

Reconstruire ce qui manque – ou le déconstruire ? Approches numériques des sources historiques [en]

Anne Baillot <anne_dot_baillot_at_gmail_dot_com>, Le Mans Université

Abstract

La réflexion présentée dans cet article s'articule autour de deux axes : d'une part, rassembler des outils théoriques pour aborder la question de la construction et de l'interprétation de sources historiques et, d'autre part, d'en éprouver la validité à partir d'une expérience de recherche spécifique numérique. La première partie est ainsi consacrée à la conceptualisation des notions de trace et d'archive, avant d'en venir en deuxième partie à la mise en pratique de telles réflexions dans le domaine numérique en partant de l'édition numérique *Lettres et textes — Le Berlin intellectuel autour de 1800* et de la plateforme associée qui contient le catalogue des manuscrits d'August Boeckh. La conclusion revient pour finir sur la place de la réflexion sur l'architecture des données dans la perspective théorique ouverte par les concepts de trace et d'archive présentés dans la première partie.

En utilisant des méthodes transversales aux différentes disciplines des lettres et sciences humaines, les humanités numériques contraignent celles-ci à une redéfinition de leur champ. Les disciplines ayant une dimension historique (histoire politique, histoire économique, histoire des idées, histoire de la littérature, histoire de la philosophie, etc.) n'échappent pas à cette confrontation. L'historicité de leurs sources pose à l'ensemble de ces approches disciplinaires des questions méthodologiques similaires.

1

Si certains corpus historiques^[1] semblent présenter une adéquation inédite avec le potentiel à tirer d'un traitement numérique (comme c'est le cas pour certaines bases de données), il existe inversement des domaines dans lesquels l'inadéquation entre la méthode numérique et le corpus ou la question à traiter fait émerger de manière particulièrement éclatante des lacunes dans les corpus historiques, que ce soit au niveau de leur constitution, de leur transmission ou dans leur nature même, dont l'objectivité, considérée à travers le spectre de méthodes nouvelles, nécessite soudain une redéfinition. Les méthodes numériques contraignent à se poser à nouveaux frais la question : qu'est-ce que construire et interpréter une source historique ? Comme le montre Jean-Philippe Genet, la critique des sources historiques telle qu'elle s'est élaborée au cours du xx^e siècle nous donne les moyens de reprendre de manière fructueuse les principes scientifiques énoncés à la fin du xix^e siècle pour les rendre opérationnels dans le cadre du travail numérique sur les sources scientifiques^[2]. Le but de la réflexion qui suit est, d'une part, d'explorer des concepts permettant d'élaborer une réponse à cette question et, d'autre part, d'en éprouver la validité à partir d'un exemple issu d'un contexte numérique. La première partie portera sur la conceptualisation des notions de trace et d'archive. La deuxième partie sera consacrée à la mise en perspective de telles réflexions dans le domaine numérique en partant de l'édition numérique *Lettres et textes — Le Berlin intellectuel autour de 1800* [Baillot depuis 2012] et du répertoire de manuscrits qui lui est associé, qui contient le catalogue des manuscrits de l'helléniste August Boeckh [Baillot depuis 2013]. En conclusion, on reviendra sur la place de la réflexion sur l'architecture des données dans l'espace de publication numérique, de manière à remettre en perspective l'usage des concepts de trace et d'archive présentés dans la première partie.

2

Conserver et détruire : les sources comme traces

Les approches numériques interrogent la notion de source et la mettent dans une perspective en partie nouvelle. Cela tient au fait qu'elles sont les héritières de traditions scientifiques diverses. Les notions de « document^[3] » et d'« archive » sont chargées d'histoires disciplinaires et de significations différentes en sciences humaines et en sciences de l'information, ce qui n'est pas sans créer des confusions. Si l'on ne parle pas de la même chose lorsqu'on parle d'archive par exemple, comment opérationnaliser ce concept à l'interface de ces deux types de recherches^[4]? Or, il n'est pas simple de faire l'effort de conceptualiser systématiquement et en profondeur l'archéologie du corpus sur lequel on se penche, *a fortiori* dans le cas de la recherche sur projets, limitée dans le temps. Mais dans le monde des humanités numériques, et précisément du fait de l'interdisciplinarité qui lui est inhérente, il est impossible de ne pas commencer par ces questions, faute de passer à côté d'éléments primordiaux pour la recherche qu'il s'agit de mener. Idéalement, deux aspects doivent être explicités pour qu'une recherche sur des sources historiques soit réalisable : la définition du corpus-source et le type de modélisation qu'il s'agit d'y appliquer, qui est la grille de lecture numérique de l'interprétation. Il s'agira ici dans un premier temps, de se demander à la fois ce qui constitue une source historique et comment le chercheur se positionne dans ce travail de constitution de la source sur laquelle il s'appuie. Le point de départ choisi est *Temps et Récit*, où Paul Ricœur part de la tradition documentaire pour faire émerger la place de la trace dans l'historiographie :

L'histoire en tant que recherche s'arrête au document comme chose donnée, même lorsqu'elle élève au rang de document des traces du passé qui n'étaient pas destinées à étayer un récit historique. L'invention documentaire est donc encore une question d'épistémologie. Ce qui ne l'est plus, c'est la question de savoir ce que signifie la visée par laquelle, en inventant des documents [...], l'histoire a conscience de se rapporter à des événements « réellement » arrivés. C'est dans cette conscience que le document devient trace, c'est-à-dire [...] à la fois un reste et un signe de ce qui fut et n'est plus. [Ricœur 1985, 13]

