

Diplôme national de master

Domaine - sciences humaines et sociales

Mention – sciences de l’information et des bibliothèques

Parcours – archives numériques

## **Les correspondances des écrivains du XIX<sup>ème</sup> siècle et leur archivage numérique : le cas de Gustave Flaubert et d’Émile Zola**

**Jérémy Da Silva**

Sous la direction de Clément Oury  
Conservateur des bibliothèques, Chef du service des données, du réseau et des  
normes –Centre ISSN international



## ***Remerciements***

*Je remercie tout d'abord M. Clément Oury, qui a accepté de diriger ce mémoire et qui m'a aidé, avec compréhension et sympathie, à ne pas perdre de vue ma problématique de recherche au gré de la réalisation de ce travail.*

*Je remercie Mme. Danielle Girard ainsi que M. Yvan Leclerc, responsables du projet d'édition numérique des correspondances de Gustave Flaubert, pour m'avoir partagé avec enthousiasme leur expérience, dans ce domaine des correspondances d'écrivain sur le Web...*

*Je remercie également M. Jean-Sébastien Macke, responsable du projet CorrELEZ, d'avoir aimablement répondu à mes questions portant sur le projet d'édition numérique des correspondances d'Emile Zola sur le Web, et d'avoir élargi en conséquence ma vision de ce sujet.*

*Je tenais à remercier un camarade de ma promotion, Jean-Baptiste Dutilleul, d'avoir été à l'origine de la rédaction ce mémoire en m'évoquant le projet du Centre Flaubert, qui a instantanément fait surgir en moi des interrogations ayant trait aux archives numériques épistolaires.*

*Enfin, je remercie ma famille et mes proches, pour leur soutien moral sans faille, porté à mon attention tout au long de ces mois de travail.*

**Résumé :**

*À partir du XIX<sup>ème</sup>, la lettre a cessé d'être un outil de conquête sociale pour les écrivains, se transformant en un objet signifiant, révélateur d'une époque. Les correspondances d'écrivain, à l'instar de celles de Gustave Flaubert et d'Émile Zola, sont devenues des fonds patrimoniaux dont la conservation et la description étaient alors un défi archivistique d'envergure, compte tenu de leur volume. L'arrivée du web a offert aux chercheurs la possibilité de numériser ces corpus, de les retranscrire et de les indexer de façon à pérenniser leur contenu à travers des plateformes d'archivage évolutives. Le partage de ces correspondances à un plus vaste public a été l'une des grandes valorisations induites par le numérique. Des gisements nouveaux de la recherche sont nés grâce à l'enrichissement et l'interopérabilité de ces archives épistolaires numériques sur le Web...*

**Descripteurs :** *Correspondance – Numérique – Archive - Plateforme – Patrimoine*

**Abstract :**

*From the nineteenth century, the letter ceased to be a tool of social conquest for writers, turning into a meaningful object, revealing a new era. The correspondence of writer, like those of Gustave Flaubert and Émile Zola became heritage funds whose conservation and description was then a major archival challenge, given their volume. The arrival of the web has offered researchers the opportunity to digitize these corpora, to transcribe and index them in order to perpetuate their content through scalable archiving platforms. Sharing these correspondences with a larger audience has been one of the major digital-driven valuations. New sources of research will be born from the enrichment of these digital epistolary archives and their interoperability on the Web ...*

**Keywords :** *Correspondence – Digital – Archiving - Platform – Héritage*

## ***Droits d'auteurs***



*Cette création est mise à disposition selon le Contrat:*

***Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France***  
*disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou*  
*parcourrier postal à Creative Commons, 171 Second Street, Suite 300, San*  
*Francisco, California 94105, USA.*



# Sommaire

<b>SIGLES ET ABRÉVIATIONS.....</b>	<b>9</b>
<b>INTRODUCTION.....</b>	<b>10</b>
<b>I. LES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE : QUELLES SPÉCIFICITÉS ET QUELS USAGES ?.....</b>	<b>13</b>
<b>I.1. L'épistolaire à l'âge de l'industrie et du romantisme.....</b>	<b>13</b>
I.1.1. La lettre au sein des transformations structurelles de la société.....	13
I.1.2. Introspection lettrée au XIX <sup>ème</sup> siècle.....	15
I.1.3. Une conservation recrudescence mais contrainte des manuscrits de lettres.....	17
<b>I.2. Les correspondances de Flaubert et de Zola.....</b>	<b>19</b>
I.2.1. Ce que nous disent ces écrivains et leurs destinataires, dans les correspondances.....	20
I.2.2. Vers une approche génétique des correspondances.....	24
I.2.3. La conservation des fonds de correspondances de Gustave Flaubert et d'Émile Zola.....	27
<b>I.3. L'appropriation des fonds de lettres de ces deux écrivains..</b>	<b>29</b>
I.3.1. Les éditions papiers des correspondances : leurs limites.....	29
I.3.2. La nécessité des collaborations archivistiques pour la contextualisation des correspondances : le cas Proust.....	33
I.3.3. La construction d'une archive épistolaire du XIX <sup>ème</sup> siècle.....	35
<b>II. LES ARCHIVES NUMÉRIQUES FACE AUX CORPUS DES CORRESPONDANCES DE GUSTAVE FLAUBERT ET D'ÉMILE ZOLA : ÉTUDE DE L'EXISTANT SUR LE WEB.....</b>	<b>39</b>
<b>II.1. Les premières approches des équipes scientifiques face aux     corpus de lettres numérisées.....</b>	<b>39</b>
II.1.1. Contexte de ces projets.....	39
II.1.2. La numérisation des manuscrits de lettres.....	42

II.1.3. Enjeux de l'« auctorialité » avec la transcription et l'annotation numérique des correspondances.....	44
<b>II.2. Indexation numérique des correspondances.....</b>	<b>49</b>
II.2.1. Description minimale du contenu des lettres numériques à l'aide des spécifications de Dublin Core.....	49
II.2.2. Perspectives d'un processus d'encodage numérique plus détaillé des correspondances d'écrivain.....	54
II.2.3. Le protocole TEI-XML à la lumière de la lettre d'écrivain à indexer : détails et développements.....	56
<b>II.3. Exploitation des corpus épistolaires de correspondances d'écrivain sur les plateformes du Web.....</b>	<b>60</b>
II.3.1. Écueils et perspectives d'une base de données épistolaire.....	60
II.3.2. Délimitation du corpus de lettres numériques et mise en accessibilité sur la plateforme du Web .....	63
II.3.3. Maintien de l'archivage numérique d'une correspondance d'écrivain : à quelles conditions ?.....	65
<b>III. PERSPECTIVES DE L'ARCHIVAGE NUMÉRIQUE DES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE SUR LES PLATEFORMES DU WEB.....</b>	<b>68</b>
<b>III.1. Extension des corpus épistolaires présents sur le Web aux Humanités Numériques.....</b>	<b>68</b>
III.1.1. Le projet Mapping the Republic of Letters et les fonctionnalités de représentations visuelle de corpus.....	68
III.1.2. Gestion de l'incertitude des données épistolaires numériques et perspectives humanistiques.....	70
III.1.3. Perspectives des Humanités Numériques pour les correspondances Gustave Flaubert et d'Émile Zola sur le Web.....	73
<b>III.2. Enjeux de l'ouverture des corpus épistolaires aux collections numériques des institutions patrimoniales sur le Web.....</b>	<b>75</b>
<b>III.3. Quels publics et quels usages pour l'archive épistolaire numérique sur le Web ?.....</b>	<b>79</b>
III.3.1. Médiation scientifique par le dépouillement des données numériques des lettres au sein de nouvelles plateformes du Web.....	80

III.3.2. Médiation pour les recherches en SHS à travers l'ouverture de la plateforme d'archivage épistolaire numérique des correspondances.....	82
<b>CONCLUSION.....</b>	<b>85</b>
<b>ANNEXE N°1.....</b>	<b>87</b>
<b>ANNEXE N°2.....</b>	<b>93</b>
<b>BIBLIOGRAPHIE.....</b>	<b>96</b>
<b>TABLE DES MATIÈRES.....</b>	<b>103</b>

# SIGLES ET ABRÉVIATIONS

**BHVP : Bibliothèque Historique de la Ville de Paris**

**BMC : Bibliothèque Municipale Classée**

**CINES : Centre Informatique National de l'Enseignement Supérieur**

**COLLEX : Collections d'Excellence**

**EAD : Encoded Archival Description**

**OAI-PMH : Open Archives Initiative - Protocol for Metadata Harvesting**

**OCR : Optical Character Recognition**

**PDF : Portable Document Format**

**SGML : Standardized Generalised Markup Language**

**SHS : Sciences Humaines et Sociales**

**TEI : Text Encoding Initiative**

**TGIR : Très Grandes Infrastructures de Recherches**

**VIAF : Virtual International Authority File**

**XML : Extensible Markup Language**

# INTRODUCTION

Une lettre est un message écrit, et peut également être nommée en tant qu'une « épître » ou qu'une « missive ». Ces deux termes correspondent, étymologiquement, à l'acte d'envoi, qui donnera lui-même naissance au terme et adjectif « épistolaire ». Celui-ci se rapporte à toute correspondance par lettres manuscrites, et provient du verbe grec « epistellein », qui signifie « envoyer à ». La correspondance évoque quant à elle un « échange régulier de lettres entre deux personnes » et renvoie, dans un second temps, à « l'ensemble des lettres alimentant cet échange ». Elle est donc tout à la fois un acte de communication et un fonds d'archives. Avec ces quelques définitions, nous remarquons que la forme de l'échange épistolaire, ses modalités de diffusion et de conservation, ne sont pas les mêmes lorsqu'il s'agira d'une « relation épistolaire », d'un « commerce épistolaire » ou encore d'un « roman épistolaire ».

Ces subtilités de langage nous montrent aussi qu'originellement, « il n'y a pas de genre épistolaire : du moins dans le sens littéraire du mot genre »<sup>1</sup>. Les affaires de genre et de style, en art, peuvent effectivement être rattachés, de prime abord, à la permanence de leur contenu.

« La lettre, au contraire, apparaît comme une forme fixe, où s'investissent des contenus différents. Forme-signe, signifiant manifeste mais signifié insaisissable, elle occupe dans l'archéologie intellectuelle d'une époque une position qu'il est nécessaire de décrire, et, si possible, d'interpréter. »<sup>2</sup>.

Mais l'épistolaire se cantonne-t-il à ces actes d'envoi et de réception. lorsque les écrivains s'en saisissent afin d'évoquer par exemple leur création littéraire ? L'écrivain qui se révèle intimement dans ses lettres, qui dépeint la société au gré des rencontres qu'il effectue, ou encore qui partage ses pensées sur les grandes questions de la vie, se met à retenir l'attention publique. Si la portée initiale de ces documents est jugée parfois « anecdotique », pour citer certains écrivains procédant à leur propre auto-critique, c'est justement à partir de là que l'on pourra postuler la « sincérité » de l'acte épistolaire et nous y intéresser. La valeur confidentielle que ces lettres acquièrent leur confère d'emblée une postérité. Ses potentiels lecteurs entendent entrer dans le cercle privé de l'artiste. « Les correspondances sont un fabuleux objet de fixation pour notre goût pour le mystérieux et le caché » et représentent, aujourd'hui encore, « une part très importante des consultations de manuscrits contemporains »<sup>3</sup> dans les bibliothèques publiques. Donc, non seulement la correspondance crée du lien entre un expéditeur et un destinataire mais elle l'entretient de surcroît à travers le temps... Et ce d'une autre façon encore que l'autobiographie ou que le journal, les deux genres littéraires se rapprochant le plus de la lettre, mode d'écriture à toutes fins irremplaçable, qui prendra sa vraie mesure seulement au XIX<sup>ème</sup> siècle, comme nous le verrons...

Abordées à partir de leur matérialité, les correspondances d'écrivains possèdent en ce sens un double statut. D'une part elles constituent des archives dont les processus de patrimonialisation sont complexes, de par leur nombre souvent exponentiel et leur éclatement spatio-temporel. Pour les spécialistes d'un écrivain souhaitant accéder à ces textes, les

<sup>1</sup> LANSON, Gustave. *Choix de lettres du XVIIIème siècle*. Paris : Hachette, 1921.

<sup>2</sup> PAGES, Alain. « Stratégies textuelles : la lettre à la fin du XIXème siècle ». *Littérature*, n°31, 1978, pp.107-116. Accessible en ligne : [https://www.persee.fr/doc/litt\\_0047-4800\\_1978\\_num\\_31\\_3\\_1168](https://www.persee.fr/doc/litt_0047-4800_1978_num_31_3_1168)

<sup>3</sup> LE BRAS, Laurence. *Les correspondances : réception ; inventaire ; communication*, Enssib : Mémoire d'étude DCB, 2002, pp.15-17

retrouver est le premier enjeu, avant même de songer à leur assemblage et à leur valorisation. Nous parlons ici des interactions entre la logique de l'archiviste privé et celle des institutions publiques de conservation, qui constituent l'approche pratique de la correspondance. Vont-elles déboucher sur un conflit d'intérêts à cause d'une divergence d'approches de ces matériaux dans les deux camps, ou bien sur une collaboration ? D'un autre côté, il faut aussi analyser les effets de la « manipulation » de ces archives sur l'interprétation de l'œuvre de l'écrivain, entre fausses pistes et reconfigurations... Celles-ci pourront acquérir une dimension littéraire, en permettant d'entrer dans l'univers d'un écrivain par la périphérie de son œuvre, sans pour autant que cette position ne soit un déclassement de leur valeur intrinsèque. Ces contradictions feront l'objet de l'approche « herméneutique », qui consiste à interroger le statut même de l'archive dans la création littéraire : dans quelle mesure, par exemple, la lettre possède-t-elle une valeur de vérité ? Joue-t-elle un rôle référentiel, ou contraire est-elle irrémédiablement prise dans un processus de fictionnalisation ? Va-t-elle dans le sens d'une personnalisation ou d'une dépersonnalisation de l'œuvre ?

La lettre manuscrite subsiste aujourd'hui, surtout dans un cadre professionnel avec la correspondance administrative, et trouve également dans la carte postale un prolongement de son ossature première. En tant que support physique d'information, elle est depuis toujours un médium porteur de sens, mais ceci à une échelle encore plus vaste, lorsqu'elle va être le fruit d'une personne influente. Cet intérêt s'accroît au XIX<sup>ème</sup> siècle avec la naissance de l'éditeur. « Le XIX<sup>ème</sup> siècle qu'on a considéré selon l'expression de Paul Bénichou comme l'époque du « sacre de l'écrivain » est aussi celle de la sacralisation de ses objets de travail »<sup>4</sup>. Les correspondances en font fondamentalement partie, avec toutes les sortes de traces laissées par l'écrivain, qu'elle recèle en surface, participant de par leur présence à un certain « sacerdoce » de l'écrivain. Le rapport de l'écrivain avec ses lettres est de plus en plus ambigu. Deux écrivains de cette époque vont condenser selon nous ces interrogations sur le statut d'archive épistolaire, à savoir Gustave Flaubert et Émile Zola. Nous tenterons donc d'établir en quoi les correspondances de ces écrivains ont à nous apprendre non seulement dans leur contenu, mais également dans leur méta-langage qu'il advient de mettre en forme.

Partant de ces nouvelles approches du document manuscrit, qui signifient de nouvelles modalités de conservation et de pérennisation du contenu de la lettre, nous nous attacherons à décrire les possibilités offertes par le numérique face à ces enjeux. Notre travail n'aura donc pas pour optique de dévoiler le substrat des correspondances de ces deux écrivains, mais de démontrer dans quelles mesures des plateformes sur le Web vont permettre de s'appropriier ces corpus uniques en leur genre, dans l'Histoire. Surtout, quels type d'interface numérique sera la plus susceptible de valoriser ces fonds de lettres dématérialisées, et partant, leurs données ?

