



**HAL**  
open science

# La littérature du XIX siècle au prisme des humanités numériques

Andrea del Lungo, Giovanni Pietro Vitali

► **To cite this version:**

Andrea del Lungo, Giovanni Pietro Vitali. La littérature du XIX siècle au prisme des humanités numériques. *Romantisme: la revue du dix-neuvième siècle*, 2021, n° 191 (1), pp.64-75. 10.3917/rom.191.0064 . hal-03375951

**HAL Id: hal-03375951**

**<https://hal.sorbonne-universite.fr/hal-03375951>**

Submitted on 13 Oct 2021

**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.

## La littérature du XIX<sup>e</sup> siècle au prisme des humanités numériques

Andrea Del Lungo et Giovanni Pietro Vitali

Dans le foisonnement actuel d'initiatives et de projets en humanités numériques, la littérature du XIX<sup>e</sup> siècle occupe sans doute une place prioritaire. Plusieurs raisons peuvent expliquer ce phénomène : la première, contingente, concerne la possibilité de numérisation de textes libres de droits. La deuxième relève de la multiplication des modes de publication à l'époque : revue, presse, recueil, volume, autant de nouveaux supports dont la variété et l'importance pour la réception ont été longtemps gommées par les éditions imprimées, ne donnant que l'état final du texte. Enfin, une dernière raison consiste dans le développement de la critique génétique, qui analyse le mouvement de création et qui trouve dans le numérique son mode éditorial idéal, notamment sur certains corpus manuscrits, très fournis, d'auteurs du XIX<sup>e</sup> siècle.

Les études dix-neuviémistes ont déjà pu bénéficier des résultats de certains projets de recherche, dont la richesse a engendré une vision quelque peu idéalisée du numérique comme la panacée à tous les problèmes éditoriaux, permettant des formes d'interrogation renouvelées des corpus. Cet article entend nuancer cette vision, à travers un relevé de l'état actuel de la recherche dans un contexte international (notamment anglophone, francophone et italoophone) qui souligne les indéniables réussites dans le domaine, ainsi que les apports en termes d'exploitation des corpus ; dans un deuxième temps, sur la base de cette enquête, seront pointés un certain nombre d'éléments problématiques plus généraux, concernant la redéfinition de la fonction éditoriale ainsi que de la notion d'auteur, le risque d'une dérive quantitative, et les difficultés auxquelles se heurte l'enrichissement critique des textes.

*Le XIX<sup>e</sup> siècle se renouvelle. Enquête sur les infrastructures et les outils numériques*

La littérature du XIX<sup>e</sup> siècle est à présent l'un des champs les plus étudiés par les humanistes numériques. La production romanesque et poétique constitue à plusieurs égards un corpus idéal, qui a donné lieu à des projets particulièrement réussis de développement des perspectives littéraires à travers l'usage du numérique : l'analyse des textes du XIX<sup>e</sup> siècle par le biais des instruments informatiques simplifie la vie des chercheurs, car l'échange et l'exploitation des textes au format numérique sont gratuits, sans aucune limitation d'usage liée aux droits d'auteur ; ces textes sont faciles à consulter en ligne, grâce à des initiatives mondiales d'échange de la mémoire littéraire, comme le *Project Gutenberg*<sup>1</sup>. Ces œuvres numériques donnent la possibilité de montrer des éditions ou des visualisations comprenant des parties significatives de texte.

Lorsqu'on évoque l'intérêt progressif des humanistes numériques pour les questions littéraires du XIX<sup>e</sup> siècle, on rencontre rapidement la nécessité d'engager une réflexion critique sur le rôle

---

<sup>1</sup> <https://www.gutenberg.org/>.

que les instruments informatiques peuvent avoir dans la réécriture d'une historiographie littéraire appropriée à notre vision contemporaine. Ce n'est pas un hasard si l'un des premiers efforts dans cette direction a été entrepris par un critique spécialisé en littérature comparée des XIX<sup>e</sup> et XX<sup>e</sup> siècles et en humanités numériques, Simone Rebora, par la publication d'un ouvrage qui se propose de retracer une historiographie littéraire du XIX<sup>e</sup> siècle, prenant en considération le rôle des instruments informatiques dans l'analyse du texte littéraire<sup>2</sup>. C'est le but de notre enquête, qui entend dresser un bilan actuel et actualisé des activités concernant les humanités numériques liées à la littérature du XIX<sup>e</sup> siècle.