Deux aspects méritent ici d'être soulignés : d'une part, la relation entre document et réalité (qualité d'invention de l'histoire comme un discours à la fois forgé et se voulant porteur d'une vérité descriptive de faits objectivables), d'autre part, la posture épistémologique qui en découle et qui suppose de poser la source comme trace^[5]. Ricœur argumente que ce dernier travail est moins le fait d'historiens praticiens, qui utilisent la trace davantage qu'ils ne la pensent, que de philosophes ou de « moralistes » [Ricœur 1985, 226]^[6] comme Levinas : « [...] l'historien, en tant que tel, ne sait pas ce qu'il fait en constituant des signes en traces. Il reste, à leur égard, dans un rapport d'usage » [Ricœur 1985, 227]. Il ne s'agit pas ici de s'engager dans une polémique disciplinaire, mais plutôt de se demander comment intégrer méthodologiquement la notion de trace dans le travail sur les sources historiques. De fait, le rapport de l'historien à la trace n'est pas nécessairement seulement un rapport d'utilisation. On peut dire que la trace constitue sa source originaire, même lorsque sa source primaire n'est pas la source originaire à proprement parler, mais une mise en œuvre de celle-ci – comme lorsqu'un historien de la littérature ne travaille pas sur le manuscrit, mais sur l'*editio princeps* d'un ouvrage, par exemple, ou qu'un archéologue étudie un artefact en dehors de l'espace des fouilles d'où il a été exhumé. L'intérêt, si ce n'est croissant, du moins marqué, pour la matérialité des objets de la recherche en sciences humaines a contribué à faire émerger des questionnements sur la constitution même de la source et son rapport à son espace de préservation. Pour Ricœur, la pensée levinasienne de la trace a ceci d'essentiel qu'elle met en évidence le fait que la trace, par définition, dérange un ordre [Ricœur 1985, 226]. Or, en dérangeant un ordre, elle en constitue un autre, ou bien elle est constituée en un tel ordre, alors même que son évanescence est sa qualité ontologique première. La trace « est fragile » [Ricœur 1985, 218–219] : cette assertion est vraie à différents niveaux. Tout d'abord, au sens où elle présente ici Ricœur, parce qu'à tout instant elle est matériellement susceptible de disparaître. Mais aussi parce qu'elle n'est jamais la chose en soi, mais seulement un signe, un symbole de quelque chose qui a été, dans un contexte donné, et est désormais dans un autre contexte ; la trace est « vestige d'un passage » [Ricœur 1985, 219]. La temporalité donne à la trace sa friabilité. Enfin, et c'est ce qu'il va s'agir d'approfondir, la trace est autant signe d'une absence que signe d'une présence^[7]. La trace est porteuse, *in nuce*, de tout ce dont elle n'est pas la trace, tout ce qui a disparu. Qui plus est, tout mécanisme d'archivage de ces traces est lui aussi inéluctablement confronté à une discrimination, une sélection, un choix « (quoi conserver ? quoi détruire ?) » [Ricœur 1985, 212]. Derrida, dans *De la grammatologie*, va plus loin dans la postulation d'une évanescence ontologiquement première de la trace :

La trace n'est pas seulement la disparition de l'origine, elle veut dire ici [...] que l'origine n'a même pas disparu, qu'elle n'a jamais été constituée qu'en retour par une non-origine, la trace, qui devient ainsi origine de l'origine. Dès lors, pour arracher le concept de trace au schéma classique qui la ferait dériver d'une présence ou d'une non-trace originaire et qui en ferait une marque empirique, il faut bien parler de trace originaire ou d'archi-trace. Et pourtant nous savons que ce concept détruit son nom et que, si tout commence par la trace, il n'y a surtout pas de trace originaire. [Derrida 1967, 59]

Dans cette perspective, il est non seulement impossible de discerner la dimension originaire de la trace, mais même essentiellement inconcevable de faire coïncider la notion de trace et celle d'origine. Ainsi, la trace n'est que construction, signe de quelque chose qui n'a aucun caractère originaire ou premier, et dont la pertinence pour la recherche historique nécessite bien d'être, à chaque instant, déconstruite pour être réévaluée.

5

La trace, vestige du passage autant que du passé, accumulée hors contexte et selon des mécanismes de sélection divers et parfois peu transparents, reste cependant l'objet principal matérialisant les sources telles qu'elles sont utilisées dans les disciplines à dimension historique et leurs recherches. Sa fragilité même se transmet à l'institution qui en garantit la transmission en l'ordonnant : l'archive.

6

Comme institutions ordonnantes, l'archive (l'unité archivistique, le dossier, la boîte) et les archives (l'institution patrimoniale) ont des structures similaires. Il s'agit, pour l'une comme pour les autres et chacune à leur échelle, de rassembler et de consigner les traces (majoritairement écrites) du passé de manière à ce qu'elles soient consultées non pas une, mais plusieurs fois, et non par une, mais par plusieurs personnes. Réactualisées dans des contextes différents, il leur est ainsi possible de déployer une variété de significations et d'exister non plus seulement en elles-mêmes, mais dans une pluralité de contextes. Archiver des traces écrites du passé, c'est donc, en ce sens, rendre possible l'émergence d'une pluralité de significations qui nécessite toujours la reconstruction d'un contexte perdu pour faire sens, reconstruction inévitablement spéculative puisque l'originaire, par définition, n'est plus. La nécessité de reconnaître ce manque est évidemment plus aisément acceptable mentalement lorsque le temps qui nous sépare de la trace est grand – l'archéologie ne peut se passer d'hypothèses de reconstruction. Inversement, plus l'histoire nous est proche, plus nous pouvons être tentés de croire que nous ne reconstruisons pas de contexte, mais que nous en saisissons la réalité d'origine. Il est plus facile de déconstruire ce qui nous est étranger que ce qui nous est proche, plus facile aussi de penser la reconstruction des pans manquants dans ce pour quoi les traces sont lacunaires que dans ce pour quoi on peut considérer ses sources comme une totalité.

