.P.E.N., 1966).

³ Nous signalons les travaux d'Eric Castex, notamment son mémoire de DEA : *Histoire et Informatique du milieu des années 50 à nos jours*, Toulouse-Le Mirail. Sa thèse, *Histoire et Informatique en France*, sous la codirection de J-F. Soulet (Toulouse-Le Mirail) et de J-P Genet (Paris I Panthéon Sorbonne) est malheureusement inachevée.

⁴ Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (Verso, 2007).

⁵Voir Peter Haber, *Digital Past. Geschichtswissenschaften im digitalen Zeitalter* (München: Oldenbourg Wissenschaftsverlag, 2011).

⁶ Citation de Robert Fogel, dans « La new economic history », *Alternatives économiques* (n° 137 - mai 1996), dont je n'ai pas retrouvé l'original. Parmi les classiques de Fogel: *Railroads and American Economic Growth. Essays in Economic History*. Baltimore, Maryland. The John Hopkins Press. 1964.

⁷ En utilisant Google ngrams et Google Books, on s'aperçoit que la'une des inspirations des *linguistics computing* est bien l'économétrie : http://books.google.fr/books?ei=BsYdUvmDOOuR1AXMoYHAAQ&hl=fr&id=jPVYAAAAMAAJ&dq=%22linguistic+computing%22&q=economic#search_anchor.

l'informatique vers les ordinateurs personnels qui ont répandu certains usages de l'informatique mais n'ont pas poussé les historiens à en approfondir l'utilisation.

La sortie des ordinateurs personnels et, quelques années plus tard, la généralisation des interfaces graphiques (Macintosh en 1984, MS Windows l'année suivante) démocratisent l'informatique, malgré le coût élevé des premiers ordinateurs personnels, y compris en histoire, d'autant plus que sont alors disponibles des tableurs, qui permettent un traitement des données statistiques. Ce qui pouvait se faire avec des fiches cartonnées dans des conditions matérielles difficiles devenait aisé avec l'informatique personnelle.

Mais, malgré, par exemple, la fondation du *Médiéviste et l'ordinateur* dans les années 1970, le tournant informatique de l'histoire n'arrive pas, même si André Zysberg estime en 1980 que de plus en plus d'historiens désirent se former à l'informatique⁹. Le peu d'historiens utilisant l'informatique est toutefois très actif à la fin des années 1980. La création d'une association et d'une revue (*Mémoire Vive*) en est la preuve. Cette activité s'est vue assez vite limitée : *Mémoire Vive* n'a publié que neuf numéros et l'association n'est plus active¹⁰. Par contre, la réflexion sur l'usage de l'ordinateur en histoire n'a jamais cessé. Certaines revues, notamment *Histoire & Mesure*, ont permis la publication d'articles de réflexion importants. Des historiens se sont engagés dans les méthodologies informatiques appliquées à l'histoire depuis la fin des années 1980, notamment autour de Jean-Philippe Genet.

Après l'arrivée des premiers supercalculateurs, l'apparition de l'informatique personnelle et de ses interfaces graphiques, un troisième grand tournant intervient : celui du réseau. L'Internet et le développement du Web à partir de 1989 changent la donne. Internet et ses possibles implications mettent progressivement en réseau des historiens – plus qu'ils ne l'étaient déjà –, des projets, des bases de données, des logiciels, des revues, des publications... Le web met également à disposition de tous, chercheurs, enseignants et étudiants de tous les niveaux, des sources secondaires et primaires, certes parfois de qualité variable.

Ce troisième grand tournant remet en avant l'utilisation de l'outil informatique en histoire, mais aussi relance des questions liées à cette utilisation : quelle conséquence pour les relations entre l'historien et ses sources, pour l'écriture et l'enseignement de l'histoire ? Cette réflexion autour de la pratique de l'histoire à l'ère numérique, nous pousse également à considérer les relations entre les historiens et tous ceux, informaticiens, archivistes, etc., qui participent à la production de l'histoire. Ainsi, devons-nous repenser la globalité du processus de production de l'histoire.