Aujourd'hui ce ne sont plus des chercheurs isolés mais le plus souvent des équipes scientifiques qui se saisissent des fonds de correspondance d'un écrivain pour en proposer, dans le meilleur des cas, une édition numérique augmentée. Mais les travaux de transcription et d'éclairages contextuels qu'ils vont effectuer amènent ces lettres vers une autre dimension, que celle de l'édition savante du XX<sup>ème</sup> siècle. Les missions de conservation culturelle des documents historiques s'en voient bouleversées, au regard de l'émergence progressive d'archives numériques épistolaires sur le Web, dont nous retracerons le cheminement... Ceci en nous appuyant sur deux projets significatifs, au cours de ces dernières années, portés par des équipes du Centre Flaubert et du Centre d'Étude sur Zola et le Naturalisme.

<sup>4</sup> SIMONET-TENANT, Françoise. « La critique génétique : définition, intérêts, limites ». Génétique, © Publications numériques du CÉRÉDI, « Ressources », 2019, p.2. En ligne : <http://publis-shs.univ-rouen.fr/ceredi/index.php?id=603>

Les notices documentaires papier ne font plus foi, « ce mode de repérage étant graduellement remplacé par les métadonnées et la recherche plein-texte »<sup>5</sup>. C'est tout l'enjeu, pour ces chercheurs, de parvenir à une indexation numérique complète pour chaque lettre, avec le respect des normes de standardisation et d'interopérabilité de leurs données à la clé. Cela en considérant toutes les entrées de la correspondance, en la transformant en un « pôle » de diffusion, à la fois littéraire et historique, sur le Web... L'enrichissement numérique des lettres d'écrivains va résulter de tous ces efforts de conciliation.

Comme nous le verrons, le XIX<sup>ème</sup> siècle, qui nous intéresse au premier plan avec les figures de Zola et de Flaubert, ne fait pas encore l'objet d'une plateforme sur le web qui regrouperait les correspondances de plusieurs écrivains emblématiques de cette période. À l'inverse, des projets tels que « Mapping the Republic of Letters »<sup>6</sup> pour le siècle des Lumières, ou encore « EpistolART »<sup>7</sup> pour la période de la Renaissance, proposent d'ores et déjà ces perspectives de mutualisation de corpus. L'intégration des Humanités Numériques à ces projets est l'une des raisons expliquant l'établissement de leur archivage numérique des correspondances. Les perspectives humanistiques nouvelles induites par ces projets seront alors à décrire. Cela nous amènera à nous demander en définitive dans quelle mesure les correspondances d'écrivains de Gustave Flaubert et d'Emile Zola pourraient-elles être valorisées à travers leur archivage numérique sur des plateformes dédiées au sein du Web.

Nous allons tout d'abord définir la lettre et la place qu'elle occupa dans le paysage non seulement littéraire mais également social du XIX<sup>ème</sup> siècle. La comparaison des correspondances de Gustave Flaubert et d'Émile Zola nous permettra d'illustrer cette problématique, en intégrant progressivement la question archivistique entourant ces fonds spécifiques de lettres. Ceci, dans le but de présenter les potentialités de la recherche littéraire et historique incluses par leur intermède. L'étude des modalités de conservation et de diffusion physique des fonds de correspondances de ces écrivains nous montrera, dans un second temps, comment les interfaces du numérique pourront se saisir de ces ensembles de documents manuscrits et patrimoniaux. Cela fera l'objet de notre seconde partie, où nous nous intéresserons aux questions de l'appropriation et de l'indexation numérique des lettres sur le Web, à la lumière de deux projets d'édition numérique des correspondances de Flaubert et Zola portés par des équipes scientifiques. L'organisation des données numériques de la lettre au sein des bases de données et des plateformes sur le Web marquera, alors, le premier moment de la constitution d'une archive numérique épistolaire. Les standards archivistiques internationaux, en la matière, seront à décrire. Au terme de cette étude de l'existant, nous dresserons les perspectives s'offrant aux équipes scientifiques porteuses de ces projets, en nous intéressant aux Humanités Numériques, mais également aux questions d'interopérabilité du corpus épistolaire et à son ouverture aux collections numériques sur le Web. La valorisation des correspondances d'écrivain pourra alors naître de ces évolutions intrinsèques et extrinsèques des corpus de lettres numériques, ainsi que de leur mise en accessibilité réelle sur le Web.

<sup>5</sup> COUTURE Carol, LAJEUNESSE Marcel. *L'archivistique à l'ère du numérique : les éléments fondamentaux de la discipline*. Québec : Presses de l'université du Québec, 2014.

<sup>6</sup> <http://republicofletters.stanford.edu/>

<sup>7</sup> [http://web.philo.ulg.ac.be/epistolart\\_bd/](http://web.philo.ulg.ac.be/epistolart_bd/)

# I. LES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE : SPÉCIFICITÉS ET USAGES

## I.1 L'épistolaire à l'âge de l'industrie et du romantisme

### I.1.1. La lettre au sein des transformations structurelles de la société

Pour bien comprendre le rapport qui s'établit peu à peu au XIX<sup>ème</sup> siècle entre l'écrivain et la lettre, un retour au siècle des Lumières, voir même à celui du classicisme, s'impose à nous. Avec la publication de la première édition des lettres de Madame de Sévigné en 1725, la frontière va devenir de plus en plus poreuse entre la sphère littéraire et le domaine de l'épistolaire. Au point de parfois confondre la nature de ces écrits qui n'est fondamentalement pas la même. « Ces transferts sont possibles parce que la lettre et la littérature partagent le même statut de communication différée »<sup>8</sup>. Telle une chroniqueuse de la mondanité sous Louis XIV, l'épistolière va dresser en quelque sorte la gazette d'une société mondaine, caractérisée par l'entre-soi et l'attente de l'autre. La lettre est d'ores et déjà une « incarnation de l'écriture de l'absence »<sup>9</sup>. C'est ainsi que progressivement, un réseau épistolaire se tisse à travers l'Europe, voir même le monde. Les lettres d'un expéditeur se retrouvent ainsi bien souvent dans la main de destinataires tierces, ce qui permet aux personnes influentes d'accroître, du même coup, leur réseau de connaissance par le seul acte d'envoi de la lettre. C'est alors l'essor de la République des lettres, c'est-à-dire de la communauté mondiale des hommes savants, en relation les uns avec les autres et qui se font part de leurs savoirs respectifs, de leurs dernières trouvailles intellectuelles... Pour rappel : la lettre était alors le seul moyen de communication écrite à distance. Aujourd'hui, ce rapport de l'écrivain à la lettre, en tant que médium exclusif, est depuis longtemps révolu. Pourtant, il perdura encore quelques décennies mais en évoluant d'une façon pour le moins ambiguë. Les coordonnées des échanges épistolaires se sont en effet progressivement resserré avec le passage au XIX<sup>ème</sup> siècle, à l'âge du « romantisme » et de la machine à vapeur... Ce qui apparaît plutôt paradoxal à première vue.

Comment, des correspondances du XVIII<sup>ème</sup> siècle, qui renvoyaient l'image de « la scénographie réglée d'une société harmonieuse » ; avons-nous glissé vers celles du XIX<sup>ème</sup> siècle, présentant un « terrain d'expression d'une parole plus centrée sur elle-même qu'adressée à l'autre »<sup>10</sup> ? Nous pensons spontanément aux grands mouvements artistiques de l'époque, qui auraient imprimé chez les écrivains, au fil des décennies, une nouvelle vision du monde ressurgissant tout naturellement dans leurs lettres... En découle l'expression de la mélancolie chez l'écrivain romantique, dont les missives interposées accentuent cette posture vouée à la solitude de l'être. Mais aussi, dans un autre registre, le repli sur soi de l'écrivain réaliste, qui consigne dans ses lettres toutes sortes d'éléments descriptifs qu'il réutilisera dans ses œuvres, aux antipodes de l'*accessoire* mondain. Ces deux types d'écrivains se rejoignaient finalement en cherchant à dépeindre dans leurs lettres à la fois leurs sentiments profonds et ceux de leurs contemporains,... La lettre expédiée à un destinataire, qui pouvait très bien s'égarer en chemin

<sup>8</sup> FERREYROLLES, Gérard. « L'épistolaire, à la lettre ». *Littératures* 2010/1, n°71, p.25.

<sup>9</sup> KHOUZEIMI, Sami. *L'interaction épistolaire au XVIIIème siècle*. Université d'Orléans, Pôle Universités Centre Val de Loire, 2013, p.8. En ligne : [ftp://ftp.univ-orleans.fr/theses/sami.khouzeimi\\_3194\\_vm.pdf](ftp://ftp.univ-orleans.fr/theses/sami.khouzeimi_3194_vm.pdf)

<sup>10</sup> DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Ecriture », 2002, p.35.

ou voir son encre se troubler, est comme une forme d'expansion de cette quête d'horizons, qui semble ne plus devoir finir chez l'artiste du XIX<sup>ème</sup> siècle. En cela, elle défie frontalement la structure du temps, mais aussi celle d'une société jusqu'ici communément organisée, où les auteurs avaient leur fonction, davantage sociale qu'artistique, à remplir. Ils étaient en cela attendus au tournant, mais d'une toute autre façon que pourront l'être leurs successeurs, pris quant à eux dans le joug d'un monde en surrégime.

Pourtant, les raisons de cette transmutation que nous allons essayer d'établir sont moins culturelles que structurelles. Nous évoquons ici les mutations socio-économiques de l'époque, et comment celles-ci ont bouleversé les spécificités mêmes de ce moyen d'expression de la lettre. Avec les grands phénomènes qu'elle vit naître, à l'instar de l'urbanisation ou encore la maîtrise de l'électricité, la société industrielle du XIX<sup>ème</sup> siècle est celle de la concentration des forces. Le quotidien de la population, à commencer par celui des cercles mondains dans lesquels évoluaient les écrivains, s'en est alors vu alors transfiguré.

« [Au XIX<sup>ème</sup> siècle] la correspondance se présente de fait et par nature comme un écrit concentrant tout ce qui relève de l'ordinaire au sens (...) large et moderne du terme : obligation d'immédiateté, poids des composantes sociales, part d'introspection, mode particulier de transmission, rôle fondamental de l'espace »<sup>11</sup> (p.70)

Le terme d'« ordinaire », évoqué ci-dessus, pourrait dans un premier temps nous faire songer à certaines futilités, échangés par exemple dans la mondanité de l'époque. Il n'en est pourtant rien. Bien que ce milieu social a continué à être exposé en arrière plan de certaines correspondances d'écrivains de ce siècle, l'emploi de ce terme à cette époque est toute autre. Ici, ce qui est ordinaire est au contraire plutôt à comprendre comme ce qui ne dépasse pas le niveau commun. Il y a aussi une dimension temporelle importante, car posé en ces termes, l'ordinaire est affilié au moment présent. Là où les savants des Lumières se projetaient quant à eux sans cesse, à travers leurs lettres, dans l'espace et le temps...

Ce nouveau rapport va s'instituer très rapidement au XIX<sup>ème</sup> siècle, par « l'interruption des techniques de communication modernes au service d'une réalité nouvelle et encore sans nom, l'actualité »<sup>12</sup>. L'explication est phénoménologique mais les conséquences vont être imprévisibles. Car cette actualité nouvelle, ou pourrait-on dire en d'autres termes cette conscience accrue du temps présent, est pleinement ressentie par les écrivains de ce siècle, qui voient leur rapport au monde bouleversée. En nous appuyant sur les exemples de Gustave Flaubert et d'Émile Zola, nous nous attacherons dans un second temps à montrer comment les écrivains présentaient effectivement une réceptivité significative aux faits qui leur étaient actuels. Les répercussions se déploieront dans leurs lettres jusqu'à s'ouvrir aux champs psychologiques, politiques et métaphysiques. Tous ces sujets d'actualité seront donc l'occasion de faire émerger le « sujet » individuel ou de le voir, parfois, s'effacer derrière eux.

Comment l'écrivain a-t-il contrecarré le règne naissant de l'informationnel au point de s'en servir comme ressource artistique ? En l'espace d'un siècle, la lettre est en effet passée d'un statut fonctionnel, en tant que support d'échange de l'information, à celui de jalon dans la construction de la subjectivité de l'écrivain. L'écrivain du XIX<sup>ème</sup> siècle a tout d'abord assumé cette mutation, aux contours avant tout sociaux, pouvant se résumer grâce à l'expression de Roger Chartier : « le réseau et le secret, le secret contre le réseau »<sup>13</sup>.

<sup>11</sup> HOOCK-DEMARLE, Marie-Claire. « Correspondances féminines au XIX<sup>ème</sup> siècle : de l'écrit ordinaire au réseau », *Clio : Femmes, Genre, Histoire*, n°35, 2012, p.70. Accessible en ligne : <https://journals.openedition.org/cliio/10507>

<sup>12</sup> Ibid. p.73.

On pourrait voir ici les prémices d'un phénomène que Serge Tisseron nommera « extimité »<sup>14</sup> afin de désigner la proportion grandissante de l'intime chez l'écrivain. L'écrivain ne va dès lors plus chercher à se fondre derrière une communauté grandissante en charge du savoir. Il s'agira pour lui d'exprimer un désir de communiquer avec autrui, à propos de son monde intérieur, pour pouvoir en quelque sorte mieux se l'approprier en retour. Et donc de considérer, que par soi-même, l'écrivain est limité dans son champ d'action et d'expression. Seule la lettre est habilitée, en tant que « medium » à part entière, à incarner de cette façon le message qu'elle véhicule

À partir de ce moment, la lettre n'est plus perçue par les écrivains comme un instrument de la conquête sociale, elle devient le réceptacle privilégié des tensions ambiantes. Le XVIII<sup>ème</sup> se contentait d'historiser les phénomènes observés et de les diffuser, lorsque le XIX<sup>ème</sup> siècle va les recevoir et les interroger directement, à-même le matériau épistolaire. Cela marque un tournant dans le rôle effectif de la lettre chez l'artiste, qui va être dès lors la pierre angulaire, pour ce dernier, de la caractérisation de l'émotion.

### I.1.2. Introspection lettrée au XIX<sup>ème</sup> siècle

Nous avons entrevu comment, après le siècle des Lumières, une certaine forme de l'épistolaire s'est faite évincée par les bruissements d'un nouveau langage, avant de se renouveler sous l'influx de ses grands écrivains. C'est ainsi que le XIX<sup>ème</sup> siècle marque un tournant dans l'histoire : c'est le premier siècle «qui s'est posé (lui-même) la question de son rapport à l'épistolaire»<sup>15</sup>. La posture des écrivains-épistoliers devient alors ambivalente. Ils n'hésitent pas à entremêler réflexions sur soi et sur le monde, dans leurs lettres, sans toutefois à aucun moment songer à les inclure comme telles, au sein d'une de leurs œuvres littéraires. Ils se cantonnent à les communiquer, à une échelle interpersonnelle, à leur(s) amant(e)s, à leur(s) ami(e)s, à leur famille, sous la forme d'un message volatile. Pourtant, nous remarquons une récurrence dans le temps de ces expressions mais également un soin particulier apporté à certains tours d'esprits, à une forme d'« affectation » de l'écrivain derrière sa plume... Tout cela nous montre que d'une certaine manière, les écrivains du XIX<sup>ème</sup> siècle se préparaient à ce que ces écrits s'inscrivent dans la grande Histoire. Pour preuve, nous évoquerons dans les prochains chapitres des traces effectives jointes au texte épistolaire (griffures du papier, lettres grossies...), comme pour prolonger une pensée qui se sentirait vite à l'étroit sur une telle surface. À partir de ce moment, «la lettre est comme dédoublée : émouvante notation de l'instant éphémère, elle est d'emblée aussi «lieu de mémoire», devient objet de collection»<sup>16</sup>. Qu'elles contiennent des confessions intimes ou non, elles n'en demeurent pas moins être des donations physiques, pourrions-nous même nous dire viscérales, de leur auteur. Elles sont en ce sens les nouvelles zones sensibles de la sphère littéraire, à laquelle elles s'incorporent à présent par touches successives.