Afin de retracer une carte conceptuelle de l'approche numérique, il est d'abord nécessaire de mentionner l'un des efforts les plus importants en ce sens, c'est-à-dire le *Distant Reading COST Action European Literary Text Collection (ELTeC)*<sup>3</sup>. L'objectif de ce projet international, financé par la communauté européenne, est ambitieux : il vise à créer un corpus plurilingue, principalement consacré à la période qui va de la fin du XVIII<sup>e</sup> au début du XX<sup>e</sup> siècle. L'objectif de la première phase du projet est de créer six sous-collections (100 romans par langue) pour la période 1850-1920 ; ensuite, cette première partie sera intégrée avec au moins quatre sous-collections ultérieures (100 romans par langue) pour la même période. Ces matériaux serviront de noyau central d'ELTeC, lequel sera intégré avec une troisième contribution d'au moins six collections ultérieures, en ajoutant encore d'autres langues qui n'avaient pas été considérées dans les deux premières contributions, mais cette fois pour la période précédente, de 1780 à 1850. Ce *modus operandi* dans la sélection du corpus permettra d'avoir une vision diachronique assez complète de l'histoire littéraire, en fournissant aussi une base empirique plus vaste pour des analyses spécifiques avec des objectifs comparatistes. L'ambition d'ELTeC est d'analyser la littérature, en suivant les suggestions critiques de Franco Moretti<sup>4</sup> et en appliquant plusieurs méthodologies d'analyse numérique sur un corpus plurilingue, déjà codifié en XML-TEI (*Text Encoding Initiative*) et de très grandes dimensions, comme on ne l'a jamais fait jusqu'à présent<sup>5</sup>. Cette *COST-Action* trouve sa naissance dans un terrain scientifiquement fertile, dans lequel les humanités numériques et les études littéraires sur le XIX<sup>e</sup> siècle sont au cœur d'une fervente activité de la part de plusieurs groupes dans toute l'Europe<sup>6</sup>. Le projet d'édition numérique des manuscrits de Jane Austen (*Jane Austen's Fiction Manuscripts Digital Edition*)<sup>7</sup>, dirigé par Kathryn Sutherland et Elena Pierazzo, en est un exemple. Ce projet rend accessible sur le web presque 1100 pages des manuscrits de Jane Austen, en permettant au lecteur de consulter et confronter les feuillets originaux en haute qualité.

Pour rester dans le domaine anglophone, un projet digne d'être mentionné, proposant une lecture distante de la littérature du XIX<sup>e</sup> siècle, est *Prismatic Jane Eyre*<sup>8</sup>, financé par l'agence de recherche britannique et dirigé par l'Université d'Oxford, sous la supervision de Matthew Reynolds, Sowon S. Park, Eleni Philippou et Giovanni Pietro Vitali, avec la contribution de nombreux chercheurs spécialisés en littérature britannique et *Translation Studies*. L'objectif est d'expérimenter une analyse comparée des traductions de *Jane Eyre* de Charlotte Brontë, dans le dessein de montrer que la traduction transforme le roman original en plusieurs romans

---

<sup>2</sup> Simone Rebora, *History/Histoire e Digital Humanities*, Florence, Firenze University Press, 2018.

<sup>3</sup> <https://www.distant-reading.net/eltec/>.

<sup>4</sup> Franco Moretti, *Distant Reading*, Londres-New York, Verso, 2013.

<sup>5</sup> La liste des textes présents dans ce corpus est constamment mise à jour sur la page github du projet (<https://github.com/COST-ELTeC>) et elle est accessible à ce lien : <https://distantreading.github.io/ELTeC/>

<sup>6</sup> <https://www.distant-reading.net/eltec/>.

<sup>7</sup> <https://janeausten.ac.uk/>.

<sup>8</sup> <https://prismaticjaneeyre.org/>.

différents, comme une lumière qui se décompose en passant à travers un prisme<sup>9</sup>. Un autre projet intéressant, qui associe édition numérique et *distant reading*, est *Melville Electronic Library (MEL)*<sup>10</sup> : coordonné par John Bryant, il a pour but d’imaginer un parcours thématique susceptible de diriger l’utilisateur à la découverte des textes, mais aussi du contexte historique dans lequel les romans de Melville se déroulent. Ces projets, comme ceux qui seront présentés ensuite, soulèvent la question cruciale de repenser la recherche littéraire sur le XIX<sup>e</sup> siècle, en disposant des instruments numériques pour l’analyse et pour la diffusion des résultats sur le web.

Il faut enfin, toujours dans le domaine anglophone, rappeler quelques projets dans lesquels les questions textuelles se mélangent avec les éléments graphiques dans une perspective archivistique, comme le *William Morris Archive*<sup>11</sup>, consacré à une édition des écrits de l’artiste britannique et visant à collecter la totalité de ses œuvres. Dans la même direction, le *William Blake Archive*<sup>12</sup> essaie de montrer le côté graphique de l’œuvre de Blake, en considérant sa carrière comme poète, peintre et graveur. Un autre exemple de la contribution que peuvent apporter les collectionneurs et les libraires à la création de parcours numériques de réinterprétation du patrimoine littéraire s’appuie sur les collections de la New York University<sup>13</sup>, une institution qui a beaucoup investi dans les humanités numériques ces dernières années : on y trouve plusieurs projets sur la littérature du XIX<sup>e</sup> siècle, comme le projet *African American Women Writers of the 19th Century*, une collection numérique d’œuvres publiées par des écrivaines noires du XIX<sup>e</sup> siècle<sup>14</sup>. Cette collection permet l’accès à la pensée, aux perspectives et aux capacités créatives des femmes noires, toutes condensées en des livres et des pamphlets publiés avant 1920. Une autre collection similaire est la *Circulating Library (ATCL)*, créée par Troy J. Bassett, qui a pour objectif de fournir des données biographiques et bibliographiques sur les auteurs, les éditeurs et les romans de l’époque victorienne (1837-1901)<sup>15</sup>.