Cette mise en réseau semble avoir réveillé l'histoire numérique. Si, aux Etats-Unis, ce réveil se concrétise dès le milieu des années 1990 par la création d'un centre de recherche comme le *Roy Rosenzweig Center for History and New Media*, de l'université George Mason¹¹, en France ce réveil se traduit par l'émergence d'une génération d'historiens, dispersés dans diverses institutions, utilisant couramment l'informatique. A Paris I autour de J.-P. Genet, à Sciences Po et à l'EHESS autour de Claire Lemerrier¹² et à l'ENS autour de Jean-Luc Pinol¹³, la caractéristique de cette génération, intégrée pour beaucoup dans le courant plus général des Humanités numériques, est aussi d'utiliser à plein et couramment le web et ses possibilités - les carnets de recherche, les réseaux sociaux, les logiciels d'édition de contenu sur le web – et de profiter de l'émergence des mouvements de l'*open source*, de l'*open data*, etc., pour la recherche comme pour l'enseignement¹⁴.

⁸ Emmanuel Le Roy Ladurie, *Le territoire de l'historien* (Gallimard, 1973).

⁹ André Zysberg, « 8 mars 1980, place aux modernistes. Table Ronde "Histoire et informatique" (ENS) », *Le médiéviste et l'ordinateur* no 4 (Automne 1980): 16-17.

¹⁰ Le site de Mémoire Vive et de l'association est disponible ici : <http://pireh.univ-paris1.fr/mv/index.html>

¹¹ <http://chnm.gmu.edu/>

¹² Voir Claire Lemerrier, « Analyse de réseaux et histoire », *Revue d'histoire moderne et contemporaine* no 2 (2005): 88–112.

¹³ Jean-Luc Pinol, « Les systèmes d'information géographique et la pratique de l'histoire », *Revue d'histoire moderne et contemporaine* n° 58-4bis, no 5 (2011): 111-126.

¹⁴ Émilien Ruiz et Franziska Heimburger, « Faire de l'histoire à l'ère numérique : retours d'expériences », *Revue d'histoire moderne et contemporaine* n° 58-4bis, no 5 (2011): 70-89.

Cette longue tradition, associée à des tendances récentes et plus générales comme le *big data*¹⁵, révèle un phénomène de fond qui s'exprime depuis les années 1960 : la mise en données de l'histoire. « Mettre en données » est aujourd'hui largement relié à la notion de numérisation. Mais elle en est distincte. D'une part, elle n'implique ni informatique, ni numérique. D'autre part, à l'ère numérique, elle dépasse largement la notion de document et peut s'en distinguer¹⁶.

Gérer l'information : la mise en données de l'histoire

Adeline Daumard et François Furet avaient beau parler de « mécanographie », lorsque l'on lit le détail de l'article, il s'agit d'un réel premier usage informatique en histoire, et, surtout, des premières bases données¹⁷ : il y avait nécessité de transformer des informations historiques en une grille chiffrée, soigneusement codifiée – en clair, même si cela se traduit à l'époque par la création de fiches perforées, la transformation des sources primaires en bases de données -, nécessité d'un travail collectif tant il est gigantesque... mais en contrepartie, des apports nouveaux pour l'histoire (sociale dans le cas de cet article) :

« La mécanisation du travail de classement et de sélection des fichiers serait une immense amélioration, qualitative et quantitative, de nos enquêtes, car les archives notariales accéderaient à la dignité statistique, sans pour autant perdre leur valeur typologique. [...] la mécanographie valoriserait ainsi non seulement le travail historique, mais nos sources mêmes. »¹⁸.

Deux ans après l'article de François Furet et Adeline Daumard, toujours dans les *Annales ESC*, Paul Garelli et Jean-Claude Gardin se penchent sur l'*Étude par ordinateurs des établissements assyriens en Cappadoce*¹⁹. Notant que publier une méthodologie d'utilisation d'un ordinateur est devenu, en dix ans, classique et monotone, remarquant que cette utilisation des ordinateurs est en passe de susciter la création d'une nouvelle discipline – qu'ils ne nomment pas : le terme français « informatique » n'est inventé que quelques années plus tard – ce second article insiste toutefois sur un aspect de l'utilisation de l'informatique qui est toujours valable aujourd'hui pour les sciences historiques :

« la mise au point d'une procédure d'exploitation automatique de telles ou telles données, trop nombreuses ou trop complexes pour qu'on les puisse aisément maîtriser par d'autres voies »²⁰.

Les auteurs détaillent trois utilisations possibles de l'informatique en histoire : les statistiques, le tri, l'injection de commandes pour manipuler les données qui sont en mémoire de l'ordinateur afin de dégager des éléments qui ne sont pas visibles aisément. Les auteurs peuvent ainsi faire une répartition géographique de ces établissements, dessiner le réseau des commerçants assyriens (avec représentations graphiques, tri des types de groupes, etc.), définir la spécialisation des marchands... Cet article se centre sur l'acquis méthodologique de ce qui était alors une enquête en cours. À ma connaissance, c'est le premier exemple, en France, d'une volonté de développer un programme informatique servant une recherche historique.