7

Dans l'article qu'il consacre aux études classiques dans le *Companion to Digital Humanities*, Gregory Crane revient sur les relations entre méthodes numériques et corpus classique [Crane 2004]. Il insiste sur le fait que des erreurs ont été commises, notamment dans les choix de logiciels qui semblaient, au premier abord, adéquats, avant de s'avérer, une fois en phase d'opérationnalisation, inadaptés. Retraçant une histoire faite de tâtonnements, d'échecs et d'itérations, il insiste sur le fait que les outils numériques n'ont pu prendre une telle ampleur dans les études classiques que parce que la tradition de recherche centenaire, voire millénaire, sur laquelle celles-ci s'appuient, était solidement structurée, entre autres en ce qui concerne le décloisonnement des différents formats d'accès à l'information^[8]. La conscience de la nécessité d'ordonner des informations parcellaires pour pouvoir en tirer du savoir est ainsi essentielle pour la mise en œuvre de méthodes numériques. Les études classiques peuvent du moins se targuer d'y être parvenu à grande échelle. Les sources issues du xx^e siècle sont un miroir aux alouettes dans ces deux sens : non seulement elles nous sont proches en termes de mentalité, mais elles nous sont parvenues en bien plus grande masse documentaire que les sources des siècles antérieurs. Il s'agit là d'une spécificité du xx^e siècle, car avec le xxi^e siècle et l'émergence de sources originaires numériques (*born-digital*), la question se déplace à nouveau. Les éléments d'identification du statut épistémologique des sources originaires numériques se déploient (par opposition aux sources analogues) sur plusieurs dimensions en même temps, prenant en compte les relations avec d'autres sources (qui y sont inscrites), les relations entre données et métadonnées, l'évolution temporelle qui existe sans être visible au premier abord. Nous ne savons toujours pas aujourd'hui comment citer une source numérique de manière fiable, de façon à savoir exactement, en la référant, de quoi on parle.

8

Mais revenons au monde analogue. Là aussi, Derrida propose une vision radicale de la fonction des archives en tant qu'instance de consignation :

9

[...] s'il n'y a pas d'archive sans consignation en quelque lieu extérieur qui assure la possibilité de la mémorisation, de la répétition, de la reproduction ou de la réimpression, alors rappelons-nous aussi que la répétition même, la logique de la répétition, voire la compulsion de répétition reste, selon Freud, indissociable de la pulsion de mort. Donc de la destruction. [...] L'archive travaille toujours et *a priori* contre elle-même. [Derrida 1995, 26–27]

Chaque consultation et interprétation de la trace archivale contribue ainsi à sa destruction ; aucune reprise à l'identique n'est possible. L'assertion est vraie à deux niveaux : tout d'abord, matériellement, la consultation – de même que la conservation sans consultation d'ailleurs – n'empêchent pas le temps de faire son office. La consultation, qui seule peut faire émerger un sens de la trace, accélère le processus de délitement matériel. Si l'on scanne un document ou un artefact qui se dégrade, on ne contribue pas à conserver la source elle-même, mais une représentation de celle-ci, une trace de la trace, pour ainsi dire. Cela revient à dévaluer la trace de départ et à générer l'illusion d'une identité entre la trace et sa représentation.

10

L'archive se délite en même temps qu'elle déploie ses mécanismes de conservation ; il est impossible de dissocier un mouvement de l'autre. Nous savons que nous ne disposons pas de toutes les traces nécessaires pour reconstruire la totalité du passé, mais, comme le demande Derrida : « Comment peut-[on] prétendre faire la *preuve* d'une absence d'archive ? » [Derrida 1995, 103]. Dans cette perspective, il est seulement possible d'admettre la co-présence d'un gouffre d'absence à côté des traces que nous percevons, comme la partie immergée d'un iceberg dont il ne serait de toute façon possible que de reconstruire une silhouette imprécise, qui ne serait elle-même que le dessin approximatif d'une réalité impossible à atteindre. Plus on chercherait à s'approcher de la trace, à embrasser l'ensemble d'une archive, à reconstituer un objet à partir de l'archive, plus on s'entourerait de chimères, moins il serait possible de construire un discours.

11

C'est à une tout autre image qu'a recours Roland Barthes dans *Le Degré zéro de l'écriture*, et une image sans doute plus encourageante pour poursuivre la recherche. Dans cet essai, c'est le texte qui est au cœur du propos, ce qui n'est pas forcément le cas pour toute recherche historique. Cependant, l'image et l'argument restent transposables à des artefacts ou autres sources pertinentes pour l'historiographie, lorsque Barthes écrit : « Toute trace écrite se précipite comme un élément chimique d'abord transparent, innocent et neutre, dans lequel la simple durée fait peu à peu apparaître tout un passé en suspension, toute une cryptographie de plus en plus dense. » [Barthes 1972, 20]. L'image de l'encre magique est tout à l'opposé de celle de l'archive inéluctablement prise dans l'autodestruction. L'encre magique fait apparaître de nouvelles strates de sens comprises comme un enrichissement et non comme un appauvrissement, inclut l'histoire et donc la temporalité dans ces strates de sens, sans essayer de les dissocier l'une de l'autre.