Pour en venir au domaine francophone, un clivage peut s’observer par rapport au monde anglophone : dans ce dernier, les modèles de recherche proches du *distant reading* théorisé par Moretti<sup>16</sup> semblent s’imposer plus facilement ; alors que dans le domaine francophone on remarque une attention particulière pour l’édition et notamment pour les implications philologiques du travail numérique, comme dans le cas du projet eBalzac<sup>17</sup>, qui vise à créer une bibliothèque numérique, en étudiant le contenu de manière dynamique. En France, on assiste ainsi au développement de projets ambitieux pour la création de bibliothèques en ligne proposant les éditions numériques de textes littéraires, comme dans le cas de Fonte Gaia. Ce projet, coordonné par Elena Pierazzo, réunit principalement des chercheurs et des bibliothécaires italiens et français : la diffusion de la culture numérique en est l’objet aussi bien que l’instrument de recherche, car l’ensemble des « objets numériques » développés dans le domaine de la recherche offre aux chercheurs la possibilité de devenir producteurs et

---

<sup>9</sup> Matthew Reynolds (dir.), *Prismatic Translation*, Londres, Legenda, 2020, 396 pp.

<sup>10</sup> <https://melville.electroniclibrary.org/>.

<sup>11</sup> <http://morrisedition.lib.uiowa.edu/index.html>.

<sup>12</sup> <http://www.blakearchive.org/>.

<sup>13</sup> <https://guides.nyu.edu/c.php?g=276589&p=1848819>.

<sup>14</sup> [http://digital.nypl.org/schomburg/writers\\_aa19/](http://digital.nypl.org/schomburg/writers_aa19/).

<sup>15</sup> <http://www.victorianresearch.org/atcl/index.php>.

<sup>16</sup> Franco Moretti, *Graphs, maps, trees: abstract models for a literary history*, Londres-New York, Verso, 2005, 119 pp.

<sup>17</sup> <https://www.ebalzac.com/>, édition génétique et hypertextuelle de l’œuvre de Balzac, dirigée par Andrea Del Lungo.

commentateurs d'éditions numériques publiées dans la bibliothèque de Fonte Gaia<sup>18</sup> (qui n'est pas expressément consacrée au XIX<sup>e</sup> siècle, mais intègre plusieurs œuvres de cette époque). Un autre projet concernant les humanités numériques, entièrement consacré au XIX<sup>e</sup> siècle, est *CORREZ - Édition des lettres internationales adressées à Émile Zola*<sup>19</sup> : coordonné par Olivier Lumbroso, il expose numériquement les lettres adressées à Zola par ses lecteurs du monde entier. La possibilité de feuilleter les textes sur un catalogue, ou de les choisir directement sur une carte géographique, permet d'imaginer une étude paralittéraire de la figure de l'écrivain français, en fournissant un matériau unique, et souvent difficile à repérer, comme les correspondances, qui peuvent aussi être téléchargées au format XML.

D'autres projets en cours à l'université de Grenoble Alpes montrent la vivacité scientifique de cet établissement, et du CNRS local, à l'égard de l'analyse numérique des matériaux historiques et littéraires de différentes périodes. Ici, on peut faire référence à des recherches dont le point focal est le XIX<sup>e</sup> siècle, comme dans les projets sur Victor Hugo et Stendhal. Ce dernier auteur est au centre de l'activité du projet *Les manuscrits de Stendhal*, qui se fonde sur le corpus d'ébauches manuscrites de Stendhal conservées à la bibliothèque de Grenoble<sup>20</sup>. L'objectif de ce projet, sous la direction de Cécile Meynard et Thomas Lebarbé, est de visualiser les feuillets et leurs transcriptions afin de publier « les journaux et les papiers » de l'auteur, aussi bien au format papier que numérique. Dans le site, la reproduction photographique de l'original du manuscrit est présentée en vis-à-vis de sa transcription linéarisée (à orthographe modernisée), et de sa transcription pseudo-diplomatique. La liste des manuscrits est consultable par recherche directe ou à travers des sections thématiques qui comprennent la liste des volumes, des corpus collectionnés, des documents et des textes publiés par les éditions de l'université de Grenoble.