Le saut important que représente cet article est aussi lié – c'est évoqué par les auteurs en conclusion – à une progression technique des ordinateurs et à l'accès à de nouvelles facilités, par la création d'Euratom à l'occasion des traités de Rome signés en 1957. Euratom, bien que quasiment sans objet avec le retour du général De Gaulle en France, était équipé et a mis ses ordinateurs à disposition de nombreux scientifiques.

¹⁵ On n'entamera pas un débat sur le terme ni sur ce qu'il recouvre. Nous préférons aujourd'hui parler de « mise en données » du monde, comme de l'histoire. Sur le big data en histoire, nous renvoyons à la lecture de Patrick Manning : Patrick Manning, « Big Data in History: a World-Historical Archive », 27 septembre 2012, <http://www.dataverse.pitt.edu/announcements/.%5Cdocuments%5CBigData.History.Manning.pdf>.

¹⁶ Voir notamment la notion de « méta-source » évoquée notamment dans : Jean-Philippe Genet, « Histoire, Informatique, Mesure », *Histoire & Mesure* 1 (1986): 7-18, doi:10.3406/hism.1986.904.

¹⁷ François Furet et Adeline Daumard, « Méthodes de l'Histoire sociale: les Archives notariales et la Mécanographie », *Annales ESC* 14, no 4 (1959): 693.

¹⁸ Ibid.

¹⁹ Paul Garelli et Jean-Claude Gardin, « Étude par ordinateurs des établissements assyriens en Cappadoce », *Annales ESC* 16, no 5 (1961): 837-876.

²⁰ Ibid., 837.

« [...] what we really need is a little pact with the devil : we know how to read texts, now let's learn how not do read them. »²¹

C'est en ces termes que Franco Moretti, professeur d'histoire de la littérature, décrit la notion de lecture distante ou distanciée. Moretti constate, notamment dans son fameux *Graphs, maps and trees*²² que l'histoire de la littérature du XIXe siècle se focalise pour l'essentiel sur un segment relativement faible de l'histoire de la littérature – les romans les plus connus des « grands auteurs ». Mais ces derniers ne forment qu'une petite partie de la littérature alors produite. Pour réintégrer dans l'histoire de la littérature la plus grande partie de la production littéraire du XIXe siècle, Moretti propose cette notion de lecture distanciée, afin d'utiliser, en histoire de la littérature, les techniques de graphes (*Graphs* – l'histoire économique de l'école des Annales des années 1960), de cartographier cette littérature (*Maps* – s'inspirer de la géographie) et d'en déduire une sorte de théorie de l'évolution de la littérature (*Trees* – de la théorie de l'évolution).

En quoi cette lecture distante est-elle applicable à l'histoire ? Je vais me permettre de prendre un exemple tiré de mon expérience²³.

Tous les historiens développent une relation spécifique aux sources primaires. Au début des années 2000, pendant ma thèse, certains documents trouvés dans les archives m'ont plus marqués que d'autres, à l'exemple des comptes-rendus des réunions entre dignitaires nazis qui ont suivi le grand pogrom de novembre 1938 en Allemagne²⁴. Des débats durant des heures pour définir ce qu'est un « banc public », ces bancs dont l'usage devait être interdit aux Juifs. Il est difficile de comprendre le régime nazi sans avoir eu une relation très proche avec ce type de documents, qui, souvent, donnent la nausée.

Mais cette proximité avec le document peut aussi être un handicap. Toujours pendant ma thèse, j'ai eu un regret. J'ai cherché à prouver que, contrairement aux propos des conservateurs alliés aux nazis, contrairement à une grande partie de l'historiographie²⁵, l'antisémitisme a joué un grand rôle dans l'alliance entre conservateurs, nazis et militaires qui a permis la naissance du IIIe Reich en janvier 1933.

J'ai donc utilisé dans ma thèse un discours, tenu en 1933 par Hjalmar Schacht, qui n'a pas été commandé ni visé par le ministère de la propagande²⁶. Prononcé devant une association féminine où sa femme est impliquée, ce discours trace une lignée directe entre Martin Luther, fondateur de la Réformation protestante au XVIe siècle et grande figure allemande, et (le catholique et autrichien) Adolf Hitler. Cette lignée est tracée, non par le biais de l'économie ou des finances (ou plutôt l'économie y est un prétexte), mais via les deux ouvrages antisémites de Luther²⁷.