La part la plus intime de l'être est ainsi passée au crible de la lettre... Au moment de l'écriture épistolaire, cela se traduit par une profusion de la note entourant l'acte épistolaire. Les écrivains n'hésitent alors pas à écrire un premier jet de leurs lettres sur des feuilles de

<sup>13</sup> CHARTIER, Roger. *La correspondance. Les usages de la lettre au XIX<sup>ème</sup> siècle*. Paris : Fayard, 1991, p.452.

<sup>14</sup> TISSERON, Serge. *L'intimité surexposée*. Paris : Hachette Littératures, 2002.

<sup>15</sup> DIAZ, José-Luis. « Le XIX<sup>e</sup> siècle devant les correspondances ». *Romantisme*, n°90, 1995, p.7. En ligne : [https://www.persee.fr/doc/roman\\_0048-8593\\_1995\\_num\\_25\\_90\\_3049](https://www.persee.fr/doc/roman_0048-8593_1995_num_25_90_3049)

<sup>16</sup> Ibid. p.10.

brouillons, avant de la soumettre à l'altérité du monde extérieur. Son ancrage dans le réel la soutient désormais toute entière ; la juxtaposition des mots jetés dans l'enveloppe ne font plus office de seule monnaie d'échange. La lettre doit désormais signifier un moment de l'artiste. La lettre reste toutefois une unité intégrée aux flux communicationnels de l'époque. Alain Pagès rappelle qu'elle est prise dans un « mouvement incessant (...) qui lui confère sa valeur d'échange », mais que désormais, « l'accumulation de l'hétéroclite » fait d'elle « le contraire d'une œuvre lisible d'abord dans son autonomie. »<sup>17</sup>. L'un des enjeux de notre étude sera de montrer comment restituer la lisibilité de la lettre, lorsqu'elle parviendra dans les mains des générations ultérieures. Car cette plongée au cœur de l'événement et d'une vie troublée entraînera des difficultés de contextualisation de l'information en son sein. C'est pourquoi les lettres et l'introspection qu'elles véhiculent devront nécessairement être intégrées au sein de constructions interprétatives qui les surpassent.

C'est comme si l'espace proprement intime, cultivé par l'écrivain, se refusait d'abord au destinataire puis au lecteur pour mieux se recréer ensuite... En cela la lettre est une forme d'écriture qui ne cesse de jouer entre le réel et l'imaginaire mais que l'on pourrait objectivement considérer comme la plus autobiographique de toutes. Nous pourrions parler à ce titre d'un « mouvement historique de "privatisation" de la subjectivité »<sup>18</sup>, qui expose à l'époque romantique et dont les correspondances seraient le premier fruit...

Il aura fallu attendre la fin du XVIII<sup>ème</sup> siècle mais surtout l'arrivée du XIX<sup>ème</sup> siècle pour assister à l'émergence du moi intime dans la lettre. Nous pouvons alors mettre en exergue la place grandissante dédiée à la confiance dans les correspondances d'écrivains, à partir du siècle des Lumières jusqu'au siècle romantique. Celle-ci va numériquement représenter 5% de la lettre entre 1700 et 1770, 9% entre 1770 et 1820, 22% entre 1820 et 1860<sup>19</sup>. L'aveu, dans ce qu'il peut avoir de « non-contrôlé » à première vue, est une forme d'expression prodigieuse pour une analyse sur le plan thématique, comme nous le verrons dans la suite de notre travail avec l'exemple de Gustave Flaubert.

L'omniprésence du pronom personnel « je » et l'expression des sentiments personnels chez les écrivains, dans la lignée romantique, sont donc une réponse stylistique aux échos du monde extérieur. « La transformation introspective de la lettre, précédée et accompagnée par la mutation de la lettre fictive en monodie, s'accorde avec le lieu commun romantique de l'incommunicabilité »<sup>20</sup>. Mais cela sera d'autant plus surprenant lorsque des écrivains, tels que Gustave Flaubert et Émile Zola, rattachés quant à eux au mouvement réaliste, vont se livrer ouvertement au sein de leurs correspondances. Leur optique romanesque consistait au contraire à s'effacer derrière leur description de la condition humaine. Ils adoptent dans leurs romans une écriture impersonnelle, visant à l'objectivité. Pourtant, ces écrivains seront parmi les plus impliqués, les plus radicaux, derrière la plume, lorsqu'il s'agira de se confier dans leurs lettres.

Ces éléments ont contribué à créer autour du XIX<sup>ème</sup> siècle l'aura d'une représentation sublimée de l'intimité, avant que sa littérisation ne devienne une valeur marchande quelques années plus tard. Les correspondances appartiennent aux Écritures du Moi, à notre patrimoine

<sup>17</sup> Ibid. p.110

<sup>18</sup> MARTIN, Philippe. *La correspondance : le mythe de l'individu dévoilé ?*. Presses universitaires de Louvain, 2015, p.56.

<sup>19</sup> GRASSI, Marie-Claire. « La correspondance comme discours privé au XVIII<sup>ème</sup> siècle ». Tiré de : BOSSIS, Mireille. *L'épistolarité à travers les siècles*. Stuttgart : Franz Steiner, 1990.

<sup>20</sup> SIMONET-TENANT, Françoise. « À la recherche des prémices d'une culture de l'intime ». Itinéraires, 2009 / 4, p.13. En ligne : <https://journals.openedition.org/itineraires/1466?lang=en>

autobiographique commun. Néanmoins, leur intrication dans le réel, en tant qu'objets d'échange, nous amène à les aborder selon une approche allant au-delà de la dimension littéraire, que l'introspection appelle pourtant. Les écrivains eux-mêmes furent les premiers à considérer que les révélations, prenant place dans leurs lettres, devaient persister dans le temps et non se détruire en même temps que leurs expéditeurs et destinataires. L'enjeu implicite est de remonter à la source de l'émotion de tout échange humain. Les correspondances d'écrivain nécessitent pour cela une exploration approfondie, sur le plan historique mais aussi linguistique, afin de dévoiler des ressources tangibles, des objets d'étude, sous la lisière de l'événement et du détail. Cette prise de conscience a donc été documentaire, avant d'être littéraire, et a été portée aussi bien par les bibliothécaires, les éditeurs ainsi que les chercheurs littéraires de l'époque. Nous ne quitterons jamais vraiment ces sphères de l'information et de la communication au cours de notre travail, car elles sont les premières façons d'aborder tout échange épistolaire, dans leur contexte opératif. Ces acteurs du XIX<sup>ème</sup> siècle se sont-ils dès lors dit que pour renouveler l'histoire des représentations et des « mentalités », il advenait de créer et d'enrichir une archive épistolaire ?

### I.1.3. Une conservation recrudescence mais contrainte des manuscrits de lettres

Nous avons évoqué l'importance acquise par les lettres chez les écrivains du XIX<sup>ème</sup> siècle, sous-tendue par leur statut singulier, inclassable, entre le document et l'œuvre artistique. Cette forme d'expression se laisse naturellement lire par celui qui la reçoit, à cet effet, mais la question va se poser en d'autres termes une fois la réception des lettres va être actée, lorsque celles-ci commenceront à s'amonceler sur le plan matériel. Nous avons montré que ces objets avaient tout lieu de vivre, en étant, dès leur origine les témoignages uniques d'une vie, d'une époque, d'un monde. Mais que peut-t-il advenir des lettres échangées entre un écrivain et son destinataire, dès lors que l'un des deux n'est plus ? Le contrat épistolaire implicite qui les liait est-t-il alors susceptible de trouver une résonance dans le monde extérieur, un prolongement de son cadre initialement interpersonnel ? Comment ces témoignages de l'instant peuvent-t-ils acquérir une dimension propre à être non seulement conservée mais plus encore, archivée ? C'est toutes ces questions que les institutions patrimoniales ont été en droit de se poser en approchant les premiers fonds de correspondances du XIX<sup>ème</sup> siècle. Car les lecteurs d'un écrivain souhaitaient découvrir ces lettres, et pour cela y accéder par n'importe quel moyen. Le rassemblement d'un maximum de lettres au sein d'un même lieu apparaissait alors comme l'unique porte d'entrée, pour que le public puisse découvrir un ensemble sélectionné et exhaustif d'échanges épistolaires, rattachés à un écrivain. Mais pour parvenir à cette fin, les correspondances d'écrivain faisaient tout d'abord face à des barrières spatio-temporelles telles que la dispersion géographique des lettres. Ces documents pouvaient potentiellement être dans les mains de tout-un-chacun, suite à des héritages familiaux ou à des acquisitions privées, ce qui ne facilitait pas leur assemblage à première vue. Au-delà de cette quête du document dans l'espace, la question du volume de s fonds de lettres commençait également à se poser du côté des institutions de conservation. Comment pouvaient-elles connaître le nombre de lettres d'un écrivain qu'il leur fallait recueillir et classer ?

Ces questions plus spécifiques, en rapport à l'épistolaire, étaient en outre à ne pas négliger. Car la pérennisation des correspondances d'un écrivain impliquait étymologiquement de recréer un dialogue à plusieurs voix entre l'écrivain et ses destinataires. Dans « correspondre », se dessine

d'une certaine façon le terme « répondre ». Les bibliothèques ont très rapidement souhaité archiver non seulement les lettres rédigées par l'écrivain, mais également les lettres envoyées à l'écrivain. Nous nommons respectivement ces deux moments de l'échange épistolaire, inséparables en soi, en tant que correspondance active et que correspondance passive. Le problème qui se posera aux institutions patrimoniales sera alors de l'ordre de la synchronicité, dans les réceptions et catalogages de ces lettres. Comme nous le verrons dans notre second chapitre avec Gustave Flaubert et Émile Zola, les correspondances impliquent des échanges épistolaires entre écrivains, et il peut alors vite devenir complexe de démêler les lettres expédiées ou reçues par ces grands noms respectifs. Cette situation entraîne un manque de lisibilité dans l'archivage des correspondances d'écrivains, ce qui pouvait entraîner certains établissements à se construire des fonds de correspondances mal définis, où plusieurs noms d'auteurs interféraient dans les catalogages. Rappelons pour finir que les contraintes esquissées pouvaient tout bonnement être de nature extérieure à ces institutions culturelles, propres au matériau de la lettre. Toute correspondance est en effet le fruit « des aléas de la transmission, du choix des éditeurs et, parfois, des nécessités matérielles. »<sup>21</sup>.

Mais une fois dépassés ces quelques écueils, proprement archivistiques, nous nous apercevons que ces fonds de lettres, au-delà d'être intrinsèquement riches, arrivaient au sein de ces institutions, la plupart du temps, assez complets. Nous avons montré les intérêts personnels que des écrivains du XIX<sup>ème</sup> siècle pouvaient avoir à ne pas se départir de ces documents les concernant au plus haut niveau, pour leur postérité. Ceci en dépit de l'éphémérité de leur support et de leur apparente rapidité d'exécution. Nous remarquons à présent que la transmission de leurs correspondances aux institutions, ou au sein de leur cercle familial dans un premier temps, s'est déroulé en très bons termes, et selon les pleines volontés de l'auteur. C'est sans doute l'une des caractéristiques les plus frappantes de l'acte épistolaire au XIX<sup>ème</sup> siècle, que le choix consenti et mûri du legs de la correspondance par l'artiste.

La présence significative des correspondances d'écrivain du XIX<sup>ème</sup> siècle au sein des bibliothèques municipales de France résulte pleinement de la libéralisation du legs, opéré le plus souvent par les écrivains eux mêmes, via leur testament. Ce geste « patrimonial » se généralise à cette période, en premier lieu pour des raisons culturelles et artistiques, que sont l'émergence de la notion moderne d'auteur en tant que sujet de l'écriture, ainsi que la valorisation du génie et de l'originalité. Nous sommes en effet, rappelons-le, à l'ère du positivisme, où tout élément est censé pouvoir s'évaluer en se basant sur l'expérience seule. Les écrivains vont considérer que les traces laissées dans leurs lettres sont tout sauf immatérielles, qu'elles sont proprement empiriques et qu'elles permettent en cela d'accéder plus facilement à leur souffle créateur.

« Le XIX<sup>e</sup> est le siècle de la consécration de l'auteur et du manuscrit, d'où une certaine abondance au sein des collections publiques et privées. La prise de conscience de l'intérêt du manuscrit monte en puissance, chez les auteurs romantiques et dans l'opinion, se traduisant par une dynamique de sauvegarde et de conservation. »<sup>22</sup>

Les legs interviennent également de plus en plus souvent du côté des descendants de l'écrivain. La motivation d'un tel geste, effectué à l'échelle familiale, était alors de faire perdurer l'œuvre de l'artiste, matériellement, puis culturellement. Avec la connaissance des expositions et autres diffusions de documents à grande échelle opérées par les bibliothèques à

<sup>21</sup> MONTENOT, Jean. « Ce que révèle la correspondance des écrivains ». L'Express, 01/07/2007. En ligne : [https://www.lexpress.fr/culture/livre/ce-que-revele-la-correspondance-des-ecrivains\\_812403.html](https://www.lexpress.fr/culture/livre/ce-que-revele-la-correspondance-des-ecrivains_812403.html)

<sup>22</sup> PINÇON, Juliette. « Les archives des écrivains, leur place en bibliothèque ». Enssib : Mémoire d'étude DCB, 2017, p.23. En ligne : <https://www.enssib.fr/bibliotheque-numerique/documents/67311-les-archives-des-ecrivains-leur-place-en-bibliotheque.pdf>

partir du XX<sup>ème</sup> siècle, les familles souhaitaient pouvoir partager dans l'immédiat ces fonds vivants de documents inédits. Nous verrons dans nos prochaines parties comment les interfaces numériques vont accroître l'attractivité d'une mise à disposition des correspondances d'écrivain, pour une mise en lumière graduelle de leur contenu.

On recense aujourd'hui plus de 150 auteurs du XIX<sup>ème</sup> siècle dont les manuscrits sont conservés dans les bibliothèques publiques françaises<sup>23</sup>. Au sein de ces fonds, les types d'archives manuscrites les plus représentées ne sont autre que les correspondances... Parmi les auteurs les plus connus de l'époque, nous citerons la présence des lettres Stendhal à la BMC de Grenoble, de Lamartine à la BMC d'Angers et Mâcon, ainsi qu'à la BNF, ou encore de Maupassant à la BMC de Rouen, à la BhVP et à la BNF, ou encore Jules Verne à la BMC d'Amiens et de Nantes. Nous évoquerons plus en détail les lieux de conservation des lettres de Gustave Flaubert et d'Émile Zola, qui sont les écrivains au centre de notre travail, dans les prochains chapitres. Il nous était important de souligner, initialement, dans quelle mesure il les grands auteurs du XIX<sup>ème</sup> siècle considèrent comme essentielle leur rapport à l'avant-texte dans leurs œuvres, qui sera scellé par leur mise à disposition publique. Avaient-ils d'ores et déjà dans l'idée de faire l'objet de collections plus spécifiques, plus pertinentes, grâce à la dation de ce pan particulièrement intime de leur vie d'écrivain ?