Deux autres projets grenoblois portent sur Victor Hugo. Le premier, *Carto-Hugo* (Cartographies hugoliennes<sup>21</sup>) vise à analyser, par le biais des manuscrits de l'auteur, les modalités par lesquelles la représentation cartographique suscite l'imaginaire romantique qui fonde l'œuvre. Ce projet, dirigé par Delphine Gleizes, met l'accent sur un cas particulier, un corpus d'ébauches et de dessins qui a conduit à la création du roman *Les Travailleurs de la mer* de 1866. Le deuxième projet, *Hugo en images*, se fonde également sur les aspects visuels liés à l'œuvre de cet auteur. Ce projet donne la possibilité de naviguer parmi les illustrations de l'*Édition Nationale des Œuvres Complètes de Victor Hugo*, publiée entre 1885 et 1895 sous la direction de Jules Lemonnier d'abord, et d'Émile Testard ensuite<sup>22</sup>. À travers cette bibliothèque virtuelle et visuelle, conçue par Camille Page, on peut profiter de plus de 2.000 gravures des illustrateurs et des graveurs qui ont contribué à illustrer la production hugolienne.

En tournant enfin le regard vers l'Italie, un projet intéressant présente une application expérimentale du système d'édition Ecdosis aux *Idilli*<sup>23</sup> de Giacomo Leopardi : il vise à montrer les modifications aux textes, selon les modalités typiques de la philologie d'auteur, avec une finalité principalement exégétique. Le projet est dirigé par Paola Italia à Bologne, qui coordonne également le *Wiki Leopardi*<sup>24</sup>, une collection des œuvres de l'auteur consultables au format numérique. Un autre projet visant à la diffusion numérique d'un auteur très important du XIX<sup>e</sup> siècle a été développé par Giulia Raboni et Donatella Martinelli : il s'agit d'une édition critique de l'œuvre de Manzoni, fondée sur l'encodage TEI-XML des textes et sur le conséquent

---

<sup>18</sup> <https://www.fontegaia.eu/>.

<sup>19</sup> <http://eman-archives.org/CorrespondanceZola/>.

<sup>20</sup> <http://manuscrits-de-stendhal.org/>.

<sup>21</sup> <http://cartohugo.elan-numerique.fr/>.

<sup>22</sup> <https://hugo-en-images.nakalona.fr/>.

<sup>23</sup> <http://leopardi.ecdosys.org/it/Home/>.

<sup>24</sup> <http://wikileopardi.altervista.org/>.

catalogage des manuscrits, des lettres et de la bibliographie de l'auteur<sup>25</sup>. Les italianistes qui s'occupent du XIX<sup>e</sup> siècle, extrêmement actifs, ont récemment créé un réseau international afin de proposer des actions conjointes sur la littérature de ce siècle, notamment italienne.<sup>26</sup>

Nous terminerons cette enquête par l'analyse de quelques initiatives qui lient les humanités numériques aux modes de publications, et notamment à la diffusion du texte littéraire dans la presse de l'époque : sujet qui se prête d'ailleurs à différentes analyses typiques de la chaîne de travail des humanités numériques, comme l'analyse de réseaux ou l'analyse des questions cartographiques-spatiales. En Belgique, plusieurs chercheurs, parmi lesquels Andrea Penso, étudient les modalités de publication du roman français, italien et anglais dans la presse nationale, dans une perspective comparatiste<sup>27</sup>. Dans le milieu universitaire, ces études trouvent un terrain fertile en France où elles sont représentées par le travail de chercheurs qui s'occupent de la poétique du support : un exemple en est *Medias19*<sup>28</sup>, projet financé par l'Agence nationale de la recherche (ANR) et par les Fonds Québécois de Recherche sur la Société et la Culture (FQRSC), dirigé par Marie-Ève Thérienty et Guillaume Pinson. La plateforme, développée entièrement en Lodel<sup>29</sup>, réunit les chercheurs intéressés par la presse du XIX<sup>e</sup> siècle : sur ce portail on trouve des réimpressions commentées de textes anciens, d'articles, de fiction, d'essais de journalisme et, à travers cette recherche, on propose aussi en ligne la première édition d'un dictionnaire de journalistes francophones du XIX<sup>e</sup> siècle. Le travail de Marie-Ève Thérienty et Guillaume Pinson, à l'issue de *Medias19*, s'est aussi étendu à des projets plus récents comme *Numapresse*<sup>30</sup>, co-dirigé par Julien Schuh, qui constitue à plusieurs égards une évolution de la recherche franco-canadienne mentionnée précédemment, dont il utilise certains matériaux : le projet vise à proposer une nouvelle histoire culturelle et littéraire qui soit capable de souligner les particularités de la presse française du XIX<sup>e</sup> siècle à nos jours. À la suite de la numérisation des journaux, de nouveaux instruments numériques ont été créés afin de combiner la contribution des études littéraires, de l'histoire de la culture et de l'analyse du discours. *Numapresse* est un projet ambitieux qui prévoit l'usage de différentes méthodologies d'acquisition et d'analyse du texte<sup>31</sup>. Ce projet offre des pistes d'analyse qui vont de la présentation des matériaux à l'étude spécifique des particularités linguistiques et thématiques des textes codifiés. L'approche de l'étude est assez complète et aborde par une perspective interdisciplinaire différentes questions linguistiques et thématiques.