Si, à l'époque, j'avais connu les techniques de l'informatique appliquée à la linguistique, j'aurais pu élargir mon argumentaire à un plus large corpus, presque intégral d'ailleurs, des discours de Schacht et comparer le champ sémantique utilisé dans les discours contrôlés par les services de Goebbels et ceux qui ne l'étaient pas, pour voir, plus précisément, dans quelle mesure (et si) l'antisémitisme a été l'un des éléments d'un socle idéologique commun entre

²¹ Franco Moretti, *Distant Reading* (London: Verso, 2013). Location 794 de l'édition Kindle.

²² Moretti, *Graphs, Maps, Trees*.

²³ Cette partie est fortement inspirée d'un billet de mon site web : Frédéric Clavert, « Lecture des sources historiennes à l'ère numérique », Frédéric Clavert, 14 novembre 2012, <http://www.clavert.net/?p=1061>.

²⁴ La Reichskristallnacht / La Nuit de Cristal.

²⁵ Vous pouvez consulter l'échange tendu entre Hans Mommsen et Peter Hoffmann en 2006 sur l'édition par Hans Mommsen et Sabine Gillmann des papiers de Goerdeler: la critique de l'édition par Hoffmann du 25 mai 2006 [<http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-German&month=0605&week=d&msg=sil6fzghulHGx238vWd4ZQ&user=&pw=>], la réponse de Mommsen du 19 juin 2006 [<http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-German&month=0606&week=c&msg=oJJJDhrs/b%2b9SvUzb8eNSg&user=&pw=>] et la réponse de Hoffmann le 5 juillet 2006 [<http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=H-German&month=0607&week=a&msg=6Fv3qRGX6inb0IbYrtihLw&user=&pw=>]. Le débat tourne notamment autour du fait que Mommsen voit Goerdeler comme un antisémite par dissimulation – étant favorable à une séparation des Juifs du corps national pour faire un résumé rapide – tandis que Peter Hoffmann insiste surtout sur le fait que c'était l'« air du temps » et qu'il est difficile de juger quelqu'un sans prendre en compte cet air du temps. Gillmann, Sabine, et Hans Mommsen. *Politische Schriften und Briefe Carl Friedrich Goerdelers*. KG Saur, 2003. Print.

²⁶ BArch Deutsche Reichsbank R2501/6985, s. 176-197 – « Luther Als Deutscher », Rede vom Dr. Schacht

²⁷ Notamment Des Juifs et de leurs mensonges

conservateurs et tout ou partie des nazis et pour mieux comprendre l'évolution du discours antisémite des conservateurs au fur et à mesure qu'évolue le régime nazi, au fur et à mesure, également, de leur éviction du pouvoir.

Aujourd'hui, il me semble que nous pouvons aujourd'hui compter sur deux échelles de lecture des sources historiques: d'une part une échelle *close reading / distant reading* et d'autre part une échelle lecture humaine / lecture computationnelle. La clé de la lecture et de l'interprétation des sources de l'histoire à l'ère numérique est dans les allers-retours constants entre *close reading* et *distant reading* et entre appréhension humaine et appréhension computationnelle de nos sources primaires.

Un premier axe de lecture: *close reading / distant reading*

Si je reprends les deux exemples d'expérience des archives décrits ci-dessus, le premier cas implique une forte proximité avec les archives, mais le second cas montre qu'il est parfois nécessaire d'avoir une relation plus distante aux sources primaires. Cette distanciation permet à l'historien non seulement de voir apparaître des éléments qu'une lecture trop proche ne lui aurait pas permis de voir, mais également de resituer le document dans un ensemble plus large et interconnectés de documents. En effet, une archive seule, en tout cas en histoire contemporaine, n'est pas toujours très significative. Nous essayons autant d'interpréter un document qu'un ensemble de documents. Nous devons également prêter attention à la manière dont les archivistes ont classé les sources primaires – nous parlons ici particulièrement des archives administratives publiques. Pourquoi trouvons-nous tel brouillon de discours dans telle série d'archives plutôt que dans telle autre? Cette question est parfois aussi importante que le contenu d'un document – ou que l'absence d'un document.

