



# Reconstruire ce qui manque – ou le déconstruire ? Approches numériques des sources historiques

Anne Baillot

## ► To cite this version:

Anne Baillot. Reconstruire ce qui manque – ou le déconstruire ? Approches numériques des sources historiques. Entre sources, données et réseaux (programme interdisciplinaire MORE-HIST), Mar 2015, Toulouse, France. 2015. <halshs-01133507>

**HAL Id: halshs-01133507**

**<https://halshs.archives-ouvertes.fr/halshs-01133507>**

Submitted on 19 Mar 2015

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

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



Distributed under a Creative Commons Attribution 4.0 International License



## Reconstruire ce qui manque – ou le déconstruire ?

### Approches numériques des sources historiques

Anne Baillot, Toulouse, mars 2015

En utilisant des méthodes transversales aux disciplines, les humanités numériques contraignent celles-ci à une redéfinition de leur champ ; l'histoire n'échappe pas à cette confrontation. Si certains corpus historiques 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 méthode numérique et corpus(/question/outil) fait émerger de manière particulièrement éclatante des lacunes dans les corpus historiques, que ce soit au niveau leur constitution, de leur transmission ou dans leur nature même, dont l'objectivité, considérée à travers le spectre de méthodes issues des sciences dures, nécessite soudain re-définition. Les méthodes numériques contraignent à se poser de manière radicale la question : Qu'est-ce que construire et interpréter une source historique ?

En première partie, je reviendrai sur l'utilité de la conceptualisation des notions de trace et d'archive dans ce contexte, avant d'en venir en deuxième partie à la mise en pratique de telles réflexions dans le domaine numérique. En conclusion, je reviendrai 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.

1

#### 1. Réflexions générales liminaires

##### Trace

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 » 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, comment en faire quelque chose ensemble ? Or, il n'est pas simple de faire l'effort de conceptualiser ce sur quoi est fondé l'ensemble du travail de recherche en sciences humaines, a fortiori la dimension historique qui y préside. En Allemagne, les luttes d'école sont fréquentes, mais en histoire littéraire par exemple, elles ne concernent la plupart du temps que les méthodes de traitement des corpus, au mieux la manière de les éditer. Rares sont les cas où l'on s'interroge en profondeur sur l'archéologie du corpus lui-même. Or, 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 qui, en sciences humaines, sont souvent négligées.

Il s'agira donc 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. Dans *Temps et Récit*, Paul Ricoeur part de la tradition documentaire pour faire émerger la place de la trace dans la tradition historiographique :

"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.,"<sup>1</sup>

Deux aspects jouent ici un rôle essentiel : 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.<sup>2</sup> Ricoeur 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 »<sup>3</sup> 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."<sup>4</sup> Je ne rentrerai pas ici dans une polémique disciplinaire ; la question intéressante me semble plutôt de savoir 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 originale, même lorsque sa source primaire n'est pas la source originale 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

---

<sup>1</sup> Paul Ricoeur, *Temps et Récit III*, Paris, 1985, p. 13.

<sup>2</sup> En ce sens, voir aussi, *ibid.*, p. 212 : « Que la trace soit un tel requisit [de toutes les productions de la pratique historique] 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, delà, 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. »

<sup>3</sup> *Ibid.*, p. 226. Cf. aussi p. 227 : « Mais ce que signifie la trace est un problème non d'historien-savant, mais d'historien-philosophe. »

<sup>4</sup> *Ibid.*, p. 227.



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.

Pour Ricoeur, 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.<sup>5</sup> Or, en dérangeant un ordre, elle en constitue un autre, ou est constituée en un tel ordre, alors même que son évanescence est sa qualité ontologique première.

La trace « est fragile »<sup>6</sup> : cette assertion est vraie à différents niveaux. Tout d'abord, au sens où le présente ici Ricoeur : 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 »<sup>7</sup>. La temporalité donne à la trace sa friabilité. Enfin, et c'est ce que je voudrais à présent approfondir, la trace est autant signe d'une absence que signe d'une présence<sup>8</sup>. 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 ?) »<sup>9</sup>.

Derrida, dans *De la grammatologie*, va encore 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. »<sup>10</sup>

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

<sup>5</sup> Ibid., p. 226.

<sup>6</sup> Ibid., p. 218-219.

<sup>7</sup> Ibid., p. 219. Voir aussi en ce sens la description de la ville de Rome par Freud dans le *Malaise dans la civilisation*.

<sup>8</sup> Cf. Jacques Derrida, „On ne peut penser la trace instituée sans penser la rétention de la différence dans une structure de renvoi ou la différence apparaît *comme telle* [...]. L'absence d'un autre ici-maintenant, d'une autre présent transcendantal, d'une *autre* origine du monde apparaissant comme telle, se présentant comme absence irréductible dans la présence de la trace...” (*De la grammatologie*, p. 68)

<sup>9</sup> Ricoeur, *Temps et Récit* III, p. 212.

<sup>10</sup> Derrida, *De la grammatologie*, p. 59.



caractère originaire ou premier, et dont la pertinence pour la recherche historique nécessite bien d'être, à chaque instant, déconstruite pour être ré-évaluée.