12

Que l'on considère l'approche des sources historiques comme un processus de destruction ou comme un processus d'enrichissement, il semble dans tous les cas difficile, voire impossible, d'échapper à un travail de reconstruction – chirurgie plastique destinée à rendre la trace interprétable –, ou à un travail de déconstruction, sans lequel le chercheur n'est pas à même d'articuler le point de vue à partir duquel il opère la reconstruction. Si l'on n'échappe ni à la reconstruction, ni à la déconstruction, alors la tâche essentielle du chercheur travaillant sur des sources historiques est bien, inévitablement, d'ordre épistémologique ; elle consiste à auto-réfléchir ses propres mécanismes et les nommer, car il s'agit là, *in fine*, du seul moyen de rendre possible le travail scientifique sur des sources qui sont par nature fragmentaires et prises dans la temporalité. Ce travail épistémologique qui, comme le met d'ailleurs en évidence Ricœur lorsqu'il distingue historien-savant et historien-philosophe [Ricœur 1985, 227], peine parfois à trouver sa place dans l'historiographie traditionnelle, où l'ancrer dans le travail numérique sur les sources historiques ? Quel est le lieu de cet auto-positionnement à partir duquel s'ouvrir à d'autres interprétations, quand il ne s'agit pas de le présenter très abstraitement comme cela vient d'être exposé, de manière pour l'essentiel spéculative, mais de l'associer à une démarche singulière d'exploration de sources et de recherche ?

13

Construire, déconstruire, reconstruire ? L'exemple de l'édition *Lettres et textes. Le Berlin intellectuel autour de 1800*

Pour aborder ces questions, cette deuxième partie partira de l'expérience de constitution de corpus dans le contexte de l'édition numérique *Lettres et textes — Le Berlin intellectuel autour de 1800* [Baillot depuis 2012]. Cette édition numérique rassemble divers types de textes signés par des auteurs différents, réunis autour d'une question de recherche, à savoir les stratégies de publication développées par les intellectuels berlinois autour de 1800 (voir [Baillot et Busch 2015]). Faire émerger des tendances au sein de telles stratégies prises dans leur ensemble suppose d'avoir des points de comparaison, de manière, notamment, à ne pas surinterpréter des phénomènes présents chez certains auteurs qui sont peut-être tout aussi bien présents chez d'autres. Les corpus édités s'étendent donc de 1760 jusqu'à la révolution de 1848^[9]. Comment le choix des sources s'est-il opéré ? Les lettres éclairent les phénomènes de genèse et de réception sous un angle que l'équipe éditoriale a décidé de recouper avec les phénomènes textuels observables dans les brouillons d'œuvre. La question des stratégies de publication chez les intellectuels suppose de prendre en compte toute une série d'aspects, dont quatre particulièrement significatifs ont été retenus : la dimension politique de la parole savante ; les différences de stratégies développées par les hommes et par les femmes dans ce contexte ; le rôle de catalyseur de discours intellectuel joué par la fondation de l'Université [Baillot 2014a] et l'inscription dans une tradition héritée des Lumières francophones [Baillot 2014b].

14

Il va de soi qu'il aurait été souhaitable de travailler avec une plus grande quantité de textes pour pouvoir faire mieux émerger les structures en question, le but étant d'aller au-delà de l'étude de cas pour pouvoir tirer des conclusions générales. Mais ce qui fait la faiblesse de cette édition fait aussi sa force : si nous n'avons pas édité davantage de textes, c'est que nous avons annoté en profondeur. L'annotation est conçue pour permettre de mettre en évidence les phénomènes de collaboration et d'intertextualité : d'une part, les entités nommées sont informées en détail (personnes, lieux, œuvres, institutions, liens avec les fichiers d'autorité autant que cela est possible) ; d'autre part, un appareil d'annotation génétique (très allégé par rapport au module *TEI Critical Apparatus* [http1]^[10]) permet de repérer les phénomènes de réécriture et de correction^[11]. La combinaison des deux niveaux d'encodage est destinée à faire émerger les types de censure observables, ainsi que de les attribuer aux instances correspondantes. Plutôt que de postuler le fait que les éditeurs adaptaient les textes (soit pour qu'ils soient moralement acceptables, soit pour qu'ils se vendent mieux) aux XVIII^e et XIX^e siècles, il est maintenant possible de montrer qui censurait quoi, comment et quand, de retracer des évolutions, de faire émerger des tendances, de suivre des circulations d'ouvrages et de matérialiser l'intertextualité. Cela n'est possible qu'à une échelle qui, en l'état, n'est pas statistiquement représentative, et qui limite la portée des conclusions générales qu'il est possible d'en tirer. Aussi faudrait-il pouvoir soit éditer au moins encore autant de textes qui ce qui est déjà en ligne, pour commencer à obtenir une masse critique acceptable, soit y agréger d'autres corpus déjà édités. L'objectif est de procéder à ces deux types d'ajouts au cours des années à venir en adaptant l'interface à mesure.