Nous avons évoqué comment des fonds de correspondances d'écrivains du XIX<sup>ème</sup> siècle pouvaient formellement se constituer. Les questions que nous nous poserons par la suite concerneront l'exploitation de tels fonds d'archives, qui incluent donc des modalités de conservation et de diffusion très spécifiques. Pour cela, nous devons nous concentrer sur un nombre limité d'écrivains, présentant des ressemblances et des contradictions éloquentes dans leur approche du matériau épistolaire. Nous avons choisi d'étudier les correspondances de Gustave Flaubert et d'Émile Zola, dans un développement « en miroir » de notre propos. Cette façon de progresser, au sein de tant de milliers de lettres archivées, devra nous montrer qu'aujourd'hui encore, « il reste beaucoup à faire dans le domaine des correspondances pour la littérature française du XIX<sup>ème</sup> siècle »<sup>24</sup>. Nous nous doutons bien qu'il ne s'agira pas de réécrire ces lettres, mais plutôt de les écrire à nouveau, avec le langage et les outils de l'homme qui sonde en permanence son passé, à la lumière d'autres grands hommes...

## I.2. Les correspondances de Flaubert et de Zola

Comme nous l'avons précédemment vu, les écrivains du XIX<sup>ème</sup> siècle sont les premiers à s'être posé la question du rapport à leurs lettres et à le coucher, paradoxalement, sur le papier. Nous avons relaté, dans notre premier chapitre, de l'importance que l'introspection a pris dans leurs lettres. Néanmoins, il pouvait parfois s'avérer délicat pour les écrivains de ratifier le fait qu'ils se livraient si profondément alors, et que ces mots pourraient connaître une trajectoire insoupçonnée, après leur envoi à leur destinataire pressenti. Gustave Flaubert et Émile Zola vont être des personnages tout à fait fascinants à étudier sous ce rapport particulier de la trace écrite dans la lettre et de son devenir mystérieux.

<sup>23</sup> Ibid. p.23.

<sup>24</sup> BRUNEAU, Jean (Ed). *Correspondance de Gustave Flaubert : tome II*. Paris : Bibliothèque La Pléiade, 1980, p.10.

### I.2.1. Ce que nous disent ces écrivains et leurs destinataires, dans leurs correspondances

Dès la fin de la rédaction de « Madame Bovary » en 1857, Gustave Flaubert a voulu bannir de sa vie l'écriture de soi. « C'est un de mes principes, qu'il ne faut pas s'écrire »<sup>25</sup>. L'homme y revient pourtant sans cesse, dans ses lettres, n'hésitant pas à orienter les sujets de conversation avec ses destinataires autour de sa personne. Pour Flaubert, « la lettre est un relais essentiel dans la constitution du sujet (...) lieu d'une véritable ontogénèse »<sup>26</sup>. Nous voyons là qu'il ne s'agit pas simplement pour lui de simplement communiquer ce qui se passe dans son existence. Il y a une volonté sous-jacente de faire de la lettre le reflet de lui-même, de se représenter non plus tellement comme un écrivain mais comme un personnage immuable dans ce monde.

« Plus que laboratoire de valeurs, la lettre apparaît d'abord et essentiellement comme le laboratoire d'une valeur : celle de la congruence des façons de dire avec la posture que l'épistolier adopte, comme être humain et comme être social. Venant à inspirer le sentiment d'un certain bonheur de la parole, cette congruence peut promouvoir la lettre au rang de petite œuvre »<sup>27</sup>.

Il était en effet rare avant le XIX<sup>ème</sup> siècle, hormis quelques exceptions comme précédemment évoqués avec Madame de Sévigné, qu'une correspondance se lise pour elle-même. C'est à partir du XIX<sup>ème</sup> siècle que la correspondance va se muer en un véritable « laboratoire esthétique », où il deviendra permis en son sein d'« d'énoncer des sentences théoriques qui sont en avance sur la pratique littéraire »<sup>28</sup>. Dans ses lettres, l'épanchement à la fois psychologique et intellectuel de Gustave Flaubert retient toute l'attention du lecteur. Cette libéralité de la parole donne bien plus qu'un simple aperçu de l'écrivain sur le plan humain et stylistique. Certes, la correspondance n'est pas un territoire entièrement personnel et heuristique, comme pouvaient l'être le journal ou les Mémoires, car la présence d'un destinataire de la lettre supposera une altérité directe dans le jugement de ce qui s'y dira. Il n'en demeure pas moins qu'au moment de l'écriture épistolaire, les paroles de l'écrivain jaillissent d'un univers virtuel au sein duquel l'écrivain confronte son expérience dans des postulations identitaires multiples. Le temps n'a-t-il plus de prise sur ces évanescences de l'esprit *flaubertien*. Il pourrait être intéressant de rapprocher les lettres datées de l'écrivain, ainsi que les biographies lui étant consacrées *post mortem*, pour évaluer dans quel contexte celui-ci était le plus a-même dans ses lettres de duper parfois les autres, tout autant que lui-même, sur son propre état général.

De par son format même, la lettre prépare le terrain du témoignage circonstancié. C'est ce que l'on pourrait appeler la *praxis* de l'artiste, à laquelle Flaubert attribue un néologisme de son cru : « l'artistisme ». De ces nombreuses interrogations sur soi et sur le monde, et, surtout,

<sup>25</sup> Lettre de Flaubert à Mademoiselle Leroyer de Chantepie, 18 mars 1857, *Correspondance, t. II*, éd. établie par Jean Bruneau, Paris, Gallimard, coll. « Bibliothèque de la Pléiade », 1980, p. 691.

<sup>26</sup> DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Ecriture », 2002, p.61.

<sup>27</sup> SCHWEIGER, Amélie. « Flaubert en toutes lettres. L'écriture épistolaire dans la correspondance et dans l'œuvre ». Mont-Saint-Aignan : Presses universitaires de Rouen et du Havre, 2012, p.9. En ligne : <https://books.openedition.org/purh/5746>

<sup>28</sup> SIMONET-TENANT, Françoise. « À la recherche des prémices d'une culture de l'intime ». Itinéraires, 2009 / 4, pp.39-62. En ligne : <https://journals.openedition.org/itineraires/1466?lang=en>

de leur partage immédiat dans les lettres, vont découler une posture singulière de l'écrivain. Flaubert adopta définitivement la sienne à partir de son voyage en Orient, avec son compagnon de route et ami Maxime Du Camp, durant les années 1849 à 1851. Il s'exprime en ces termes : « Il n'y a rien de plus beau que ces femmes vous appelant. Si j'eusse baisé, une autre image serait venue par-dessus celle-là et en aurait atténué la splendeur. (...) Je n'ai pas toujours mené avec moi un *artistisme* si stoïque »<sup>29</sup> Comme tout un symbole, ce principe est énoncé, pris dans le remous de considérations, dira-t-on, primaires. Pourtant, l'usage que Flaubert a fait de ce terme postule ici une attitude, où le geste artistique serait tout entier fondé sur une forme de démenti, voire de déni de l'autobiographique. Et c'est l'écriture épistolaire qui révèle ce moment si caractéristique de l'invention de l'écrivain par lui-même. Ainsi, « Écrire des lettres, pour Flaubert, ce serait faire l'épreuve et la preuve de son être écrivain »<sup>30</sup>.

Son statut d'épistolier prodigue s'inscrit de fait dans une démarche artistique totale, se suffisant pour ainsi dire à elle-même. Pourtant, la lettre d'un écrivain reste toujours affiliée, par certains aspects, à l'œuvre littéraire de Flaubert. C'est ainsi qu'« Une lettre peut apporter une information de première main, décisive pour dater ou pour interpréter, sans constituer pour autant un document de genèse à proprement parler »<sup>31</sup>.

Flaubert laisse également libre cours à son opinion personnelle dans ses lettres. C'est semble-t-il sans retenue et sans concession qu'il la proclame à ses correspondants. La nature strictement privée de cette écriture épistolaire favorise à ce titre la sincérité de l'expression, jusqu'à l'outrance du propos, parfois. La lettre engendre une fictionnalisation de l'existence des écrivains. Toutefois, la liste des sujets qui feront inévitablement retomber le propos épistolaire dans le réel ne manquent pas. Nous pouvons par exemple évoquer le rapport de l'écrivain à la politique. Citons une nouvelle fois Flaubert dans ses lettres : « Je ne sais si la forme nouvelle du gouvernement et l'état social qui en résultera sera favorable à l'Art. »<sup>32</sup>. L'homme est un libéral conservateur, qui s'en prend indifféremment à droite et à gauche à toute idéologie qui menace l'individu dans sa liberté d'entreprendre, de posséder, de penser et de s'exprimer. Pourtant, la *Correspondance* montre qu'il a suivi de près les événements, et qu'il s'est passionné pour l'avenir de la république. Il fréquente la mondanité parisienne et échange d'ailleurs à ce sujet avec un certain... Émile Zola. La concordance artistique entre les deux hommes leur apparaissait assez limpide ; Zola essayant alors d'amener les conceptions réalistes de Flaubert plus loin, avec ses romans dits « naturalistes ». Les échanges redoublés entre les deux hommes dans leurs lettres montraient alors que Flaubert cherchait à savoir quelles tournures pouvaient prendre les choses.

Car derrière toutes ces apparentes contradictions de l'homme se trame, une fois n'est pas coutume, la réalité. Elle nous enseigne qu'au XIX<sup>ème</sup> siècle, la figure de l'écrivain épistolier est avant tout celle d'un intellectuel explorant sa propre pratique. Seules les lettres en définitive étaient en mesure de nous représenter l'écrivain à la tâche. « À s'écrire « en l'Art », la correspondance s'est convertie en discours de la méthode »<sup>33</sup>. Nous constatons ainsi un glissement vers « une

<sup>29</sup> Lettre de Flaubert à Louis Bouilhet, envoyée à Syène, le 13 mars 1850.

<sup>30</sup> GRASSI, Marie-Claire. *Lire l'épistolaire*. Paris : Armand Colin, 1998,

<sup>31</sup> DE BIASI, Pierre-Marc. «Correspondance et genèse : Indice épistolaire et lettre de travail : le cas Flaubert ». *Genèse et correspondance*. Paris : Éditions des Archives contemporaines, 2012.

<sup>32</sup> Lettre de Flaubert à Louise Colet, Croisset, mars 1848. BRUNEAU Jean (ed.). *Correspondance de Flaubert, tome I : Janvier 1830-Mai 1851*. Paris : Gallimard (Bibliothèque de la Pléiade), 1980, p.492.

<sup>33</sup> SCHWEIGER, Amélie. « Flaubert en toutes lettres. L'écriture épistolaire dans la correspondance et dans l'œuvre ». Mont-Saint-Aignan : Presses universitaires de Rouen et du Havre, 2012, p.92. En ligne : <https://books.openedition.org/purh/5746>

communication horizontale – inféodée à la pratique »<sup>34</sup>. A t-elle pour autant significatif un nivellement de l'expression, une uniformisation de la pensée ?

Car il est curieux de remarquer que le nombre de biographies consacrées à Flaubert est assez restreint, par rapport à d'autres écrivains, et surtout, par rapport à l'aura qui entoure aujourd'hui l'écrivain dans la sphère littéraire. Sans doute pourrions-nous postuler que c'est l'achèvement et la restitution publique de cette *Correspondance* qui a rendu la tâche des biographes on ne peut plus compliquée ; l'écrivain s'étant déjà écrit lui-même, en définitive, mieux que quiconque.

Tous ces éléments sont au fondement de synergies historiques essentielles, que l'étude de la correspondance d'un autre écrivain, contemporain à Flaubert, nous révélera en la personne d'Émile Zola. Flaubert était tourné vers son intériorité et ne semblait pas résolu à se poser, dans ses lettres, en tant que témoin d'une société qu'il n'affectionnait pas particulièrement. Ce manque d'intérêt assumé pour son époque est une posture aux antipodes de celle qu'adopta le chef de file du naturalisme, notamment dans ses lettres...

Dès lors que l'on plonge dans les correspondances de l'auteur des *Rougon-Macquart*, se dessinent les grandes facettes d'un milieu social, de la sociabilité des gens de lettres, peintres, journalistes de la seconde moitié du XIX<sup>ème</sup> siècle. C'est également un microcosme de la vie politique complexe de la France de cette époque qui nous apparaît, à travers de rares échanges épistolaires avec Raymond Poincaré ou Georges Clémenceau. La découverte du cheminement social de l'écrivain se poursuit en lisant ses lettres adressées à des personnalités du théâtre, des chroniqueurs, des éditeurs. C'est souvent par la médiation de la correspondance que l'écrivain gère à distance ses livres et leur réception. Vecteurs et témoins de cette activité médiatique, ces échanges ont beaucoup à nous apprendre sur la « circulation littéraire » au XIX<sup>ème</sup> siècle. N'oublions pas les lettres échangées avec ses pairs, artistes venant de tous horizons, tels que Paul Cézanne dont il fût le grand ami (beaucoup de lettres furent malheureusement égarés entre les deux hommes), en passant par le compositeur Alfred Bruneau, jusqu'au peintre-graveur Fernand Desmoulins. Les lettres échangées avec le photographe Nadar, où il y est sujet de séances de pose, sont particulièrement intéressantes pour l'histoire de la mise en scène de Zola en tant que « figure » d'écrivain.. L'épistolaire fait alors naître une nouvelle forme de discours sur l'art et sur les artistes, avec une terminologie qui lui est propre. Mais la pierre angulaire de cette mosaïque relationnelle est bien sûr constituée des lettres adressés par Zola à ses proches, se montrant tour à tour homme, mari, amant, et père avant d'être un écrivain plein d'ambition. Ces lettres se caractérisent ainsi par un emploi élevé de noms propres et de chiffres qui assurent leur ancrage dans le temps, l'espace, et offrent une immersion dans la quotidienneté d'une époque. C'est comme si la médiation des destinataires, plus ou moins proches de l'écrivain, confortait dans le même temps ce contrat d'artiste passé avant tout avec avec lui-même... La correspondance est souvent le lieu unique du discours critique et théorique. Même si l'assertion est surtout pertinente pour Gustave Flaubert, elle l'est également pour Émile Zola, et notamment dans les moments où l'intensité des événements politiques se bousculent... Nous avons évoqué l'implication de l'homme en société, d'où résultera un matériau neuf, proprement expérimental, que Flaubert ne parvint par exemple pas à obtenir dans ses lettres. « La correspondance est la meilleure trace du choc entre pensée et histoire » et permet de rétrospectivement « se rapprocher d'une généalogie réelle des productions intellectuelles »<sup>35</sup> d'une époque.

Car si l'aventure épistolaire de Zola est devenue unique en son genre, c'est sans doute grâce aux lettres reçues par l'écrivain qui parachevèrent cette plongée dans le réel, couronnée par la célèbre

<sup>34</sup> Ibid. p.125.

<sup>35</sup> PROCHASSON, Christophe. « Les correspondances : sources et lieux de mémoire de l'histoire intellectuelle ». *Varia*, 8 / 1991, p.3. En ligne : <https://journals.openedition.org/ccrh/2824>

affaire Dreyfus. Émile Zola est engagé depuis plus de 35 années dans les affaires publiques françaises, avant d'asséner son « J'accuse », ce 13 janvier 1898, contre les accusations de trahison patriotique faites aux colonels Dreyfus sous la III<sup>ème</sup> République. Une prise de position idéologique qui ne s'est donc pas établie en un jour, comme peuvent en témoigner les échanges épistolaires entre l'écrivain et les journaux de l'époque tels que « Le Figaro » ou encore « L'aurore », dans lequel son article fût d'ailleurs publié. Nous sommes alors dans le cœur de l'actualité, précédemment évoquée.