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 la recherche historiographique. Sa fragilité même se transmet à l'institution qui en garantit la transmission en l'ordonnant : l'archive.

### Archive

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 une, mais plusieurs fois et non par une, mais 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. Voilà qui 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.

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.<sup>11</sup> Il insiste sur le fait que des erreurs ont été commises notamment dans les choix de softwares qui semblaient, au premier abord, adéquats, avant de s'avérer, une fois en phase d'opérationnalisation, totalement 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, notamment en

---

<sup>11</sup> *A Companion to the Digital Humanities*, dir. par Susan Schreibman, Ray Siemens et John Unsworth, Oxford, 2004; article de G. Crane ibid., p. 46-55 "Classics and the Computer: An End of the History"



ce qui concerne le décloisonnement des différents formats d'accès à l'information.<sup>12</sup> 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<sup>ème</sup> siècle sont un miroir aux alouettes dans ces deux sens : non seulement elles nous sont proches en termes de mentalité, mais elles sont bien plus complètes qu'elles ne le sont pour les siècles antérieurs. Il s'agit là d'une spécificité du XX<sup>ème</sup> siècle, car avec le XXI<sup>ème</sup> siècle et l'émergence de sources originellement numériques (digital born), l'approche se déplace à nouveau. Les éléments d'identification du statut épistémologique des sources originellement 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 à l'œil nu. Nous ne savons toujours pas aujourd'hui comment citer une source numérique de manière fiable, de manière à savoir exactement, en la référençant, de quoi on parle.

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

« 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. »<sup>13</sup>

Chaque consultation et interprétation de la trace archivale contribue ainsi à sa destruction ; aucune reconstitution 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 en effet à dévaluer la trace de départ et à générer l'illusion d'une identité entre la trace et sa représentation.

---

<sup>12</sup> Ibid. p. 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.”; p. 48: “ He[= 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.

<sup>13</sup> Derrida, *Mal d'archive*, p. 26-27.



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 demande Derrida : « Comment peut-[on] prétendre faire la *preuve* d'une absence d'archive? »<sup>14</sup> Dans cette perspective, il est seulement possible d'admettre la coprésence 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.

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. Bien sûr, pour Barthes, 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 me semblent 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."<sup>15</sup>

6

L'image de l'encre magique est tout à l'opposé de celle de l'archive inéluctablement prise dans l'auto-destruction. 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.

Que l'on considère l'approche des sources historiques comme un processus de destruction ou comme un processus d'enrichissement – la vérité se situe probablement à mi-chemin entre les deux –, il semble dans tous les cas difficile, voire impossible, d'échapper à un travail de reconstruction – chirurgie plastique destinée à rendre la trace interprétable –, ni d'ailleurs 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 de l'historien est bien, inévitablement, d'ordre épistémologique, et 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 Ricoeur lorsqu'il distingue historien-savant et historien-

---

<sup>14</sup> Derrida, *Mal d'archive*, p. 103.

<sup>15</sup> Roland Barthes, *Le degré zéro de l'écriture*, p. 20.



philosophe,<sup>16</sup> peine 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 je le fais ici, de manière pour l'essentiel spéculative, mais de l'associer à une démarche singulière de recherche ?

## 2. « Lettres et textes » comme repose-trace

Pour illustrer mon propos, je vais revenir sur l'expérience de constitution de corpus dans le contexte de l'édition numérique *Lettres et textes. Le Berlin intellectuel autour de 1800*<sup>17</sup>. Cette édition numérique rassemble des types de textes différents signés par des auteurs différents, réunis autour d'une question de recherche générique, à savoir les stratégies de publication développées par les intellectuels berlinois autour de 1800.<sup>18</sup> Faire émerger des tendances 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 s'étendent donc de 1760 jusqu'à la révolution de 1848 ; certains auteurs résident à Berlin, d'autres y ont résidé ou sont en contact avec des Berlinoises. Les lettres éclairent les phénomènes de genèse et de réception sous un angle qu'il s'agit de recouper avec les phénomènes textuels observables dans les brouillons d'œuvre. La question des stratégies de publication chez les intellectuels peut être déclinée en un nombre considérable 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<sup>19</sup> et l'inscription dans une tradition héritée des Lumières (francophones)<sup>20</sup>.

Il va de soi que j'aurais aimé pouvoir travailler avec davantage 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 de 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

<sup>16</sup> Cf. Ricoeur, *Temps et Récit III*, p. 227.