Dresser un bilan des études sur le XIX<sup>e</sup> siècle à travers les humanités numériques est nécessaire mais il s'agit en même temps d'une opération transitoire, en raison de la grande quantité de projets en cours ou qui vont être développés. Ce qui semble évident, lorsqu'on observe la réalité française des humanités numériques, c'est que la présence de l'édition numérique dans ces études joue un rôle central. Au sein de l'aperçu qu'on a présenté jusqu'ici, on observe une claire dichotomie dans le traitement numérique de la littérature du XIX<sup>e</sup> siècle : d'une part l'approche typique du milieu universitaire français, qui se concentre sur le texte ; de l'autre, une façon de traiter les problématiques littéraires avec des méthodologies typiques du *distant reading*,

---

<sup>25</sup> Le site est actuellement en construction : <http://www.alessandromanzoni.org/>.

<sup>26</sup> <http://ottocentismi.org/>.

<sup>27</sup> Dirk Vanden Berghe, Andrea Penso, Sandra Parmegiani, « Mapping the Italian reception of English novels in the Long Eighteenth Century. A digital approach », *Quaderni d'Italianistica*, 39-2, 2018, pp. 35-59.

<sup>28</sup> <http://www.medias19.org/>.

<sup>29</sup> <https://www.lodel.org/>.

<sup>30</sup> <http://www.numapresse.org/>.

<sup>31</sup> Le projet Numapresse travaille sur un corpus de textes issus de la numérisation, l'encodage (XML-ALTO), lemmatisation (SpaCy) et l'exploitation. Cet aspect se focalise sur l'analyse des textes par *topic modelling* et du *text mining*, en utilisant un paquet pour R nommé Tidy Text.



comme dans le cas de *Prismatic Jane Eyre* ou de *COST action Distant Reading for European Literary History*, plus utilisée hors du contexte francophone.

L'impulsion vers l'édition numérique dans le domaine francophone est certainement due à une attention particulière de l'université française pour la philologie, mais on peut invoquer une autre raison. Afin de développer des analyses typiques du *distant reading* à travers des instruments numériques, il est nécessaire de posséder les textes au format numérique ; or, surtout pendant la phase de construction de beaucoup d'instruments d'analyse numérique dont on profite aujourd'hui, les premiers textes accessibles librement et en grande quantité étaient en anglais<sup>32</sup>. Les ressources textuelles, principalement en anglais, ont facilité des approches textuelles de type interprétatif, alors que dans le monde francophone, où les textes étaient difficiles à repérer, l'attention a été portée sur la collection de corpus en français, avec un intérêt particulier pour les pratiques acquisitionnelles du texte, comme la XML-TEI.

Les possibilités à venir sont nombreuses, surtout si les différents projets menés jusqu'ici pouvaient croiser leurs expériences ou mettre en commun leurs méthodes et leurs résultats, suivant le nouvel esprit de partage et de collaboration typique des humanités numériques.

### *Apports et difficultés de l'édition numérique*

L'enquête menée dans la première partie permet d'établir un bilan provisoire des projets de recherche en humanités numériques appliqués à la littérature du XIX<sup>e</sup> siècle, qui concernera ici plus spécifiquement le domaine francophone dans sa composante éditoriale (sans vouloir développer un débat autour des théories du *distant reading*, impossible à mener dans le cadre d'un article).

De toute évidence, l'apport majeur de l'ensemble de ces projets et de ces réalisations est de permettre un accès généralement libre et gratuit, au niveau planétaire, aux corpus, c'est-à-dire aux textes littéraires mais aussi à d'autres documents d'ordre génétique, contextuel ou bibliographique. À la simple consultation des textes, qui constitue déjà une avancée considérable en termes de diffusion du savoir, s'accompagne d'ailleurs la mise à disposition d'outils de recherche variés : moteurs de recherche lexicaux, concordanciers, étiquetages morfo-syntaxiques, bases de données, autant de moyens non pas inédits (car la plupart d'entre eux participent de la tradition des études littéraires), mais dont le numérique a multiplié la puissance d'interrogation des corpus. L'accessibilité que plusieurs sites donnent aux sources, par exemple aux fichiers XML-TEI, rend aussi possible leur exploitation dans d'autres projets de recherche, à des fins quantitatives : l'encodage permet de mesurer potentiellement tout élément textuel, des fréquences lexicales aux structures syntaxiques, des types de discours à leur distribution suivant les personnages, etc. L'édition numérique s'adresse évidemment à un public vaste et hétérogène, composé de lecteurs fortuits comme de chercheurs avisés : cette richesse est aussi un élément de difficulté, car la détermination d'un lectorat se révèle plus compliquée que dans le champ de l'imprimé.