Cette correspondance active, dans tous les sens de son acception, montre un homme qui se bat pour ses opinions. C'est « un phénomène rarissime, peut-être unique dans notre histoire : la parole d'un grand écrivain s'est transformée en un levier d'opinions qui a fait exploser les pesanteurs »<sup>36</sup>. L'écrivain dispose alors, parmi ses confrères, de l'audience publique la plus forte, de la verve polémique la plus prononcée ainsi que d'une expérience dans la confrontations de ses idées. La correspondance passive intervient alors, en dressant un panel des réactions dans le monde face à ce basculement historique sans précédent. Les lettres des lecteurs de Zola, adressées à l'écrivain, proviennent aussi bien d'inconnus que de personnalités influentes. Des lettres venues de tous les continents lui arrivent par dizaines, chaque jour, alors qu'il est en exil en Angleterre. Nous nous trouvons ici face à une mobilité à grande échelle mais unilatérale de la lettre, aux antipodes de la République des Lettres du XVIII<sup>ème</sup> siècle et ses échanges diffus dans toute l'Europe notamment<sup>37</sup>. Elles sont un formidable document sur la réception du naturalisme et de l'affaire Dreyfus dans le monde entier, qui offrent aux spécialistes la possibilité de recherches nouvelles, devant nous donner une image nouvelle, plus précise, . Nous avons basculé dans le cadre des transferts culturels ouverts, à propos desquels cette correspondance offrait déjà des perspectives novatrices. Les confidences ne sont définitivement plus feintes au sein des lettres, qui sont une « traduction de l'état d'âme de l'épistolier à un moment précis et appréciation de celui du destinataire au même moment, avec pour conséquence, tantôt un effort de conciliation et tantôt un essai d'intimidation »<sup>38</sup>... L'étude de la correspondance de Zola s'inscrit tout à fait dans cette dynamique avec une confrontation exacerbée des points de vue, qui va permettre de restituer la dynamique du présent « multiple » de l'écrivain-épistolier. Nous constatons ici que « la glissance temporelle de l'épistolaire est liée à ses caractéristiques de métonymie et de présence simultanée »<sup>39</sup>. Certes, Zola est l'une des personnalités médiatiques les plus interviewées de son temps, mais il est aussi l'un des plus sollicités au-delà de ce cadre conventionnel. Sa correspondance le fait ainsi s'engager, peut-être malgré lui, dans une « stratégie de visibilisation polémique », d'un homme qui « construit sa réputation à coup de buttoirs »<sup>40</sup>. Et d'encrier.

La correspondance « zolienne » trahit toutefois derrière tous ces événements, à plusieurs reprises, la tentation d'une écriture retirée, en solitaire, sur laquelle nous reviendrons prochainement. Car tout comme Gustave Flaubert, Émile Zola est conscient que ces échanges épistolaires perdureront, et devront refléter en conséquence sa nature profonde, intime...

« Pour Zola, comme pour Balzac ou Flaubert, elle [la lettre] témoigne de la quête inlassable du document qui parcourt la dynamique de l'œuvre »<sup>41</sup>. Il nous importe à présent de retracer

<sup>36</sup> MITTERAND, Henri. *Zola, tel qu'en lui-même*. Paris : Presses Universitaires de France. 2009, p.155-156.

<sup>37</sup> Voir le schéma de l'Annexe N°2 (p.95)

<sup>38</sup> DUCHÊNE, Roger, *Comme une lettre à la poste*. Paris : Fayard, 2006, pp.222-223.

<sup>39</sup> STANLEY, Liz. « The Epistolarium : On Theorizing Letters and Correspondences », *Auto/Biography*, n°12, 2004.

<sup>40</sup> GIRAUD, Frédérique. « S'imposer parmi ses pairs. Le travail réputationnel d'Émile Zola à l'assaut de la sphère restreinte du champ littéraire ». *Terrains&Travaux* N°26, 2015/1, p.25. En ligne : <https://www.cairn.info/revue-terrains-et-travaux-2015-1-page-23.htm>

l'émergence de cette obsession documentaire sous-jacente à l'écriture épistolaire au XIX<sup>ème</sup> siècle, en dévoilant le rapport matériel que nos deux écrivains établissaient avec leurs lettres...

## I.2.2. Vers une approche génétique des correspondances

Nous avons précédemment montré que l'approche à adopter face à un fonds épistolaire était pratique, avec les enjeux d'une classification des lettres d'écrivains faisant sens. Elle sera dans un second temps génétique, comme nous allons le montrer. Cette méthode d'analyse des écrits, notamment littéraires, se propose de restituer le processus de création du texte, à travers notamment ses brouillons et autres ébauches manuscrites. Le plus étonnant, avec des écrivains tels qu'Émile et Gustave Flaubert, est sans doute le fait qu'ils aient posé les premiers jalons de ce que devrait être l'archivage de leurs lettres dans les siècles futurs.

La lettre n'est donc pas officiellement un manuscrit de travail de l'écrivain. Comme nous l'avons vu, les circonstances, les normes et les conventions sociales derrière sa production restent prégnantes, à son insu souvent. En cela, elle prête toujours plus ou moins sujet à caution, et reste son propre brouillon. Il sera donc difficile pour elle de trouver sa place au sein du dossier génétique consacré à un auteur. La première empreinte à laquelle vont s'intéresser les généticiens face à une lettre d'écrivain va être la signature, à des fins évidentes d'identification et de certification du document au sein du fonds de sa correspondance. Mais au-delà de cette trace caractéristique, toutes les formes de l'écrit laissées sur une lettre vont constituer une source de recherche. Elles indiquent, en un coup d'œil, si l'écrivain cherchait alors ses mots, en raturant, s'il souhaitait donner à un mot une importance toute particulière en le faisant se distinguer des autres dans le texte...

« Le généticien s'interroge sur les choix continuels d'un écrivain qui corrige, rature, évalue ce qu'il vient d'écrire, hésite entre plusieurs possibilités. Le texte est dans beaucoup de cas le fruit de tâtonnements et de réajustements successifs et il n'y a rien de magique dans cette opération souvent laborieuse »<sup>42</sup>.

Les études « génétiques » des œuvres de Flaubert, conduites notamment à l'Institut des textes et manuscrits modernes du CNRS, « ont progressivement permis de donner tout son sens à ce qui pouvait paraître le paradoxe d'une œuvre cherchant la phrase « inchangeable », en révélant « l'infini travail de la rature et de la reprise »<sup>43</sup> propre à l'écrivain. Lorsque les lettres manuscrites sont dévoilées au public, les diverses phases d'écriture de Flaubert sont d'autant plus percées à jour. L'habitude du style d'un écrivain, la connaissance de ses tics de langage mais aussi de son imaginaire se combinent comme un système de possibles et permettent de s'orienter plus vite vers l'interprétation de ses productions. Cette perspective sémiotique peut aller tout à fait au-delà du tracé de la plume, jusqu'à prendre en compte la composition de la page d'une lettre pour ce qu'elle est : une trace unique de l'esprit. Flaubert

<sup>41</sup> PAGES, Alain. « Émile Zola ou la transparence épistolaire ». *Genèse et correspondance*. Paris : Éditions des Archives contemporaines, 2012.

<sup>42</sup> SIMONET-TENANT, Françoise. « La critique génétique : définition, intérêts, limites ». *Génétique*, © Publications numériques du CÉRÉDI, « Ressources », 2019, p.7. En ligne : <http://publis-shs.univ-rouen.fr/ceredi/index.php?id=603>

<sup>43</sup> NEEFS, Jacques. « Les modernités de Gustave Flaubert, écrivain » *Europe : revue littéraire mensuelle*, 2018, p.4

l'avouait : « Jamais je ne jette aucun papier. C'est de ma part une manie »<sup>44</sup>. L'écrivain pousse cette « manie », ce fétichisme, jusqu'à l'identification de sa personne au matériau de son travail. À travers les lettres qu'il rédige, il devient l'« homme-plume », se rêvant avec un corps équivalent à la masse des feuilles qu'il a noircies. Comment en est-il arriver à ces images mentales, face à de simples bouts de papier ? Nous pouvons essayer d'y répondre en partant une nouvelle fois d'un point de vue sociétal, avec la naissance de l'éditeur au XIX<sup>ème</sup> siècle, et par la même du recueil de Correspondance. C'est à l'époque de l'écrivain réaliste que cet objet, ne s'étant pas encore réellement matérialisé, a joué un rôle capital dans l'approche de leur propre fonds de lettres. Cette idée de réunir toutes les lettres d'un même artiste au sein d'un même document a germé dans l'esprit de Flaubert ou de Zola. Elle leur fournissait alors « une trace visible (...) un contact, (...) l'illusion d'une résurrection possible par *l'épuisement du détail* »<sup>45</sup> dans ce pan inconnu de leur œuvre.

Revenons alors à la phrase de Flaubert: « Je suis un homme-plume. Je sens par elle, à cause d'elle, par rapport à elle et beaucoup plus avec elle [...]. Je crèverai obscur ou illustre, manuscrit ou imprimé. »<sup>46</sup>. Nous voyons qu'au-delà des projections mentales de l'écrivain, l'archivage de toutes les traces écrites de la lettre, à l'échelle personnelle, obéissait déjà à un impératif éthique et esthétique. La dernière phrase de la citation nous montre que Flaubert entendait conserver les preuves du travail artisanal de son style en premier lieu. Peu lui importait si ces dernières seraient soumises ou non à la standardisation des caractères imprimés de la « littérature industrielle ». Cette expression, chère à un autre écrivain du 19<sup>ème</sup> siècle en la personne de Saint-Beuve, nous rappelle une nouvelle fois rôle naissant de l'éditeur. Celui-ci allait alors être le premier généticien de son temps, extérieur à l'écrivain, en matière épistolaire, au sein d'une société où espace public et littérature se conventionnaient réciproquement. Dès lors, les scripteurs n'étaient plus recrutés exclusivement dans les mêmes groupes sociaux, comme au XVIII<sup>ème</sup> siècle. De cette libéralisation de la pratique éditoriale a découlé la divulgation progressive du geste de l'écrit épistolaire, qui s'assimilait dorénavant dans l'inconscient populaire à une forme de l'écrit allant tout à fait de soi, chez l'écrivain.

Les manuscrits des lettres valent donc en tant que pièces justificatives de la lenteur d'exécution et que témoins d'une conception neuve du style qui doit se conquérir mot à mot. Une telle fonction peut d'ailleurs étonner chez cet « anti-moderne » qu'était Flaubert . Elle dévoile la recherche de l'écriture à laquelle se livre l'écrivain et marque une discontinuité de la lettre avec le pan littéraire de l'œuvre, figé à partir du moment des publications. En nous appuyant sur ces éléments récurrents au sein de sa *Correspondance*, nous pouvons dire aux côtés de Roland Barthes que Gustave Flaubert est celui qui a « constitué définitivement la littérature en objet »<sup>47</sup>. Avant de se renvoyer entre elles, ses lettres se parcourent. Nous pouvons dire que c'est son œuvre épistolaire qui fait définitivement de Flaubert l'écrivain de l'intransitivité.

Ce statut est l'une des raisons qui expliquera le souhait de l'écrivain de conserver en versions manuscrites, les lettres qu'il recevra de ses destinataires, ou celles qu'il venait de rédiger, copiées sur des brouillons avant leur envoi. Il se constitue alors sa propre collection épistolaire. Les auteurs du 19<sup>ème</sup> siècle, parmi lesquels en premier lieu Flaubert avaient une «tendance à la thésaurisation de ces «écrits du jour» conçus comme un gisement à exploiter le moment venu»<sup>48</sup>. Nous reviendrons à ces potentialités incluses des lettres conservées à portée

<sup>44</sup> Lettre de Flaubert à Louise Colet, envoyée le 26 avril 1853.

<sup>45</sup> PAGES, Alain. « Stratégies textuelles : la lettre à la fin du XIX<sup>ème</sup> siècle ». *Littérature*, n°31, 1978, pp.109-110. Accessible en ligne : [https://www.persee.fr/doc/litt\\_0047-4800\\_1978\\_num\\_31\\_3\\_1168](https://www.persee.fr/doc/litt_0047-4800_1978_num_31_3_1168)

<sup>46</sup> BRUNEAU, Jean (Ed). *Correspondance de Gustave Flaubert : tome II*. Paris : Bibliothèque La Pléiade, 1980, p.42.

<sup>47</sup> BARTHES, Roland. *Le degré zéro de l'écriture*. Paris : Seuil, 1953, p.4.

<sup>48</sup> DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Écriture », 2002, p.56

de main de l'écrivain. Mais il nous faut avant cela établir le rapport matériel développé entre un écrivain et ses lettres, au fondement de toute approche génétique de l'épistolaire. Le fait que Flaubert ne définisse son propre fonds de lettres manuscrites en tant qu'*archives* qu'à une seule reprise<sup>49</sup>, au sein de sa correspondance, et ce dans les dernières années de sa vie, est significatif. Et d'autant plus pour un homme ordonné comme il pouvait l'être dans ses phases de travail rédactionnel. Les spécialistes de Flaubert, tels que Pierre-Marc de Biasi, insistent à ce titre sur l'aspect inachevé et sur l'esthétique du « non-finito », qui prédominent dans l'œuvre de l'écrivain. Flaubert avait pour habitude de laisser certains passages de ses romans nous renvoyer leur avant-texte, et a fortiori vers les lettres, en considérant que son œuvre ne serait achevée qu'à sa mort. D'où sa réticence à aborder le terme d'*archives* de son vivant...

« Véhicule du secret, la lettre en représente l'apparition et le retrait, mime le jeu de la dissimulation par la précarité de son existence textuelle. La main qui censure, au XIX<sup>ème</sup> siècle, sait que la suppression qu'elle opère s'inscrit dans la forme d'un texte déjà marqué par l'inachevé et le discontinu. »<sup>50</sup>.

Au regard de sa correspondance et de la mémoire textuelle qu'elle véhicule, Flaubert apparaît donc également comme l'écrivain de l'immobilité et de la dissimulation. Pour lui, les traces écrites sont sans cesse à conquérir. C'est d'ailleurs dans cette optique qu'à partir du XIX<sup>ème</sup> siècle, les écrivains vont instaurer un « contrat tacite de négligence, répudiant a priori de leurs correspondances l'effort, intellectuel aussi bien que calligraphique »<sup>51</sup>. Ces points que nous avons soulevé, concernant le rapport de Flaubert à sa correspondance, dénotent là encore dans leurs grandes lignes, avec l'approche épistolaire adoptée Émile Zola. Néanmoins, le généticien qui se retrouvera face à ces deux fonds respectifs de correspondances serait en mesure de les appréhender d'une façon équivalente, car elles présentent également certaines similitudes dans leur méta-langage textuel...

Pour les études génétiques, Zola constitue le parangon de l'écriture dite « préparatoire ». Les correspondances relatent de ce travail effectuée dans une absolue transparence. Il est à toutefois à noter que la logique de l'écriture programmatique n'est pas une invention du siècle de Zola et Flaubert. Elle appartient au patrimoine anthropologique de la pensée qui se projette, s'extériorise dans des langages multiples et se ressaisit pour mieux se relancer. En ce sens, arraché à un contexte historique précis (le XIX<sup>ème</sup> siècle), à un mouvement littéraire périodisé (le réalisme) ou à un genre dominant au sein de ce mouvement (le roman naturaliste), ce procédé scriptural est un réservoir conséquent de données sur la création littéraire et ses points d'ancrage dans la réalité.