<sup>17</sup> <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/> (consulté le 17.03.2015)

<sup>18</sup> Les grandes lignes conceptuelles de cette édition sont élucidées dans : „Berliner Intellektuelle um 1800“ als Programm. Über Potential und Grenzen digitalen Edierens », co-écrit avec Anna Busch et publié en septembre 2015 sur literaturkritik.de ([http://www.literaturkritik.de/public/rezension.php?rez\\_id=19678&ausgabe=201409](http://www.literaturkritik.de/public/rezension.php?rez_id=19678&ausgabe=201409) consulté le 17.03.2015)

<sup>19</sup> Cf. "Qu'est-ce qui fait l'intellectuel ? Les professeurs de l'université de Berlin et leur patriotisme (1810-1820)", in : France-Allemagne. Figures de l'intellectuel entre révolution et réaction (1780-1848), dir. par A. Baillot et A. Yuva, Presses du Septentrion, 2014 [<http://halshs.archives-ouvertes.fr/halshs-00879149>]

<sup>20</sup> Cf. "Les corpus français sont-ils allemands? Éditer des textes de la Prusse fédéricienne en Allemagne", in: Philologie im Netz Beiheft 7/2014 (dir. par Christof Schöch et Lars Schneider), p. 49-63 [<http://web.fu-berlin.de/phn/beiheft7/b7t03.pdf>].





de collaboration et d'intertextualité : d'une part, les entités nommées sont informées en détail (personnes, lieux, œuvres, institutions, mises en lien 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<sup>21</sup>) permet de repérer les phénomènes de réécriture et de correction.<sup>22</sup> La combinaison des deux niveaux d'encodage est destiné 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, la plupart du temps) aux XVIIIème et au XIXème siècle, 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 n'est en l'état pas statistiquement représentative ; 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.

Avant d'en venir à la construction numérique d'une telle masse documentaire, je voudrais préciser les dimensions qu'elle fait ressortir sous l'angle de mon propos de première partie :

- 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<sup>23</sup>).
- les sources proviennent en effet de différentes archives et bibliothèques, et constituent, ainsi mises ensemble, une archive virtuelle singulière.
- cette édition est plus souvent saluée par les archivistes et bibliothécaires que par les chercheurs (en Allemagne du moins)
- c'est un corpus ouvert, qui peut toujours être enrichi et ne sera a priori jamais achevé ; l'annotation, elle aussi, est ouverte et susceptible d'être approfondie ; l'ensemble tel qu'il se présentera dans quelques mois lorsque le financement sera épuisé, constituera donc en quelque sorte un échantillon, une heuristique destinée à continuer d'être développée
- cette dimension dans le fond très expérimentale et qui m'attire les foudres d'une partie importante de ma communauté scientifique est compensée par le fait que ce

<sup>21</sup> Cf. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>.

<sup>22</sup> La documentation des choix d'encodage de cette édition est téléchargeable depuis la page : <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/about?de>.

<sup>23</sup> Concernant la définition du concept d'intellectuel, voir l'introduction à *France-Allemagne. Figures de l'intellectuel entre révolution et réaction 1780-1848*, dir. avec Ayse Yuva, Villeneuve d'Ascq/Lille, 2014.



« repose-traces » que j'ai fabriqué à partir de ma question de recherche, contient quelques perles, des manuscrits jusqu'ici inédits et de grande valeur du point de vue du canon, et sur lesquels j'ai fait, avec mon équipe, un travail pionnier

Commençons par le plus simple, à savoir les métadonnées. Pourquoi est-ce que les archives et les bibliothèques reconnaissent la valeur de cette ressource, pourquoi cela leur est-il égal qu'elle 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 correcte de travailler dans l'univers de sources parcellaires qui est le nôtre consiste à mettre si ce n'est en commun, du moins en relation, ce qu'a chacun. Le plus important, ce sont 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, le 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.

Je ne veux pas dire par là qu'il faudrait 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 est intéressant. 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 une épine dans le pied pour la Deutsche Nationalbibliothek, au moins autant que son enrichissement. Les prémisses de saisie sont également problématiques du point de vue de la recherche. Par définition, les notices d'autorité sont gérées par les bibliothèques et saisissent en priorité des données bibliothécaires. Pour les personnes, cela signifie que ce sont principalement les personnes ayant publié des livres qui seront mises à l'honneur.

Les fichiers d'autorité, donc, perpétuent le canon et ne permettent pas fondamentalement de correctif historiographique puisque se trouvent exclues de ces ressources les personnes « secondaires », notamment les femmes. Or, il faut attendre le XX<sup>ème</sup> siècle pour que les



femmes publient plus souvent sous leur nom, ou du moins sans se cacher 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 d'autres ressources : non, dans ce cas, la recherche doit contribuer à structurer le fichier d'autorité. Il faut envoyer un mail à la Gemeinsame Normdatei, mettre en lien le prêtre-nom masculin et l'auteur femme, justifier des sources, et ce pour, finalement ne faire entrer qu'un *auteur* de plus (certes de sexe féminin) au registre des traces biographiques des temps passés, en renforçant donc la domination existante du texte imprimé et en perpétuant la tradition historiographique.

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 ?).

J'en viens pour finir à la question la plus complexe. Si tant est qu'on soit en mesure de travailler avec le manque, 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 media pour être recevable ? Combien d'articles faudra-t-il que j'écrive pour que le principe même de mon édition numérique soit audible pour la communauté scientifique ?

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 media, 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 - peut-être faudrait-il ici aller chercher du côté du *Sorge* heideggerien une conceptualisation qui rende compte de cette temporalité. D'autre part, l'archive numérique doit accepter sa propre finitude. Il n'y aura sans doute jamais que 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 gigabytes de scans numériquement « invisibles ». Où est le plein, où est le creux : voilà ce que nous avons à nommer.