Cette accessibilité historiquement inédite, si elle contribue à la démocratisation de la connaissance, ne va pas sans problèmes. Le principal en est la surabondance d'information – problème commun à l'ensemble de l'archive web, par son immensité –, ainsi que la redéfinition

---

<sup>32</sup> Pour donner un exemple, le premier texte non anglais ajouté au Projet Gutenberg a été le numéro 1000, c'est-à-dire *La Divine Comédie* de Dante en août 1987. Marie Lebert, *A Short History of eBooks*, Toronto, University of Toronto, 2009, p. 6 (<https://www.gutenberg.org/files/29801/29801-pdf.pdf>).

de la fonction éditoriale dont les garanties d'authenticité et de qualité ont constitué pendant des siècles le fondement de la culture de l'imprimé. La question est bien connue : l'utilisateur d'internet voulant lire ou exploiter par les ressources numériques un roman du XIX<sup>e</sup> siècle se retrouve face à un foisonnement de sources : une pléthore de sites s'ouvre devant lui, proposant la lecture ou le téléchargement de textes qui sont livrés à l'état brut, sans aucun enrichissement ni appareil critique, dans des versions non référencées, fautives, voire corrompues. Même le lecteur avisé, choisissant des sites de référence, trouverait des difficultés devant le caractère instable de certains textes, qui est le fruit de leur histoire éditoriale. Prenons l'exemple de *La Comédie humaine* : les éditions papier donnent à lire, depuis plus d'un siècle, la version de l'édition Furne publiée de 1842 à 1848, en intégrant les corrections que l'auteur avait effectuées sur son exemplaire personnel (version dite « Furne corrigé »), en conformité avec le précepte philologique qui assigne la priorité à la dernière version établie du vivant de l'auteur. Or, à l'exception du site eBalzac, celle-ci est introuvable en ligne, même sur des sites prestigieux et institutionnels : le CD Acamédia de 1999, dont le contenu est accessible depuis le site de la Maison de Balzac, propose la numérisation de l'édition Furne sans les corrections de Balzac, dans une version assez fautive à cause de la présence, par endroits, de nombreuses coquilles ; le site d'e-books Éféle<sup>33</sup> donne également le texte de l'édition Furne, dans une version plus fidèle ; Wikisource choisit en revanche la leçon de l'édition posthume chez Houssiaux (1855-1874), qui n'intègre toujours pas les corrections de Balzac. Le paradoxe est flagrant : les textes de Balzac qui circulent en ligne ne sont pas les mêmes que ceux qu'on lit dans les éditions imprimées.

La réponse à cette difficulté ne réside pas, on vient de le voir, dans la réputation ou la fiabilité des sites, mais plutôt dans la structuration et le référencement des données que proposent certains projets de recherche de plus grande ampleur. Nous donnerons ici deux exemples. Le premier est le site pionnier sur Flaubert de l'Université de Rouen, piloté par Yvan Leclerc, qui donne accès à l'intégralité des manuscrits de *Madame Bovary*, de *Bouvard et Pécuchet* et de la correspondance, avec une transcription diplomatique en vis-à-vis de chaque feuillet<sup>34</sup>. L'apport majeur du site, outre évidemment l'accessibilité de ce corpus très complet, est de proposer des tableaux génétiques qui présentent, pour chaque séquence du texte, la stratification chronologique de sa genèse, suivant un parcours de création parfois accidenté, qui multiplie les plans, les scénarios, les brouillons, les notes ou les passages abandonnés, pour aboutir au manuscrit définitif, puis à celui du copiste, et au texte publié. Le corpus des manuscrits, ordonné et structuré, devient ainsi pleinement accessible à l'utilisateur qui peut alors observer la complexité extraordinaire du travail sur la forme effectué par l'écrivain. Le deuxième exemple, le projet de recherche eBalzac, porte cette fois sur la génétique de l'imprimé : le site issu du projet propose une version philologiquement très exacte de *La Comédie humaine*, qui intègre les corrections de Balzac sur son exemplaire personnel ; de surcroît, il donne accès aux versions antérieures, c'est-à-dire aux états imprimés qui présentent des corrections auctoriales, à partir des premières éditions en feuilleton, en revue ou en volume. L'ensemble du site est conçu graphiquement avec une séparation verticale au centre qui mime la page du livre et qui permet surtout d'articuler deux espaces en vis-à-vis : une colonne présente le texte numérisé, et l'autre permet l'ouverture en mode image de la page du support d'origine. Sa rubrique « génétique » propose la comparaison des multiples états imprimés des textes de *La Comédie humaine*, dans lesquelles les variantes sont surlignées suivant quatre types d'opérations : suppressions, insertions, remplacements et déplacements, avec la possibilité aussi de les chiffrer ou de

---

<sup>33</sup> <http://efele.net/ebooks/>.

<sup>34</sup> <https://flaubert.univ-rouen.fr/>.



retrouver une variante spécifique<sup>35</sup>. Ce système résout le problème du relevé des variantes des éditions imprimées au moyen d'une visualisation claire de la totalité des opérations de réécriture : le devenir de l'œuvre devient ainsi lisible. On découvre par l'analyse de ces comparaisons un autre Balzac, orfèvre de la forme avant Flaubert, en quête du terme exact, de la bonne structure syntaxique, d'un rythme fluide.