Les grands romanciers du XIX<sup>ème</sup> siècle, dont Zola en est le parangon, « ont fait de l'espace une coordonnée essentielle de leur représentation du réel, matériel et social, tant au plan de la semiosis qu'au plan de la mimesis »<sup>52</sup>. Les indications concernant le lieu d'envoi et de réception d'une lettre sont en ce sens un formidable témoignage de l'« ethnographie de proximité »<sup>53</sup> esquissée notamment dans l'œuvre romanesque de ces auteurs. Sur le plan de la génétique textuelle, la correspondance complète en ce sens les dossiers préparatoire de l'œuvre de Zola. La matière première de l'artiste se trouve désormais au coin de sa rue. L'heure est à la micro-histoire, à la consignation du quotidien et à son relais diffracté dans des enveloppes de plus en plus fines. La lettre condense toutes les caractéristiques d'une captation et d'une expression de l'infime de la vie. Une génétique du tracé de l'écrit pourrait t-elle, en ce sens, révéler la nature littéraire, tout autant

<sup>49</sup> Lettre de Flaubert à Georges Charpentier, le 20 mai 1877.

<sup>50</sup> PAGES, Alain. « Stratégies textuelles : la lettre à la fin du XIX<sup>ème</sup> siècle ». *Littérature*, n°31, 1978, p.112. Accessible en ligne : [https://www.persee.fr/doc/litt\\_0047-4800\\_1978\\_num\\_31\\_3\\_1168](https://www.persee.fr/doc/litt_0047-4800_1978_num_31_3_1168)

<sup>51</sup> DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Ecriture », 2002, p.38.

<sup>52</sup> MITTERAND, Henri. *Zola, tel qu'en lui-même*. Paris : Presses Universitaires de France. 2009, p.60.

<sup>53</sup> Ibid. p.70.

qu'historique, de ces lettres écrites sur le terrain, par Zola ?

Soulignons que la correspondance « oblige à penser que la variable pertinente pour étudier l'écrit de l'homme (...) n'est pas la forme mais l'espace dans lequel ils s'inscrivent : espace matériel de leur production (on n'écrit pas certaines lettres sous le regard des siens ; on choisit ses supports et ses temps de correspondance) mais aussi espace supposé ou réel des réceptions »<sup>54</sup>.

Une fois encore, la correspondance de Zola doit s'appréhender à un niveau spatio-temporel, ce qui n'est pas aussi prégnant dans celle de Flaubert, où le propos de la lettre prédomine sur le contexte de sa production.

Cela nous enseigne que l'étude de la correspondance d'un écrivain ne peut s'effectuer qu'à la lumière de la trajectoire de vie de son auteur et de celles de ses contacts épistolaires. C'est ainsi que « l'énoncé circule, sert, se dérobe, permet ou empêche de réaliser un désir, est docile ou rebelle à des intérêts, entre dans l'ordre des contestations et des luttes, devient thème d'appropriation ou de rivalité »<sup>55</sup>. Cette approche du texte épistolaire nous intéresse ici, car le présupposé de l'analyse génétique des correspondances sera en définitive l'étude de la lutte de l'écrivain contre le papier. Ou au contraire de son délestage derrière cette surface-écran... Les deux écrivains que nous analysons adoptent génétiquement ces attitudes respectives, en fonction de leur état d'esprit du moment. Les lettres écrites par Zola en temps de crise politique ou celles écrites par Flaubert dans sa retraite orientale ne sauraient renvoyer les mêmes images, à la fois intérieures et extérieures. Mais pour que le langage développé en leur sein puisse alors être pris comme un objet, comme un méta-langage, décomposé en niveaux distincts, décrit et analysé, il faudra qu'il existe un référentiel, qui ne soit autre que le sujet, la personne unique de l'écrivain. Et c'est la capacité rétrospective, propre au généticien étudiant manuscritement ces correspondances après la mort de leurs auteurs, qui permettra la mise en place d'un tel système. L'objet-lettre et sa matérialité feront office de cadre expressif, à travers lequel la dimension discursive de la lettre pourra être réactualisée, comme nous le verrons dans les prochaines parties. Du point de vue linguistique, cette appropriation devient si effective que Roland Barthes affirmait à son sujet que « la naissance du lecteur doit se payer de la mort de l'auteur ». Le destinataire de l'époque, tout comme le lecteur découvrant aujourd'hui la lettre d'un écrivain, est le dernier réceptacle où l'unité du texte épistolaire se produit. L'auteur n'est définitivement plus le garant du sens de son « œuvre » dès lors que ses lettres seront archivées et diffusées, car celles-ci dévoilent *de facto* des pans entiers de sa conception, à commencer par sa dimension génétique. Il y a dès lors une sorte de prééminence absolue du texte épistolaire qui prend le pas auteur sur le langage même de l'écrivain, dans ce qui pourrait s'apparenter à un net rejet de l'idée d'intentionnalité. Cela aura des conséquences quant à la conservation effective des lettres et leur archivage, d'où de nouveaux langages viendront émergeront.

### I.2.3. La conservation des fonds de correspondances de Gustave Flaubert et d'Émile Zola

Nous devons à présent parler de la trajectoire effective des lettres de ces deux écrivains, après leur disparition, afin de préparer le terrain de leur exploitation passée et actuelle. Les deux hommes souhaitaient donc que leur fonds épistolaire trouve leur propre dimension à travers le temps ; c'est pourquoi, à aucun moment, ils ne s'opposèrent à la diffusion publique de leurs lettres. Ce qui n'a pas tout le temps été le cas du côté de leur famille, pour diverses raisons, comme nous le verrons.

<sup>54</sup> BOSSIS, Mireille. *L'épistolarité à travers les siècles*. Stuttgart : Franz Steiner, 1990, p.167.

<sup>55</sup> FOUCAULT Michel, *L'archéologie du savoir*, Paris : Éditions Gallimard, 1969, p.138.

L'héritière de Gustave Flaubert, Caroline Franklin Grout, a tout d'abord fait don des manuscrits de l'écrivain en 1914 à la Bibliothèque Nationale parmi lesquels se trouvait une partie de la correspondance. Puis, à son décès en 1931, par legs testamentaire elle transmis à l'Institut de France le reste du fonds de lettres écrites et reçues par Gustave Flaubert, dont elle avait héritée. Il s'agit d'un fonds de près de 3000 lettres, aujourd'hui conservées à la bibliothèque de l'Institut à Paris. Les lettres reçues par l'écrivain ont été transféré par l'Institut à la Bibliothèque Spoelberch de Lovenjoul à Chantilly. Celles-ci ont alors rejoint une collection qui tire son nom du vicomte ayant légué en 1905, à la bibliothèque, un ensemble de manuscrits témoignant de la vie littéraire du XIX<sup>ème</sup> siècle... De tels agencements historiques nous montrent une fois encore que la lettre, seule, est une « forme sans signifié, c'est-à-dire non lisible par elle-même », et qu'en cela, elle ne « prend sa signification qu'intégrée à un niveau supérieur, soit d'ordre matériel, soit d'ordre intellectuel. »<sup>56</sup>. Dans cette quête de sens, le département des manuscrits de la Bibliothèque Nationale a entrepris, dans les années 1950, de classer et de cataloguer ce fonds en 12 volumes reliés. Nous n'étions pas là encore dans une volonté d'exploiter les lettres mais de leur insuffler une signification documentaire.

Mais si une partie conséquente de la correspondance active de Gustave Flaubert se trouve encore en ce cadre, il n'y eu toutefois, au fil des années, aucune centralisation de ce fonds physique<sup>57</sup>. La raison principale est le volume global du fonds de lettres. Précisons que, quantitativement, la Correspondance de Flaubert représente environ 5000 pages de textes publiés. Soit un volume quasi équivalent à celui de son œuvre littéraire toute entière, bien que la plupart des échanges épistolaires ait été amputé à travers le temps, comme nous le verrons par la suite.

Selon les destinataires à qui Flaubert écrivait, qui ont conservé ou légué à leurs tour ces lettres, chaque échange a connu une trajectoire archivistique singulière au fil des ans. Nous pouvons évoquer à ce titre, la dation en 1991 à la BNF de la correspondance de Flaubert avec son éditeur Michel Lévy, la conservation à la BMC de Rouen des lettres adressées à Léonie Brainne ou encore l'Archivio Campello (Rome), où sont conservées cinquante-cinq lettres de Flaubert à la princesse Mathilde. Mais se focaliser sur les bibliothèques et instituts culturels serait au combien réducteur, si l'on veut tenter de recenser l'ensemble des lettres manuscrites rédigées par Gustave Flaubert ou par Émile Zola, à travers le monde. Comment ne pourrions-nous ne pas prendre en compte les catalogues de ventes privées et les collections particulières, où sont susceptibles d'émerger de nouvelles lettres, conservées jusqu'alors par les héritiers des proches ainsi que des connaissances de ces auteurs ?

La correspondance active de Zola est conservée à ce titre dans de nombreux fonds publics et privés du monde entier. Toutefois, ces lettres écrites par Émile Zola ont pu bénéficier d'une forme de centralisation « physique » avant celles de Flaubert, leurs copies ayant été archivés aux locaux des Programmes de recherche sur Zola (basés à Paris et à Toronto). Cette initiative avait alors requis la collaboration des institutions patrimoniales et des possesseurs de lettres de l'écrivain à travers le monde. Les spécialistes de l'écrivain ont pu, à partir de ces copies, construire des inventaires à la lumière de leur connaissance sur la vie de Zola, pour chaque lettre, faits sur fiche. La correspondance passive de l'écrivain fait quant à elle partie de la collection "famille Émile-Zola". Cela signifie que seuls ses héritiers étaient habilités à décider dans un premier temps si oui ou non, certaines lettres reçues par Émile Zola pouvaient être diffusées au public.

Une partie conséquente de ce fonds a fait l'objet, de la part de la descendance de Zola, d'une dation à la BNF en 1932. Ce fonds représentait un total de 4466 lettres expédiées à l'intention d'Émile Zola, qui ont pu rejoindre le département des Manuscrits de la BNF. Comme nous l'avons

<sup>56</sup> PAGES, Alain. « Stratégies textuelles : la lettre à la fin du XIX<sup>ème</sup> siècle ». *Littérature*, n°31, 1978, p.116 Accessible en ligne : [https://www.persee.fr/doc/litt\\_0047-4800\\_1978\\_num\\_31\\_3\\_1168](https://www.persee.fr/doc/litt_0047-4800_1978_num_31_3_1168)

<sup>57</sup> Pour accéder à une liste non-exhaustive des lieux de conservation des lettres écrites par Gustave Flaubert : <https://gallica.bnf.fr/html/und/litteratures/flaubert-correspondance?mode=desktop>

expliqué, elles sont un témoignage essentiel pour la compréhension de la vie et de l'œuvre de l'écrivain et devaient être préservées en cela à titre historique. Plusieurs destinataires récurrents peuvent alors rapidement être identifiés dans ces collections, ce qui peut par exemple donner lieu à des sectionnements opérés de la part des bibliothécaires à l'échelle interpersonnelle. La sélection de certains échanges épistolaires, dès l'étape de la récolte des manuscrits, est ainsi à l'origine de certains gisements de la recherche. Il ne s'agit donc pas d'une mise à l'écart d'un corpus de lettres extraites de l'ensemble, mais d'une mise en lumière d'un certain moment de la correspondance.

Les derniers fonds de lettres arrivés à la BnF concernant l'écrivain, sont les correspondance d'Émile et d'Alexandrine Zola avec également celles Alfred et Philippine Bruneau ; soit trois cent vingt-huit lettres ayant fait leur entrée dans les collections publiques en février 2014, grâce à une dation en paiement de droits de succession<sup>58</sup>. Nous verrons dans quelles mesures ces fonds présentaient une matière indéniable à exploiter pour les éditeurs, d'hier comme d'aujourd'hui, avec le concours des archivistes ayant travaillé à l'extraction et à valorisation de ces lettres. Car l'appel à une conservation centralisée des lettres d'un écrivains est un défi perpétuel, lorsque l'on postule la difficulté, parfois sentimentales, pour ces possesseurs privés de se séparer de ces témoignages familiaux, avant d'être historiques, à leur échelle.

Mais le public est de plus en plus sensible à ces enjeux d'archivage coordonné et collectif des correspondances d'écrivains du XIX<sup>ème</sup> siècle, en envisageant la perspective d'une diffusion de ces données inédites à grande échelle. Au cours des quinze dernières années, les catalogues de ventes publiques et ceux des marchands d'autographes ont révélé des lettres inédites de Zola et de Flaubert. Des collectionneurs, des conservateurs de bibliothèques universitaires ou publiques, les descendants de collègues ou amis de ces écrivains ont fait parvenir aux institutions patrimoniales des copies des lettres qu'ils venaient d'acquérir ou de découvrir. En ce qui concerne Flaubert, nous pouvons citer, parmi les plus illustres collectionneurs des documents originaux de sa correspondance, le docteur Besançon ou encore le colonel Sicklès... La diversité des lieux de conservation des lettres d'écrivain du XIX<sup>ème</sup> siècle nous enseigne la complexité à se renseigner sur un corpus épistolaire quel qu'il soit. Le nombre de lettres conservées mais aussi l'authenticité des autographes présents sur les lettres, sont autant de variables d'ajustements que les chercheurs devront prendre en compte avant le dépouillement du fonds épistolaire d'un écrivain. Mais ces précautions d'emploi en valent assurément la chandelle, comme nous le constaterons. Elles s'inscrivent dans la volonté même de ces auteurs réalistes qu'étaient Zola et Flaubert, expression corollaire à l'adjectif « pragmatiques », pourrions-nous dire, qui présentaient déjà les gisements inclus par l'édition posthume de leurs correspondances...

« En principe, je ne suis pas du tout opposé à la publication posthume de la correspondance des écrivains. Quelques auteurs ont critiqué les lettres de Balzac, de Flaubert ; pour moi, j'ai lu ces lettres avec le plus grand intérêt, je les ai trouvées admirables dans leur simplicité !... Sous ces phrases non travaillé son sent passer un souffle de passion violente ou amère, on devine les souffrances de la vie qui ont torturé l'âme de Balzac... de Flaubert ! »<sup>59</sup>.

<sup>58</sup> SANDRAS, Agnès. VIAL, Charles-Éloi. « Zola à la Bibliothèque nationale ». Genesis : Manuscrits – Recherche – Invention, 42 / 2016, p.149. En ligne : <https://journals.openedition.org/genesis/1660>

<sup>59</sup> Tiré d'un article de la « Nouvelle Revue Internationale » datée du 15 juillet 1896.

## I.3. L'appropriation du corpus épistolaire du XIX<sup>ème</sup> siècle

### I.3.1. Les éditions papiers des correspondances : leurs limites

Avant de nous intéresser à la valorisation actuelle des corpus de correspondances d'écrivains du XIX<sup>ème</sup> siècle, nous nous intéresserons dans un premier temps à l'exploitation physique qui fut la leur, lorsque le Web en était encore à ses balbutiements...

Pour envisager l'exploitation d'un tel corpus, il fallait alors, pour commencer, songer aux barrières éventuelles à la diffusion de ces documents, autres que celles purement pratiques que nous avons déjà listé. Les chercheurs devaient prendre conscience de la nature réelle des lettres et de leur portée historique, en les considérant comme un fonds d'archives uniques en leur genre. Les archives sont « l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale, et par tout service ou organisme public ou privé, dans l'exercice de leur activité »<sup>60</sup>. Nous allons voir que des clauses juridiques protègent l'accès à ces documents, dont la plupart contiennent des éléments sur la vie privée de l'écrivain et de ses contemporains. De surcroît, l'indifférenciation des lettres au sein de leur corpus sera également à prendre en compte dans cette perspective.