15

Pour revenir sur la dimension de construction numérique d'une telle masse documentaire, soulignons les dimensions qu'elle fait ressortir au regard des réflexions faites en première partie. Tout d'abord, il s'agit d'un corpus construit de toutes pièces à partir d'une question de recherche et non à partir d'une tradition ou d'une autorité. Il n'est ni centré sur un auteur, ni sur une tradition historique (un tout dont l'unité proviendrait de la conservation), ni sur une construction idéologique propre à l'époque (comme l'aurait été le fait d'axer le propos ne serait-ce que sur le romantisme, notion qui apparaît très peu de temps après dans le discours d'époque, contrairement à l'usage de la notion d'intellectuels autour de laquelle s'articule l'édition et la recherche afférente^[12]). De plus, les sources choisies proviennent de différentes archives et bibliothèques, et constituent, ainsi mises ensemble, une archive virtuelle singulière. Il s'agit d'un corpus ouvert, qui peut toujours être enrichi et ne sera *a priori* jamais achevé. L'annotation, elle aussi, est ouverte, et elle est susceptible d'être approfondie ; l'ensemble constitue en quelque sorte un échantillon, une heuristique destinée à continuer d'être développée. Cette dimension très expérimentale, problématique pour une partie importante de la communauté scientifique, est pour ainsi dire compensée par le fait que cette mosaïque de textes fabriquée à partir de la question de recherche contient quelques perles, à savoir des manuscrits jusqu'ici inédits et de grande valeur du point de vue du canon historique ou littéraire, et sur lesquels ce travail éditorial a été pionnier.

16

Le premier aspect à interroger pour comprendre ce travail de construction d'une métasource, au sens proposé par Jean-Pierre Genet^[13], concerne les métadonnées. Pourquoi, contrairement à sa réception parfois fraîche par la communauté scientifique, les archives et les bibliothèques reconnaissent-elles la valeur de cette ressource ? Pourquoi cela leur importe-t-il peu que cette édition ne prétende pas faire le tour complet d'un corpus ou d'un sujet ? Parce que les archives et les bibliothèques savent en général que rien ne sera jamais complet et que la seule façon de travailler dans l'univers de sources parcellaires consiste à mettre, si ce n'est en commun, du moins en relation, ce dont dispose chacun. Le plus important, ce seront donc les métadonnées, car ce sont elles qui permettront la connexion avec d'autres ressources, mais aussi parce que, d'un point de vue bibliothécaire ou archivistique, l'architecture des métadonnées renseigne sur l'approche adoptée. Le format, les types de champs, le niveau de détail avec lequel ces champs sont renseignés, bref, la structure des métadonnées représente la carte de visite de la perspective de recherche dans laquelle elles ont été conçues et implémentées. Elles documentent l'architecture, le mode d'emploi de la construction du corpus (compris ici comme métasource). Cela ne signifie pas pour autant que la recherche devrait complètement basculer vers un fonctionnement bibliothécaire ou archivistique. Tout d'abord, parce qu'il règne là aussi un joyeux chaos en termes de standards, qui ne garantit pas vraiment l'interopérabilité souhaitable. Ensuite, parce que, bien que certains formats permettent fort bien de composer avec les informations manquantes, par exemple en intégrant des substituts (*surrogates*) à l'endroit *ad hoc* (que ce soit au niveau d'un champ, d'un document ou d'un ensemble de documents), la communication entre les ressources ne permet pas d'échapper à des inexactitudes ou, par exemple, à des saisies concurrentes et redondantes. Le cas des doublons dans les fichiers d'autorité de personnes en témoigne. En effet, lorsque deux autorités (deux bibliothèques) ont chacune procédé à une saisie biographique portant sur la même personne avec des données très légèrement différentes, comment procéder à la correction ? Quelle sera l'autorité correctrice ? La régulation est inscrite dans la structure elle-même puisque chaque saisie est accompagnée d'une indication concernant le degré de fiabilité de l'information saisie ; ainsi, les données recoupées plusieurs fois ont par définition un statut d'autorité supérieur à celles qui ne le sont qu'une fois. Il n'en reste pas moins que l'élimination des doublons biographiques reste problématique pour la recherche et son travail sur des sources historiques.

17

Les prémisses de saisie sont également problématiques du point de vue de la recherche^[14]. Par définition, les notices d'autorité sont gérées par les bibliothèques et saisissent en priorité des données pertinentes d'un point de vue bibliothécaire. Pour les personnes historiques, cela signifie que ce sont principalement les personnes ayant publié des livres qui seront mises à l'honneur. Mais même au sein de cette catégorie de personne privilégiée dans les fichiers d'autorité de personnes, c'est-à-dire les écrivains, la saisie n'est pas homogène.

18

Les fichiers d'autorité perpétuent en bonne partie le canon et ne permettent pas fondamentalement de correctif historiographique puisque se trouvent en général, ou en tout cas à un niveau statistiquement significatif, exclues de ces ressources les personnes « secondaires », notamment les femmes. Or, il faut attendre le xx^e siècle pour que les femmes publient plus souvent sous leur nom, ou à tout le moins sans être dissimulées derrière un homme ou un nom d'homme. Pour tous les corpus antérieurs à cette période, il est donc nécessaire non pas de se contenter d'un lien de la ressource scientifique vers le fichier d'autorité pour pouvoir agréger les données issues d'autres ressources : dans ce cas, la recherche doit contribuer à structurer le fichier d'autorité, à mettre en lien, par exemple, le prête-nom masculin et l'auteur femme, justifier des sources, et ce, pour faire entrer un auteur de plus (certes de sexe féminin) au registre des traces biographiques des temps passés, en restant donc dans le fond dans la tradition historiographique de la domination existante du texte imprimé. Des acteurs aussi essentiels au monde du livre que le sont les éditeurs jouent en comparaison un rôle secondaire dans ce travail de référencement.