Un constat, assez inattendu, émerge de cette enquête ainsi que de la description des deux sites que nous venons de mentionner : la plupart des projets en humanités numériques littéraires, y compris de ceux qui participent du *distant reading*, sont *sur auteur*. Ils exploitent donc, par les nouvelles technologies, des corpus limités à la production d'un écrivain, avec une visée d'exhaustivité qui se rapproche souvent des entreprises traditionnelles des œuvres complètes, comme c'est le cas des manuscrits de Flaubert ou de Stendhal, des éditions de Balzac ou de la correspondance de Zola. Ceci n'implique évidemment pas un retour au lansonisme, ou à une vision téléologique de l'histoire littéraire : au contraire, c'est une reconfiguration de la notion d'auteur qui est ici proposée, commune d'ailleurs à la création contemporaine nativement numérique dont les exemples majeurs sont le site de François Bon *Le Tiers Livre*<sup>36</sup>, véritable atelier d'une œuvre en devenir, ou le blog d'Éric Chevillard *L'Autofictif*<sup>37</sup>, qui explore par le fragmentaire de nouvelles formes d'écriture intime. Sous des formes différentes, ces sites redéfinissent une notion d'auteur qui intègre la fonction éditoriale et qui établit une forme inédite d'autorité (et de propriété intellectuelle) au moment même où la figure de l'auteur est saisie dans son être écrivain (ou « écranvain », selon la terminologie de Gilles Bonnet<sup>38</sup>). Les sites sur les auteurs du XIX<sup>e</sup> siècle, s'ils n'ont évidemment pas de visée créative, contribuent à une redéfinition de l'auctorialité qui procède dans le même sens : celui de participer au domaine des études sur auteur, avec toutes les implications critiques que cela comporte en termes de monumentalisation ou de canonisation, et de proposer à la fois une saisie multiple de la figure de l'auteur, par la représentation que le numérique permet de la mobilité de l'œuvre et de son devenir. Quoi qu'il en soit, le « réseau » que forme l'ensemble de ces projets sur auteur est loin de la vision de l'intertexte comme croisement d'écritures multiples, issues de plusieurs cultures et affranchies de l'autorité auctoriale, que Roland Barthes avait promue dans son article déclarant précisément la mort de l'auteur (1968) – si vivant encore, de nos jours, sur le web.

Pour avancer vers la conclusion, nous ne ferons qu'évoquer le problème relatif à la fragilité de l'archive web, déjà abordé par Milad Doueïhi<sup>39</sup>, qui concerne aussi, dans le domaine de l'édition numérique, la pérennisation des données produites dans le cadre des projets de recherche, exposée au risque de l'évolution rapide des formats, des logiciels et des systèmes d'exploitation. On touche ici à une question d'ordre économique : les sites éditoriaux nécessitent une mise à jour constante que les financements de la recherche, généralement limités dans le temps, rendent difficile. Il serait aussi possible de recourir à cette même raison, d'ordre économique, pour expliquer la difficulté à faire avancer et aboutir des projets qui, dans le domaine des humanités numériques, sont très coûteux en termes surtout de ressources humaines.

---

<sup>35</sup> Actuellement, la comparaison entre le Furne et le Furne corrigé est possible pour tous les textes : la comparaison avec des états antérieurs est déjà disponible pour quelques textes et sera à terme étendue à l'ensemble des romans de *La Comédie humaine*. Soulignons que le site Variance, de l'université de Lausanne, utilise le même système et propose des comparaisons génétiques de trois textes de Balzac et de *L'Éducation sentimentale* de Flaubert (<http://variance.ch/>).

<sup>36</sup> <https://www.tierslivre.net/>.

<sup>37</sup> <http://autofictif.blogspot.com/>.

<sup>38</sup> Voir Gilles Bonnet, *Pour une poétique numérique. Littérature et internet*, Paris, Hermann, 2017, p. 7-18.

<sup>39</sup> Milad Doueïhi, *La Grande conversion numérique*, Paris, Seuil, 2008, p. 205-238.

En effet, la réalité actuelle de l'édition numérique n'est probablement pas à la hauteur des grandes attentes aux allures borgésiennes que les nouvelles technologies ont pu susciter depuis deux décennies : accès à la bibliothèque interminable des textes du monde, « navigation » dans le vaste océan de l'hypertexte, nouvelles modalités d'édition qui signeraient la fin d'une ère, celle de la clôture du texte, à laquelle nous contraindrait l'imprimé. Le constat est assez clair : en dépit des indéniables réussites que nous avons pu détailler dans cet article, l'édition numérique reste souvent plus pauvre, et parfois moins fiable que les bons vieux livres, en termes notamment d'apparat critique. La question est donc moins économique que scientifique : elle porte sur l'exploitation des relations hypertextuelles ou des ressources multimédiales qui rendraient accessibles les savoirs, les sources, les sens et discours qui traversent un texte littéraire (on revient, par-là, à la vision barthésienne). Or, sur ce point, le numérique pose en réalité davantage de problèmes qu'il n'en résout, dans la mesure où l'équilibre entre la clôture du texte et l'ouverture virtuellement infinie de l'hypertexte reste difficile à trouver, d'un point de vue matériel mais aussi ontologique.