Nous avons vu la diversité de provenance des lettres, au moment de leur envoi, ainsi que la dispersion géographique a posteriori. Les éditeurs devaient s'interroger sur la provenance de chaque document, en se basant sur l'environnement connu de l'écrivain. La mobilisation de la lettre ne sera en effet pas la même si celle-ci était jusqu'alors restée dans la sphère privée de l'écrivain, en possession de ses descendants, ne comportant donc pas le filtre institutionnel et social d'autres lettres, plus conventionnelles et formelles.

La communication des archives épistolaires nécessite également quelques précautions d'usage au regard du droit de divulgation et du respect de la vie privée, et notamment dans le cas des lettres inédites, susceptibles de contenir de nouvelles informations à caractère personnel. Il est alors nécessaire d'obtenir, à défaut de l'accord de l'auteur, celui de ses ayants droit. Les descendants de la famille de Gustave Flaubert et d'Émile Zola disposent ainsi d'un droit moral de divulgation du scripteur et de ses ayants droit<sup>61</sup>. De surcroît, ces personnes peuvent à tout moment demander une autorisation de reproduction des lettres manuscrites en question, afin de s'assurer que leur diffusion respecte le contenu de la correspondance... À la lettre. Cette clause s'inscrit dans le droit au respect de la vie privée<sup>62</sup> de toute personne mentionnée (de son vivant et postérieurement, si atteinte à l'honneur ou à la réputation) au sein des lettres publiées. Les institutions en possession de ces documents devaient bien souvent redoubler de vigilance au moment de la mise à disposition de ces lettres, notamment dans les échanges avec les éditeurs. Cela s'avère toujours d'autant plus vrai aujourd'hui avec les possibilités de dématérialisation de la lettre, qui la rendent perméable aux manipulations à grande échelle, comme nous le constaterons. C'est pourquoi « la plupart du temps, les bibliothèques font preuve de pragmatisme, en choisissant de mettre en ligne des documents

<sup>60</sup> Cf : Article L.211-1 du Livre II du Code du patrimoine.

<sup>61</sup> Cf : Article L111-3 du Code de la Propriété intellectuelle.

<sup>62</sup> Cf : Article 9 du Code Civil.

qu'elles peuvent ensuite retirer en cas de litige »<sup>63</sup>. Or, l'édition en papier ne permettait pas ces actualisations du contenu des correspondances et leur malléabilité.

Les correspondances de Gustave Flaubert et d'Émile Zola ont fait toutes les deux l'objet de publications en plusieurs tomes. Les éditions du CNRS et les Presses de l'Université de Montréal ont fait édité, entre 1978 et 2010, environ 5000 lettres écrites par l'écrivain, équivalent à la quasi intégralité de sa correspondance active. Ces volumes sont aujourd'hui difficilement trouvables, en dehors des bibliothèques universitaires. Notons qu'une première édition de la correspondance de Zola avait été publiée chez Eugène Fasquelle, en 1907. Celle-ci contenait deux volumes contenant près de trois cent cinquante lettres, recopiées par Alexandrine Zola, veuve de l'écrivain, à partir des originaux que les correspondants de son époux lui avaient communiqués... Une seconde édition de plus de six cent lettres est parue en 1929, sans pour autant quitter le cercle décisionnel familial. Elle était cette fois-ci l'œuvre de Maurice Leblond, gendre de Zola. Ces éditions fragmentaires montraient d'une part le souhait d'une livraison immédiate des souvenirs de l'écrivain, devant la curiosité des lecteurs. L'épistolaire pouvait alors être perçu comme une denrée de consommation, plutôt que comme un fonds d'archives à faire fructifier, avec le peu de recul qu'avaient alors ces acteurs littéraires sur le contenu réel des correspondances. C'est pourquoi les équipes éditoriales de la seconde moitié du XX<sup>ème</sup> siècle ont créé la publication d'une « correspondance générale » de l'écrivain du XIX<sup>ème</sup> siècle. En s'attendant à produire un appareil critique approfondi de l'œuvre épistolaire de Gustave Flaubert ou encore d'Émile Zola, ces équipes désacralisaient du même coup leur correspondance, sans donc plus aucun objectif mercantile ou voyeuriste. Concernant l'édition consacrée à Zola, nous pouvons retenir l'apport des notices biographiques des personnes citées par Zola ; l'index analytique des périodiques cités par Zola ; l'index des œuvres de Zola, et l'index des noms propres, cités dans les introductions, les lettres et les notes, l'index analytique des périodiques cités par Zola ou encore l'index des œuvres de Zola... De plus, une ouverture à la correspondance passive de l'écrivain était opérée à l'aide de citations intégrales ou partielles, des résumés ou des allusions à certaines lettres envoyées par des destinataires illustres de Zola.

« L'éditeur d'une correspondance générale ne s'apparente donc plus comme précédemment au collectionneur avisé qui recherche la pièce exceptionnelle, mais au maniaque qui accumule sans distinction et compulsivement tout ce qui touche à son sujet »<sup>64</sup>.

Du côté de Gustave Flaubert, les Bibliothèque de la Pléiade ont fait paraître la correspondance de l'écrivain en 5 tomes, dirigés successivement par Jean Bruneau et Yvan Leclerc. Ce corpus représentait environ 4000 lettres rédigées par l'auteur de « Madame Bovary ». Un index général des noms propres et des correspondants y figurait. Avant cela, une partie des lettres de Flaubert avaient été publiées dans la Nouvelle édition augmentée Louis Conard. Cette édition, qui a fait date, se composait de 9 volumes (ou séries), parus entre 1926 et 1933, et de 4 suppléments (1954). Concernant l'édition de la correspondance passive, nous pouvons noter les Lettres de Maupassant à Flaubert, reproduites d'après Guy de Maupassant, dans une édition établie par Jacques Suffel, « Le Cercle du bibliophile ». Mais les exemples en la matière n'étaient pas vraiment aboutis, car trop parcellaires. L'établissement de ces éditions tendait d'ores et déjà à l'exhaustivité de la présence des documents, connus et

<sup>63</sup> PINÇON, Juliette. « Les archives des écrivains, leur place en bibliothèque ». Enssib : Mémoire d'étude DCB, 2017, p.50. En ligne : <https://www.enssib.fr/bibliotheque-numerique/documents/67311-les-archives-des-ecrivains-leur-place-en-bibliotheque.pdf>

<sup>64</sup> CHOTARD Loïc. « Correspondances : une histoire illisible ». *Romantisme*, n°90, 1995, pp. 7-26. En ligne : [https://www.persee.fr/issue/roman\\_0048-8593\\_1995\\_num\\_25\\_90](https://www.persee.fr/issue/roman_0048-8593_1995_num_25_90)

recensés par les institutions patrimoniales, à l'intégralité de chaque texte épistolaire reproduit et à la fiabilité des datations.

Toutefois, ces éditions présentaient leurs limites. Ce qui était par exemple dommageable, avec les 11 tomes de l'édition papier originelle des lettres de Zola, c'est qu'une partie non négligeable des lettres destinées à Alexandrine Zola n'avait pas été intégrée à ces volumes. La faute à des suites de questions d'héritage complexes, survenues entre les descendants de l'écrivain. Il en fût de même pour les lettres à Jeanne Rozarot, maîtresse de l'écrivain et mère de ses deux enfants. En laissant dominer le propos public et les négociations d'affaires, ces discussions familiales ont empêché toute collaboration culturelle pendant plus d'un siècle sur ces corpus, ce qui faussa dans une moindre mesure fausser la tonalité d'ensemble de ces éditions papiers.

Se posait aussi la question de la complétude du fonds de la correspondance d'un écrivain. La correspondance de jeunesse de Zola a été largement amputée, pour cause de perte des manuscrits et d'une réécriture opérée par la veuve de l'écrivain, Alexandrine Zola. Le cercle familial de l'écrivain avait-t-il, alors, tout intérêt à nuancer certains propos de Zola ? Le généticien qui fait face à des lettres au propos étonnamment lisse provenant de l'écrivain ne peut pour autant pas se permettre des modifications d'usage qui outrepasseraient totalement son rôle d'intermédiaire artistique et culturel. Était-t-il pourtant honnête, sur le plan intellectuel, de présenter une correspondance active sans doute remaniée, dans laquelle les propos de l'auteur ne seraient pas en cohérence avec leur intention originelle ? Cette question s'est posée à un niveau supplémentaire face à la correspondance passive de ces deux écrivains, avec un nouvel écueil émergeant face à la diffusion éditoriale des correspondances... Celui d'une décontextualisation des lettres de ces écrivains.

Il est aujourd'hui acquis, chez les spécialistes de Gustave Flaubert, que celui-ci était désireux de laisser à la postérité ses œuvres romanesques, en premier lieu. Comme nous l'avons montré, il était allé jusqu'à nier son principe biographique dans ses propres lettres, et par corollaire, le pan épistolaire de son œuvre. Mais il lui arriva d'aller au bout de sa démarche, d'une façon pour le moins radicale, dans la manière de se départir de ces documents. L'écrivain a donc organisé, en toute lucidité et à plusieurs reprises au cours de sa vie, la destruction d'une partie des lettres qu'il avait reçues, avec l'accord de certains de ses correspondants, qui supprimaient leurs propres lettres reçues en retour. C'est donc une partie importante de la correspondance active et passive de Flaubert qui restera pour ainsi dire lettre morte. Mais comme ces actes d'auto-dafé épistolaires étaient plutôt rares au XIX<sup>ème</sup> siècle, il nous faut citer d'autres causes de la décontextualisation d'une correspondance d'écrivain en prenant l'exemple d'Emile Zola cette fois-ci.

Dans les archives familiales de l'écrivain, c'est-à-dire dans les collections privées des héritiers de l'auteur, plus de 15000 autres lettres reçues par l'écrivain ont été conservées communément. Nous avons déjà cité leur situation géographiquement éparse, lors de l'envoi de ces lettres qui constitueront la correspondance passive de l'écrivain. Celles-ci provenaient alors du monde entier : Angleterre, Hollande, Allemagne, Suisse, États-Unis... Elles ont été adressées à Zola par des directeurs de journaux, par des étrangers, par des admirateurs et des adversaires inconnus, par des éditeurs, mais également par des amis intimes, par des parents proches... Tous ces éléments concourent à une indifférenciation de l'information entourant cette correspondance, en définitive. L'édition de ces 20000 lettres est un projet très ancien mais qui n'était pas réalisable au format papier pour des raisons évidentes d'économie, et d'efficacité dans la disposition informationnelle de cette immense correspondance. Des correspondants illustres de Zola ont toutefois vu leurs lettres éditées, parfois dans une correspondance croisée avec Zola. Il y eut ainsi la publication de trois « lots » de lettres, de 1958 jusqu'à 1980, constitués de morceaux choisis de cette correspondance passive, avec des lettres qui présentaient des informations sur l'actualité littéraire, théâtrale et politique du XIX<sup>ème</sup> siècle.

Ont été intégrés à ces éditions, notamment 46 lettres de Guy de Maupassant ou encore 43 lettres d'un certain Gustave Flaubert, adressées à Émile Zola... Il n'en demeure pas moins que les milliers de lettres d'inconnus écrites à l'attention de l'écrivain restaient inédites au public.

Face à ce dernier problème, les spécialistes de l'écrivain, qui entendaient diffuser tous les documents le concernant, ne pouvaient pour autant se résoudre à laisser ne serait-ce qu'une lettre de ce fonds sur le côté. « De ces 20000 messages, il n'y en a pas un qui soit sans intérêt pour l'historien (...) Le moindre mot d'une de ces lettres peut fournir une précieuse information à un biographe ou à un historien des échanges culturels nationaux ou internationaux. »<sup>65</sup>. Éditer alors la correspondance passive aurait nécessité une publication en 50 tomes « alors que de toute évidence ces textes ne forment qu'une partie de la vaste polyphonie de l'ensemble »<sup>66</sup>. Face au volume immense de ces papiers ne s'est posé à aucun moment la question de la suppression d'un nombre régulier de documents afin de rendre l'entreprise envisageable. Les spécialistes de l'écrivain, les bibliothécaires et les archivistes soupçonnaient-ils déjà l'arrivée des plateformes d'archivage sur le Web et leurs capacités d'encodage numérique de l'information condensée au sein de la lettre ?

Après avoir abordé l'importance générale des correspondances de ces écrivains, nous voyons, à une échelle plus réduite, qu'un réel gisement documentaire se présentait alors aux yeux des chercheurs, dès lors que des corpus spécifiques de lettres étaient créés. Fallait-il encore que tous ces documents soient mutualisés et en libre accès.

« Avant l'ère de l'informatique il aurait été futile de songer à aborder ce genre d'entreprise (...) Une fois la décision prise, ce ne serait plus qu'une question d'organisation. On répartirait en groupes distincts la masse des lettres, et l'on aborderait le travail systématiquement »<sup>67</sup> ....

Les potentialités du Web, à travers l'archivage numérique des correspondances d'écrivain que nous présenterons dans un second temps de ce travail, répondra à cette optique d'un traitement presque automatisé de la lettre. Il nous fallait avant cela rappeler que « les grandes éditions de correspondances présentent un ensemble d'interférences, de paroles, de connexions, qui constitue en lui-même une forme nouvelle d'intelligibilité historique » et participent à « l'histoire non écrite de la littérature en train de se faire »<sup>68</sup>

Le dernier enjeu était alors de mettre la correspondance de ces écrivains devant les chercheurs, non pas sous un format standardisé de la publication, cette fois-ci, mais sous un ordre manuscrit. Une documentation capitale sur l'histoire de la culture du XIX<sup>ème</sup> siècle pouvait alors être restituée dans son ensemble opératif, mais cela ne s'est pas déroulé sans poser de nouvelles contraintes archivistiques.

### 1.3.2. La nécessité des collaborations archivistiques pour la contextualisation des correspondances : le cas de Marcel Proust...

Nous avons aperçu les enjeux d'une mobilisation collective et documentée des lettres d'un écrivain du XIX<sup>ème</sup> siècle. Afin de mieux dévoiler les difficultés de ce processus, nous nous

<sup>65</sup> WALKER, John A. « Zola destinataire : 20 000 lettres à éditer ». Cahiers naturalistes, n°61. 1987, p.209. En ligne : <https://gallica.bnf.fr/ark:/12148/bpt6k9788419p>

<sup>66</sup> Ibid. p.210.

<sup>67</sup> Ibid. p.214.

<sup>68</sup> CALLU, Florence, DEBRAY GENETTE, Raymonde, NEEFS, Jacques. *L'œuvre de l'œuvre : étude de la correspondance de Flaubert*. Paris : Presses Universitaires de Vincennes, 1993, pp.6-7.

appuierons sur la correspondance d'un autre grand écrivain français, situé chronologiquement à cheval entre le 19<sup>ème</sup> et le 20<sup>ème</sup> siècle : Marcel Proust.

Un éditeur américain, Phillippe Kolb, a consacré une grande partie de sa vie à dater et à annoter le fonds de la correspondance de l'écrivain, composé d'environ 5000 lettres. Celles-ci furent publiées entre 1970 et 1993, au sein de 21 volumes d'une édition papier que Phillippe Kolb a entièrement mené par lui-même. De ce travail au long souffle ont résulté près de 40 000 fiches biographiques, chronologiques, bibliographiques et toponymiques, écrites à la main, qui éclairaient considérablement la correspondance de l'écrivain. L'annotation de Philip Kolb se voulait extrêmement érudite et précise. Toutefois, un tel volume d'information constitue en soi une limite, à première vue infranchissable, en ce qui concerne l'accès immédiat à une correspondance contextualisée.