19

Ce cas particulier est intéressant dans la mesure où il montre tout d'abord la domination de la trace textuelle dans notre rapport aux sources historiques. Qui plus est, les ressources numériques étant elles-mêmes essentiellement textuelles, toute cette épistémologie encore à développer ne pourra pas vraiment se passer d'être une épistémologie textuelle. Enfin, on voit bien ici qu'il est moins facile de s'affranchir du canon que ce que l'on voudrait bien prétendre, alors même que voilà des décennies que la recherche est censée s'être affranchie de la domination de l'auteur (mais l'auteur est-il vraiment mort ?).

20

Venons-en, pour finir, à la question la plus complexe. Si tant est qu'on soit en mesure de travailler avec le manque,

21

avec la lacunarité des traces qui font nos sources – ce qui suppose de lier activité de recherche et activité d’archivage de manière très étroite – comment et où définir cette posture singulière ? La ressource numérique ne devrait-elle pas être en mesure de se passer d’une notice explicative formulée *via* un autre média pour être recevable ? Combien d’articles faut-il écrire pour que le principe même d’une édition numérique soit accessible à la communauté scientifique ? On le voit, le travail de modélisation, de sa documentation, de dissémination de cette modélisation et de sa documentation, est au cœur de l’approche numérique des sources historiques. Pas davantage que l’archive analogue, mais peut-être d’une manière rendue plus explicite par l’exigence fondamentale de mise en réseau, l’archive numérique ne peut se passer d’une réflexion sur la posture de celui qui archive la source pour comprendre ce qui est archivé. La tâche de l’historien numérique est à la fois de constituer la source (selon des principes de mise en relation et de mutualisation des ressources propres à Internet), de s’interroger sur la manière dont il a constitué la source, et d’explicitier l’ensemble de ce processus en formulant autant que possible les choix de sélection, les biais – une exigence qui nécessite également de repenser les moyens dont nous disposons pour communiquer l’ensemble de ces informations, entre édition, article, documentation, ouvrage, etc.

Conclusion

La question des sources historiques numériques éclaire ainsi dans un même mouvement la manière de les construire – qui reste, éminemment, humaine^[15] – et la manière de les diffuser.

22

Les formats numériques sont flexibles ; ils ne sont pas encore enfermés dans des traditions qui compartimentent différents niveaux d’écriture et de recherche. C’est à nous de nous en saisir de manière aussi créative que rigoureuse, pour permettre d’y faire émerger et figurer, à côté des ressources elles-mêmes, les fondements autant heuristiques qu’herméneutiques qui président à la constitution de ces ressources. Le changement de média, de l’artefact ou du papier au numérique, rend la prise de conscience du caractère éminemment construit de toute ressource numérique plus aisé ; dans quelques générations, il sera devenu si naturel qu’on l’oubliera. C’est de ce geste de méfiance naturelle, d’étrangeté vis-à-vis de l’outil numérique que nous pouvons et devons nous saisir pour ré-interroger notre posture scientifique. L’archive numérique ne sera jamais qu’incomplète, reflet en cela de son *alter ego* analogue. D’une part, elle est elle-même aux prises avec une « longue durée » encore à définir, prisonnière d’une temporalité qui la détruit si on n’en prend pas soin. D’autre part, l’archive numérique doit accepter sa propre finitude. Il n’y aura sans doute jamais qu’un pays riche en pétrole et pauvre en littérature comme la Norvège qui pourra prétendre numériser la totalité de son patrimoine livresque. La question du rapport à la quantité doit être posée, là aussi, de manière à ce que la recherche puisse s’en saisir, c’est-à-dire en acceptant que le « big » de « big data » soit différencié et mis en œuvre différemment : une annotation très complexe d’un texte, voire même seulement des métadonnées riches, représentent une quantité d’information plus significative que ne le sont des gigaoctets de scans numériquement « invisibles ». Où est le plein, où est le creux ? Voilà ce que nous avons à nommer.

23

Notes

[1] « Corpus historiques » désigne ici des corpus, principalement de textes, ayant une dimension historique, et non les seuls corpus étudiés dans le cadre de recherches strictement historiographiques.

[2] Voir [Genet 1994, 8] : « L’idéal [dix-neuviémiste] de la science historique allemande est la publication critique des sources d’où doit découler, presque nécessairement, une interprétation juste. [...] La critique des écoles historiques du xx^e siècle a peu à peu réintroduit dans le champ de l’histoire des questionnements théoriques [...] ; le travail de l’École des Annales, notamment, est trop important ici pour ne pas être mentionné. En même temps, le statut privilégié de la source historique en soi a été battu en brèche : on a réalisé que tout pouvait devenir source historique, à condition d’être construit comme tel. [...] L’informatique nous permet [...] de reprendre le concept de source et de le réévaluer. »

[3] Sur ce point, voir l’excellent travail de définition dans [Bertrand et Jacobs 2011, 128], ainsi que [Pédauque 2006].

[4] [Bertrand et Jacobs 2011, 129] notamment à la note 11, montrent bien comment tenir cette tension entre l’inspiration théorique et une approche pragmatique de la modélisation.