En outre, il faut souvent savoir prendre de la distance par rapport au document et considérer une période toute entière (et ses nombreux documents) pour comprendre un processus. Ainsi, entre juin 1941 – le déclenchement de l'opération Barbarossa – et la conférence de Wannsee en janvier 1942 – où la Solution finale est explicitement exposée à un cercle plus large de dignitaires nazis –, dispose-t-on de peu de documents sur cette phase particulièrement importante du déclenchement de la Shoah²⁸. Seul le fait de considérer l'ensemble des documents permet à l'historien de comprendre le processus qui a mené au génocide. Ici, il faut qu'il y ait une certaine distanciation vis-à-vis des documents considérés individuellement pour les comprendre globalement.

Mais cette lecture distanciée des sources primaires, non sans certains risques, peut aller plus loin par une lecture assistée par ordinateur. C'est cette dernière qui permet de replacer des documents dans un ensemble nettement plus vaste afin de faire ressortir des éléments qui ne sont pas visibles par une lecture humaine d'un corpus trop restreint de sources primaires. C'est l'intérêt du second axe de la lecture des sources primaires à l'ère numérique, second axe qui potentialise le premier. Je n'entends pas opposer une lecture humaine à une lecture computationnelle. L'axe lecture humaine / lecture computationnelle mesure le degré, finalement, de médiation computationnelle existant entre l'historien et ses sources primaires. Il ne faut cependant pas sous-estimer l'humanité de la médiation computationnelle, elle-même conçue par des informaticiens, souvent avec l'aide des chercheurs en sciences humaines et sociales.

Un second axe de lecture: lecture humaine / lecture computationnelle

Avec l'émergence des nouvelles technologies, nous pouvons aujourd'hui nous reposer sur des techniques de plus en plus éprouvées permettant de faire ressortir des documents des éléments que nous n'avions pas vu par une lecture classique, humaine, ou de confirmer ce que la lecture humaine avait relevé. Ces techniques peuvent s'appliquer à des documents individuels²⁹ ou à des corpus de documents³⁰, dont l'étendue peut varier de quelques centaines à bien plus³¹.

²⁸ Voir l'intense débat entre Édouard Husson et Florent Brayard, discussion qui a mené à des considérations intéressantes sur le document historique: Florent Brayard, « Shoah : l'intuition et la preuve. Retour sur le processus décisionnel. », La vie des idées, 12 février 2009, <http://www.laviedesidees.fr/Shoah-l-intuition-et-la-preuve.html>. et Édouard Husson, « Comprendre les origines de la Shoah. Réponse à Florent Brayard. », La vie des Idées, 26 février 2009, <http://www.laviedesidees.fr/Comprendre-les-origines-de-la.html>.

²⁹ Frédéric Clavert, « Visualiser, C'est Questionner Les Sources », Frédéric Clavert, 2 octobre 2012, <http://www.clavert.net/visualiser-cest-questionner-les-sources/>.

³⁰ Stéfan Sinclair et Geoffrey Rockwell, « Introduction to Distant Reading Techniques with Voyant Tools, Multilingual Edition », in Digital Humanities 2012, 2012, <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/introduction-to-distant-reading-techniques-with-voyant-tools-multilingual-edition/>.

³¹ Parmi les « corpus » qui peuvent comprendre une masse très importante de documents pourront à terme – ou peut-être même déjà maintenant – les sites web. Ainsi, le site de l'organisation Internet Archive (<http://www.archive.org/>) promet l'accès à 80 terabytes de données pour la recherche: Internet Archive, « 80 Terabytes of Archived Web Crawl Data Available for Research », Internet Archive Blogs, 26 octobre 2012,

Prenons un exemple: Lauren F. Klein, à la conférence *Digital Humanities 2012* à Hambourg, a montré comment elle pouvait avancer sur l'histoire de l'esclavage aux États-Unis par une analyse de réseau et des techniques de visualisation³². La problématique est la suivante: Thomas Jefferson, comme beaucoup d'autres bourgeois états-uniens de son époque, n'évoque ses esclaves que par leurs prénoms. Grâce aux méthodologies numériques et en utilisant l'édition numérique des papiers Jefferson, elle a pu compenser cette absence. Ici, une lecture distante des archives via une appréhension computationnelle de ces documents alternant avec une lecture humaine (car elle a dû vérifier un grand nombre de données) permet de parler des esclaves malgré l'absence de leur nom de famille.