Nous prendrons une dernière fois l'exemple du projet eBalzac : outre les aspects éditoriaux et génétiques déjà mentionnés, son objectif est aussi de concevoir une édition hypertextuelle qui donnerait accès à la bibliothèque virtuelle de l'écrivain, grâce à un logiciel de détection d'homologies de séquence, sur base lexicale, susceptible de faire émerger des citations, des emprunts, des allusions à un ensemble de textes littéraires ou scientifiques qui ont nourri l'œuvre balzacienne. Or, cette partie expérimentale du projet s'est heurtée aux difficultés de quantification inhérentes au numérique : dans la délimitation des textes-source, potentiellement infinis, mais aussi dans le tri des résultats, trop nombreux pour être exploitables ou pour permettre une visualisation pertinente. De ce point de vue, la « page » d'un site internet limite en réalité la quantité d'informations qu'elle peut accueillir (et ce même en laissant une colonne entière pour l'hyperannotation en vis-vis du texte), et pose à plusieurs égards les mêmes problèmes, en termes de quantité d'annotations, que la page du livre. La réponse à ce problème est évidemment l'hyperlien, qui permet la relation à une autre page du web, selon une logique résiliaire et non linéaire ; mais dans un site éditorial, l'hyperlien qui ferait naviguer l'utilisateur vers d'autres ressources extérieures s'exposerait au risque du non-retour de l'utilisateur en question : le but du site serait ainsi manqué, car il s'agit avant tout de faire lire le texte objet de l'édition.

La principale difficulté que doit affronter l'édition numérique est donc celle de l'enrichissement critique, pour l'instant relativement peu développé, alors que le numérique promettait, suivant une vision un peu naïve, la résolution de tous les problèmes éditoriaux qui se posent dans les éditions imprimées, concernant l'établissement du texte, les variantes, les notes, par le développement de ressources hypertextuelles et multimédiales. À l'ensemble des difficultés matérielles que nous avons évoquées, il faut alors ajouter une question ontologique : est-ce vraiment souhaitable, dans le cadre d'une édition numérique, de fournir à l'utilisateur un appareil critique enrichi ? Pour simplifier à outrance, de lui montrer par un hyperlien tel tableau que décrit l'auteur, de l'amener dans tel site où se déroule l'action, de lui faire écouter telle sonate qu'entend un personnage<sup>40</sup> ? Et comment trier une masse d'informations virtuellement infinie ? Le tournant du numérique pourrait permettre de nous affranchir pour toujours de la page, mais aussi du texte, comme ce qui est défini par son début et sa fin, clos sur lui-même (selon un héritage structuraliste) et sacralisé par sa concrétisation dans l'espace du livre, avec l'imaginaire culturel qui l'accompagne<sup>41</sup>. Suivant cette logique, qui rompt avec la culture

---

<sup>40</sup> Pour un exemple d'édition enrichie, voir le *Candide* de Voltaire de la BNF (<https://candide.bnf.fr/>).

<sup>41</sup> Voir Andrea Del Lungo, « Problèmes de l'annotation numérique », dans *Des notes et des textes. Étude sur l'annotation*, Franc Schuerewegen (dir.), CRIN (*Cahiers de Recherche des Instituts Néerlandais de langue et de littérature françaises*), 67, 2020.

éditoriale traditionnelle, ce serait plutôt à l'utilisateur de mettre à profit ses compétences de navigateur dans l'archive du web, où chacun trouve son chemin par rapport à ce qu'il cherche, mais aussi à ses attitudes et son vécu.

La réintroduction d'une forme de subjectivité paraît en effet nécessaire afin de combattre le risque d'une dérive quantitative à laquelle s'exposent les études littéraires, les humanités numériques fournissant des outils qui pourraient être conçus comme des instruments de mesure au sein d'une approche prétendument objective du texte. L'enjeu actuel, que doivent affronter les recherches littéraires dans le domaine, consiste à réaliser une alliance du quantitatif et de l'herméneutique<sup>42</sup> dont la clé réside dans la place laissée à la subjectivité critique, mais aussi dans une nouvelle conception de l'édition : sans suivre davantage la chimère de l'exhaustivité que fait miroiter le numérique, celle-ci gagnerait peut-être à se défaire de la logique fortement auctoriale qui a jusqu'ici présidé aux projets en humanités numériques littéraires, pour s'ouvrir plus amplement à la logique de l'hypertexte.

(Sorbonne Université, CELLF et UVSQ, DYPAC)

---

<sup>42</sup> Voir à ce propos le récent article de Franco Moretti, « Les études littéraires entre herméneutique et quantification », RIEF (Revue Italienne d'Études Françaises), 10, 2020 (<https://journals.openedition.org/rief/5359>).