[5] En ce sens, voir aussi [Ricoeur 1985, 212] : « Que la trace soit un tel *requisit* [de toutes les productions du travail historien] pour la *pratique* historique, il suffit, pour le montrer, de suivre le processus de pensée qui, partant de la notion d'archives, rencontre celle de document (et, parmi les documents, celle de témoignage) et, de là, remonte à sa présupposition épistémologique dernière : la trace précisément. C'est de ce *requisit* que la réflexion sur la *conscience* historique repartira pour son investigation de second degré. »

[6] Cf. aussi [Ricoeur 1985, 227] : « Mais ce que signifie la trace est un problème non d'historien-savant, mais d'historien-philosophe. »

[7] Cf. [Derrida 1967, 68] : « On ne peut penser la trace instituée sans penser la rétention de la différence dans une structure de renvoi où la différence apparaît *comme telle* [...]. L'absence d'un autre ici-maintenant, d'un autre présent transcendantal, d'une *autre* origine du monde apparaissant comme telle, se présentant comme absence irréductible dans la présence de la trace... »

[8] [Crane 2004, 46] : « Classicists have for thousands of years been developing lexica, encyclopedias, commentaries, critical editions, and other elements of scholarly infrastructure that are best suited to an electronic environment. Classicists have placed great emphasis on systematic knowledge management and engineering. » ; [Crane 2004, 48] : « [David Packard] observed that software and systems were ephemeral but that primary sources such as well structured, cleanly entered source texts were objects of enduring value. » Les sources primaires ici évoquées sont, du point de vue de mon argumentation, des représentations de traces.

[9] La relation à Berlin (sujet central de la recherche par les conditions socio-historiques qui y sont réunies à cette période) est représentée à différents niveaux : certains auteurs résident à Berlin, d'autres y ont résidé ou sont en contact avec des Berlinoises.

[10] Cf. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.

[11] La documentation des choix d'encodage de cette édition est téléchargeable depuis [http2].

[12] Concernant la définition du concept d'intellectuel, voir l'introduction à *France-Allemagne. Figures de l'intellectuel* (voir [Baillot 2014b]).

[13] [Genet 1994, 8] : la métasource est « l'ensemble structuré des informations mises en forme et transmises à l'ordinateur et traitées par lui ».

[14] Sur toutes les questions relatives à l'usage des fichiers d'autorité dans les éditions savantes (numériques), voir [Stadler 2012].

[15] Voir ainsi [Genet 2011, 7].

Works Cited

Baillot 2014a Baillot, Anne. "Qu'est-ce qui fait l'intellectuel? Les professeurs de l'université de Berlin et leur patriotisme (1810-1820)". In Anne Baillot et Ayşe Yuva (éd.), *France-Allemagne. Figures de l'intellectuel entre révolution et réaction (1780-1848)*. Presses du Septentrion, Lille. 2014 [<http://halshs.archives-ouvertes.fr/halshs-00879149>; dernière consultation le 10.07.17].

Baillot 2014b Baillot, Anne. "Les corpus français sont-ils allemands? Éditer des textes de la Prusse fédéricienne en Allemagne". In *Philologie im Netz Beiheft 7* (2014): 49-63 [<http://web.fu-berlin.de/phin/beiheft7/b7t03.pdf>; dernière consultation le 10.07.17].

Baillot depuis 2012 Baillot, Anne (éd.). *Lettres et textes — Le Berlin intellectuel autour de 1800* (depuis 2012) [<http://www.berliner-intellektuelle.eu/?fr>; dernière consultation le 10.07.17].

Baillot depuis 2013 Baillot, Anne (éd.). *August Boeckh-Nachlassprojekt* (depuis 2013) [<http://www.berliner-intellektuelle.eu/boeckh>; dernière consultation le 10.07.17].

Baillot et Busch 2015 Baillot Anne et Busch Anna. "“Berliner Intellektuelle um 1800” als Programm. Über Potential und Grenzen digitalen Edierens". *Literturkritik.de*, numéro spécial "Romantik Digital" (septembre 2015) [http://www.literaturkritik.de/e/public/rezension.php?rez_id=19678&ausgabe=201409; dernière consultation le 10.07.17].

Barthes 1972 Barthes, Roland. *Le Degré zéro de l'écriture*. Seuil, Paris (1972).

Bertrand et Jacobs 2011 Bertrand, Paul et Jacobs, Christophe. "“Digital Humanities” et critique historique documentaire: “digital” ou “critical turn”? ". In Jean-Pierre Genet et Andrea Zorzi (éd.), *Les Historiens et l'informatique. Un métier à réinventer*, École française de Rome (2011): 125-139.

Crane 2004 Crane, Gregory. "Classics and the Computer: An End of the History". In Susan Schreibman, Ray Siemens et John Unsworth (éd.), *A Companion to the Digital Humanities*, Oxford (2004): 46-55.

Derrida 1967 Derrida, Jacques. *De la grammatologie*. Éditions de Minuit, Paris (1967).

Derrida 1995 Derrida, Jacques. *Mal d'archive. Une impression freudienne*. Galilée, Paris (1995).

Genet 1994 Genet, Jean-Pierre. "Source, Métasource, Texte, Histoire". In Francesca Bocchi et Paul Denley (éd.), *Storia & Multimedia. Proceedings of the Seventh International Congress of the Association for History & Computing*, Bologna (1994): 3-17.