Cette articulation entre lecture humaine et lecture computationnelle n'est pas nouvelle. Elle a déjà existé dans le domaine de l'histoire économique quantitative. Si l'on revient au premier chapitre du livre de Franco Moretti qui explique comment implémenter la manière dont les Annales exploitaient des données quantitatives pour l'écriture de l'histoire économique dans les années 1960. Cette approche quantitative a été critiquée dès les années 1970 – soit bien avant Moretti. Paul Veyne insiste sur le particularisme et s'oppose aux simplismes des concepts. Michel de Certeau dénonce une forme d'illusion réaliste de la prétention des historiens à faire de l'histoire. Le mouvement italien de la micro-histoire, partant de l'individu pour étudier l'ensemble de la société qui l'entoure, a aussi contesté cette approche quantitative fondée sur les grandes séries statistiques de l'histoire. Certains hérauts de l'histoire quantitative eux-mêmes se tournent vers une histoire plus locale, comme Emmanuel Leroy-Ladurie lui-même.

Toutefois, l'ère numérique change à nouveau la donne par le phénomène de mise en données évoqué plus haut. Cette mise en données remet au goût du jour les méthodologies quantitatives, mais pas uniquement. Car les données, si elles s'expriment de manière « numériques » puisqu'elles sont une suite de 0 et 1, ne représentent pas que du quantitatif. Ainsi, apparaissent aujourd'hui des méthodes pour faire ressortir des éléments qualitatifs d'un corpus numérique regroupant une très grande masse de documents / de données.

La thèse ici défendue est qu'il faut articuler une lecture humaine des sources primaires – en clair, le séjour classique en centre d'archives et la critique interne et externe des documents que l'on y trouve – avec une lecture computationnelle. Cette dernière est une lecture assistée par ordinateur, permettant, par le biais d'un ordinateur, l'appréhension – y compris l'interprétation – de masses d'archives (numérisées ou nées numériques) non exploitables par simple lecture humaine.

*

Comment entretenir une relation humaine aux archives tout en étant capable de la dépasser? Comment exploiter des données numérisées, impersonnelles d'une certaine manière, sans tomber dans les biais dans lesquels l'histoire quantitative est parfois tombée?

Le seul moyen est d'être capable de pratiquer une double lecture, en parallèle, des archives, du moins lorsqu'elles sont disponibles en version numérique. Garder cette relation de proximité avec les sources avec une lecture humaine, proche, qui nous permet de garder le contact avec des acteurs et leur environnement. Savoir à la fois améliorer la lecture proche et prendre de la distance au moyen d'une médiation computationnelle. Prendre de la distance, c'est pouvoir aussi comprendre les interactions entre les différentes sources, se dégager de la source unique, « sensationnelle », pour voir le global.

Conclusion

Le numérique ne modifie pas uniquement la manière de lire les sources primaires, mais également l'écriture de l'histoire. L'un des éléments fondateurs du web (et pas uniquement du web) est l'hypertexte, conceptualisé par Ted Nelson dans les années 1960. En insérant des liens sur des mots, expressions, images ou tout autre élément d'une page web, l'hypertexte est multidimensionnel, permettant une infinité de lectures. La narration de l'histoire et le processus de sa fabrication sont ainsi largement remis en cause.

Cette évolution des lectures et des écritures de l'histoire redéfinit la position des chercheurs dans la chaîne globale de la production de l'histoire. Leur collaboration avec l'archiviste, le bibliothécaire, l'éditeur évoluent et s'instaurent de nouvelles relations avec les informaticiens. Ainsi, le problème n'est pas tant celui de l'historien-programmeur, mais le rôle de l'historien dans cette chaîne de production de l'histoire et sa capacité à comprendre comment fonctionnent les différents éléments de cette chaîne. L'historien doit désormais acquérir de nouvelles compétences techniques, de design

<http://blog.archive.org/2012/10/26/80-terabytes-of-archived-web-crawl-data-available-for-research/>.

³² Lauren Frederica Klein, « Social Network Analysis and Visualization in 'The Papers of Thomas Jefferson' », in *Digital Humanities 2012*, 2012, <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/social-network-analysis-and-visualization-in-the-papers-of-thomas-jefferson/>.

(graphique comme de l'information) et de chef de projet. Membre d'une équipe interdisciplinaire, il doit donner cohérence à cet ensemble en se faisant penseur de la globalité de la production de l'histoire et non plus seul producteur de l'histoire.

Ceci implique un renouvellement de l'enseignement du numérique en histoire, qui ne fera pas l'économie d'une interrogation des « littératies » de l'historien et particulièrement de son alphabétisation numérique.