Genet 2011 Genet, Jean-Pierre. "Introduction". In Jean-Pierre Genet et Andrea Zorzi (éd.), *Les Historiens et l'informatique. Un métier à réinventer*, École française de Rome (2011): 1-9.

Pédauque 2006 Pédauque, Roger T. *Document et Modernités* (2006).

Ricœur 1985 Ricœur, Paul. *Temps et Récit III*. Seuil, Paris (1985).

Stadler 2012 [Stadler 2012] Stadler, Peter. "Normdateien in der Edition". *Editio*, 26 (2012): 174-183.

http1 [<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>; dernière consultation le 10.07.17].

http2 [<http://tei.ibi.hu-berlin.de/berliner-intellektuelle/about?de>; dernière consultation le 10.07.17].



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Rebuilding what is missing – or deconstructing it? Digital approaches to historical sources [fr]

Anne Baillot <anne_dot_baillot_at_gmail_dot_com>, Le Mans Université

Abstract

This paper presents first a theoretical framework on the concepts of trace and archive, and in a second step an operationalization of a digital approach to historical material based on these concepts. The second part is illustrated by the practical experience of developing the digital scholarly edition *Letters and Texts. Intellectual Berlin around 1800*. The conclusion addresses the question of data architecture based on the confrontation of the theoretical framework and the practical experience.

Note on Translation

For articles in languages other than English, DHQ provides an English-language abstract to support searching and discovery, and to enable those not fluent in the article's original language to get a basic understanding of its contents. In many cases, machine translation may be helpful for those seeking more detailed access. While DHQ does not typically have the resources to translate articles in full, we welcome contributions of effort from readers. If you are interested in translating any article into another language, please contact us at editors@digitalhumanities.org and we will be happy to work with you.

1

Works Cited

- Baillot 2014a** Baillot, Anne. "Qu'est-ce qui fait l'intellectuel? Les professeurs de l'université de Berlin et leur patriotisme (1810-1820)". In Anne Baillot et Ayşe Yuva (éd.), *France-Allemagne. Figures de l'intellectuel entre révolution et réaction (1780-1848)*. Presses du Septentrion, Lille. 2014 [<http://halshs.archives-ouvertes.fr/halshs-00879149>; dernière consultation le 10.07.17].
- Baillot 2014b** Baillot, Anne. "Les corpus français sont-ils allemands? Éditer des textes de la Prusse Frédéricienne en Allemagne". In *Philologie im Netz Beiheft 7* (2014): 49-63 [<http://web.fu-berlin.de/phn/beiheft7/b7t03.pdf>; dernière consultation le 10.07.17].
- Baillot depuis 2012** Baillot, Anne (éd.). *Lettres et textes — Le Berlin intellectuel autour de 1800* (depuis 2012) [<http://www.berliner-intellektuelle.eu/?fr>; dernière consultation le 10.07.17].
- Baillot depuis 2013** Baillot, Anne (éd.). *August Boeckh-Nachlassprojekt* (depuis 2013) [<http://www.berliner-intellektuelle.eu/boeckh>; dernière consultation le 10.07.17].
- Baillot et Busch 2015** Baillot Anne et Busch Anna. "'Berliner Intellektuelle um 1800' als Programm. Über Potential und Grenzen digitalen Edierens". *Literaturkritik.de*, numéro spécial "Romantik Digital" (septembre 2015) [http://www.literaturkritik.de/e/public/rezension.php?rez_id=19678&ausgabe=201409; dernière consultation le 10.07.17].
- Barthes 1972** Barthes, Roland. *Le Degré zéro de l'écriture*. Seuil, Paris (1972).
- Bertrand et Jacobs 2011** Bertrand, Paul et Jacobs, Christophe. "'Digital Humanities' et critique historique documentaire: 'digital' ou 'critical turn'?" In Jean-Pierre Genet et Andrea Zorzi (éd.), *Les Historiens et l'informatique. Un métier à réinventer*, École française de Rome (2011): 125-139.
- Crane 2004** Crane, Gregory. "Classics and the Computer: An End of the History". In Susan Schreibman, Ray Siemens et John Unsworth (éd.), *A Companion to the Digital Humanities*, Oxford (2004): 46-55.

Derrida 1967 Derrida, Jacques. *De la grammatologie*. Éditions de Minuit, Paris (1967).

Derrida 1995 Derrida, Jacques. *Mal d'archive. Une impression freudienne*. Galilée, Paris (1995).

Genet 1994 Genet, Jean-Pierre. "Source, Métasource, Texte, Histoire". In Francesca Bocchi et Paul Denley (éd.), *Storia & Multimedia. Proceedings of the Seventh International Congress of the Association for History & Computing*, Bologna (1994): 3-17.

Genet 2011 Genet, Jean-Pierre. "Introduction". In Jean-Pierre Genet et Andrea Zorzi (éd.), *Les Historiens et l'informatique. Un métier à réinventer*, École française de Rome (2011): 1-9.

Pédauque 2006 Pédauque, Roger T. *Document et Modernités* (2006).

Ricœur 1985 Ricœur, Paul. *Temps et Récit III*. Seuil, Paris (1985).

Stadler 2012 [Stadler 2012] Stadler, Peter. "Normdateien in der Edition". *Editio*, 26 (2012): 174-183.

http1 [<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>; dernière consultation le 10.07.17].

http2 [<http://tei.ibi.hu-berlin.de/berliner-intellektuelle/about?de>; dernière consultation le 10.07.17].



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.