« Ces fiches constituant une sorte d'hypertexte de la correspondance (puisque chaque nom de personne, de lieu, chaque événement politique, chaque date, chaque citation, etc., mentionnés dans une lettre font l'objet d'une fiche), il s'avéra vite frustrant de ne pas pouvoir lier à chaque fiche la lettre qu'elle annote »<sup>69</sup>

Leur numérisation allait permettre, quelques décennies plus tard, de lier les notes des lettres aux fiches numérisées, au sein d'une archive numérique sur le Web nommée « The Kolb-Proust Archive for Research »<sup>70</sup>. Cet exemple nous montre qu'un fonds de correspondance pourra non seulement s'enrichir quantitativement au fil du temps, avec l'arrivée de lettres inédites, écrites ou reçues par l'écrivain, mais également d'un point de vue purement documentaire. Chaque annotation faite à propos d'une lettre présente un intérêt intellectuel, et de ce fait archivistique. Il s'agira de décrire « l'emboîtement d'écritures autour de l'écriture » contenue dans la lettre afin de les faire passer « d'un statut de note de recherche ou d'aide-mémoire rédigé par et pour un seul individu à celui de notice documentaire à l'usage d'un public varié »<sup>71</sup>.

Ces éléments de conservation spécifiques que sont les annotations peuvent tout aussi bien être l'œuvre du destinataire, du collectionneur ou de l'institution patrimoniale jusqu'à laquelle est parvenu le legs d'une correspondance d'écrivain. Les matériaux sur lesquels ces annotations sont susceptibles d'avoir été apposées sont, dans le cas de lettre, le support même de la feuille, l'enveloppe ou bien un objet associé à la lettre (prenons l'exemple de la fleur séchée glissée dans l'enveloppe, pour les écrivains les plus « romantiques » du 19<sup>ème</sup> siècle...). Il faut alors composer avec ces éléments afin que la description matérielle de la lettre soit indexée le plus correctement possible, pour que le lecteur puisse retrouver visuellement la texture de ce document manuscrit.

C'est donc à partir du moment où les lettres d'un écrivain sont envisagés, par les chercheurs, comme des documents à intégrer au sein de fonds d'archives, que leur renseignement se révèle impératif. Le maintien de leur l'intégrité et la pérennisation de leur contenu textuel et visuel, à travers l'espace et le temps, est à ce prix. Pour adopter un comportement de recherche devant ces lettres réunies devant nous, il faut avoir conscience que « l'unité n'est pas l'archive, le fonds ou la série [de lettres, en l'occurrence], mais c'est le comportement archivistique »<sup>72</sup> lui-même. Nous n'avons plus à faire à de simples textes

<sup>69</sup> LERICHE, Françoise. « Quelle édition pour quel public ? ». *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008, p.63. En ligne : <https://journals.openedition.org/recherchestravaux/89>

<sup>70</sup> En ligne : <http://kolbproust.library.illinois.edu/proust/search>

<sup>71</sup> SZYLOWICZ Caroline, « Proust, Kolb, Kolb-Proust : emboîtement d'écritures autour de l'écriture », Document numérique (Vol.5), 2001, pp.135-154).

<sup>72</sup> BERT, Jean-François. RATCLIFF, J. Marc. *Frontières d'archives : recherches, mémoires, savoirs*, Paris : Archives contemporaines, 2015, p.18.

littéraires mais bien à des archives, que l'on peut qualifier dans un premier temps de *génétiques*.

Le rôle de l'archiviste, au-delà de celui de l'éditeur, sera donc fondamental dans la valorisation d'un corpus épistolaire. L'exemple de Phillippe Kolb nous enseigne qu'il est souvent difficile, à l'échelle individuelle, de songer à toutes les facettes de la restitution publique d'une œuvre si faramineuse, de par sa taille, et mystérieuse, de par son éclatement spatio-temporel. Il devenait palpable pour ces chercheurs, avec un matériau polymorphe comme seule pouvait l'être la lettre d'écrivain, que l'archivage d'un fonds de correspondance dépendait entièrement du regard qu'on porterait collectivement sur celui-ci.

Lorsqu'on décide de se saisir du matériau de la lettre, on sera tributaire à la fois de son existence, de son accessibilité mais également de son classement. Si ces manipulations sont planifiées a minima, elles pourront produire un sens de lecture de la correspondance de ces écrivains. Elles influenceront ainsi, en retour, les usages possibles que l'on peut faire de ces lettres particulières, qui ne sont dès lors plus des supports de communication, déifiant le présent dans leur course effrénée... Les lettres s'historicisent.

Seule la passion indéniable pour un écrivain saurait outrepasser ces contraintes logistiques. Cela nous montre que pour que de tels travaux de référencements puissent être exploitées, le regard du groupe prime sur l'appréhension d'un corpus épistolaire du XIX<sup>ème</sup> siècle.

L'enjeu est de maintenir un discours cohérent sur les lettres et dans leurs relations entre toutes celles écrites et reçues par un écrivain. Il y' a par donc des limites inhérentes à l'épistolaire dans cette entreprise archivistique, en tant que matériau de travail spécifique. La correspondance est un genre réticulaire. Une lettre n'est jamais isolée ; elle se situe dans un réseau de lettres, de correspondants, de lieu et de dates qui doit être pris en compte pour leur articulation au sein du corpus. Par ailleurs, le discours épistolaire est elliptique. C'est un des discours les plus exigeant à annoter puisque le discours épistolaire s'élabore en prenant en considération ce que le destinataire est censé avoir comme référents communs avec l'auteur de la lettre. Ainsi, « parce qu'elle se joue des discours constitués, elle [la lettre] devient l'instrument idéal d'un savoir vivant (...) l'indispensable outil formel d'une vaste réflexion épistémologique »<sup>73</sup>. Il faut considérer qu'une lettre d'écrivain en soi ne sera toujours qu'une face d'un discours à reconstituer avec la lecture effective des lettres de ses correspondants. L'objectif pour les archivistes est de stabiliser ce discours épistolaire à travers la mobilisation collective des correspondances.

Ainsi, « ce n'est pas le « temps qui passe » qui crée les archives d'un certain type de savoirs, les savoirs « concurrentiels », mais bien le rendu public, c'est-à-dire un regard social qui attribue de l'irréversibilité au texte présenté et, aux yeux de l'auteur, qui change le statut de ses éléments »<sup>74</sup>. L'archive qui accueille la correspondance d'un écrivain est amenée à devenir un réceptacle de la mémoire , mais elle ne peut plus assumer ce rôle dès lors qu'elle est décontextualisée. Cette dimension du traitement de la lettre ne peut pas être survolée. La correspondance d'un écrivain est sans cesse sous le joug d'une massification documentaire et de l'affluence de nouvelles lettres, au sein du corpus. Vérifier l'authenticité de chaque document épistolaire est donc impératif, avant toute opération d'enrichissement du fonds d'archives. Ces aventures singulières que furent les éditions papier de correspondances, au XX<sup>ème</sup> siècle, nous enseignent donc que la disponibilité d'un corpus épistolaire ne peut

<sup>73</sup> DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Ecriture », 2002. p.41-43.

<sup>74</sup> BERT, Jean-François. RATCLIFF, J. Marc. *Frontières d'archives : recherches, mémoires, savoirs*, Paris : Archives contemporaines, 2015, p.20.

s'obtenir qu'au gré de partenariats archivistiques entre institutions ou entre équipes scientifiques.

### I.3.3. La construction d'une archive épistolaire du XIX<sup>ème</sup> siècle

Nous avons montré les recoupements qui pouvaient s'établir entre le fonds de lettres d'un écrivain et les archives patrimoniales, si nous concevons celles-ci comme un ensemble de documents historiques, à recontextualiser. «L'archive constitue la mémoire des nations comme, à l'échelle d'une vie ; les lettres que nous gardons témoignent de ce qu'ont choisi nos souvenirs»<sup>75</sup>. La conservation et l'établissement d'un corpus de lettres à travers les siècles n'a rien d'évident, comme nous venons de le souligner au sein de toute cette partie. Il n'en demeure pas moins qu'une fois sa constitution acquise, un fonds de correspondance devra répondre à de nouveaux enjeux, de pérennisation et d'enrichissement pour prétendre à être une archive épistolaire.

L'archivage de ces documents ne commence véritablement qu'une fois que les lettres envoyées et destinées à un même écrivain sont réunies. Face au fonds physique d'une correspondance, l'approche à adopter ne pourra être ensuite qu'à la fois sélective et ouverte. Elle devra concilier la rigueur du traitement documentaire de la lettre, tout autant que son intégration au sein de projets de valorisation des correspondances. Au regard biographique du chercheur en littératures devra répondre la vision prospective de l'archiviste et l'œil classificateur du bibliothécaire. Nous parlons ici d'un régime qui serait « de l'ordre de l'institution plus que de la localisation spatio-temporelle », et qui définirait « des possibilités de réinscription et de transcription (mais aussi des seuils et des limites) plus que des individualités limitées et périssables»<sup>76</sup> en nous appuyant sur le matériau de la lettre.

L'analyse linguistique et le comportement archivistique vont de pair comme nous l'avons souligné dans cette partie. La structure du texte épistolaire est à la fois très libre dans son contenu, mais aussi très cadrée dans sa forme. Ces objets textuels se caractérisent en effet par une structure spécifique: en-tête et date dans la plupart des cas, mais également une formule d'adresse obligatoire, une formule d'adieu, une signature, qui les classent comme «lettres» au premier regard, avant même la lecture du texte proprement dit. C'est pourquoi les informations d'une lettre ne se laisseront appréhender qu'en prenant un certain recul sur elles-mêmes, dont le généticien puis l'archiviste seront les garants.

De plus, si l'épistolaire n'est pas un genre littéraire à proprement parler, celui-ci inclut toutefois des sous-registres : les lettres de condoléances, de félicitations, accusé de réception, ordre de ventes à la Bourse, lettre de vœux, dédicaces, lettres administratives... Ceux que nous avons cités restent relativement codifiés – même si l'épistolier peut introduire en eux des écarts par rapport aux normes sociales. En revanche, d'autres « genres » épistolaires sont moins conventionnels, tels que les nouvelles quotidiennes, les échanges de vues amicaux ou les négociations avec l'éditeur, qu'Émile Zola et Gustave Flaubert ont bien souvent réalisés à titre utilitaire... Or, pour qu'il y ait un corpus de lettres, il faudra que cet ensemble présente une cohérence à travers l'objectivation du contenu de chaque lettre, de l'usage concret qui était le sien au moment de sa rédaction... Afin de nous représenter cette entreprise, nous pouvons

<sup>75</sup> FURET, François. « Histoire quantitative et construction du fait historique. ». *Annales. Economies, sociétés, civilisations*. 26<sup>e</sup>année, N.1, 1971, p.67. En ligne : [https://www.persee.fr/doc/AsPDF/ahess\\_0395-2649\\_1971\\_num\\_26\\_1\\_422459.pdf](https://www.persee.fr/doc/AsPDF/ahess_0395-2649_1971_num_26_1_422459.pdf)

<sup>76</sup> FOUCAULT Michel, *L'archéologie du savoir*, Paris : Éditions Gallimard, 1969, p.136.

d'ores et déjà postuler que tous ces sous-genres évoqués sont des énoncés, au sens où Michel Foucault l'entendait, en terme de construction sémantique.

« Entre la langue qui définit le système de construction des phrases possibles, et le corpus qui recueille passivement les paroles prononcées, l'archive définit un niveau particulier: celui d'une pratique qui fait surgir une multiplicité d'énoncés comme autant d'événements réguliers, comme autant de choses offertes au traitement et à la manipulation. »<sup>77</sup>

Nous nous situons ici-même au fondement de l'approche herméneutique des correspondances d'écrivains, qui tendra à interpréter significativement chaque énoncé jalonnant la lettre. « De par sa spécificité textuelle, le matériau empirique constitué par les lettres exige en effet l'articulation de plusieurs mouvements interprétatifs. »<sup>78</sup>. Car comme nous l'avons soulevé, cette période historique du XIX<sup>ème</sup> siècle a marqué non seulement un tournant dans l'approche qu'a eu l'écrivain face à la lettre, mais plus encore dans le système d'expression artistique général... « L'outillage discursif des artistes qui écrivent des lettres est complexe, sophistiqué, codifié. »<sup>79</sup>. C'est bien là tout l'intérêt d'une étude complète de la lettre qu'on manipulerait comme un objet historique, intriqué dans un format matériel et intégrée à un contexte de production précis, qu'il faudra reconstruire, tels qu'à son origine.

« Dans sa rencontre avec l'archive, en construisant un corpus de lettres, l'historien comme avant lui les auteurs, héritiers ou archivistes, redistribue et redéfinit des unités de savoir, il inaugure un lieu de recommencement. Le contenu ne devient accessible qu'à travers la transparence de l'architecture qui a assemblé des matériaux en objets historiques. »<sup>80</sup>

Ce processus particulier va entraîner les spécialistes de ces écrivains du XIX<sup>ème</sup> siècle à remettre en question temporairement leur rapport aux correspondances à travers le prisme de leur historicisation. Pour promouvoir cette invention complexe mais au combien passionnante qu'est la correspondance générale, leur méthodologie consistera à segmenter et à hiérarchiser chaque lettre en une somme de données quantifiables et enrichies. Comme le sous-entendaient les structuralistes, c'est là le prix à payer pour une mise à disposition effective du texte épistolaire. Cette « désanctuarisation des fonctions de conception » devra permettre aux lettres de ne plus être dépendants du support unique de leur manuscrit et de tomber sous la « logique d'une multidiffusion »<sup>81</sup>.

Que seront les enjeux d'une archive épistolaire ? L'archivage du fonds de lettres d'un écrivain doit recontextualiser l'ensemble des ces textes d'un point de vue historique, géographique et linguistique, en apportant au lecteur des éclairages, par exemple sous la forme d'annotations explicatives. Il devra permettre en outre de diffuser ces documents anciens, rares du fait de leur unicité, pour la plupart inédits, en respectant les droits de divulgation précédemment entrevus. Enfin, l'archivage devra assurer la pérennisation de ces textes anciens, anciennement épars, car n'ayant pas fait l'objet d'une coordination patrimoniale

<sup>77</sup> Ibid. p.171.

<sup>78</sup> TOUATI, Paul. « De la médiation épistolaire dans la construction du savoir scientifique ». Revue d'anthropologie des connaissances, 2010/3 (Vol 4, n°3), pp.451-475. En ligne : <https://www.cairn.info/revue-anthropologie-des-connaissances-2010-3-page-451.htm?contenu=plan>

<sup>79</sup> DAUPHIN, Cécile. « Les correspondances comme objet historique ». Sociétés et représentations, n°13, 2002, p.43. Accessible en ligne : <https://www.cairn.info/revue-societes-et-representations-2002-1-page-43.htm>

<sup>80</sup> DAUPHIN, Cécile. « Les correspondances comme objet historique ». Sociétés et représentations, n°13, 2002, p.46. En ligne : <https://www.cairn.info/revue-societes-et-representations-2002-1-page-43.htm>

<sup>81</sup> HOLZEM, Maryvonne. LABICHE, Jacques. *Dessillement numérique : éaction, interprétation, connaissances*. Bruxelles : P.I.E-Peter Lang S.A., 2017, p.76.

particulière. Nous avons vu que du point de vue matériel, les manuscrits de lettres sont d'ores et déjà, pour la plupart d'entre eux, conservés dans des conditions irréprochables sur le plan matériel et intellectuels par les institutions patrimoniales. Conditions de conservation qui s'imposaient face à la fragilité du support de la lettre, mais qui ne sont plus forcément opératives à notre ère actuelle, marquée par la dématérialisation des supports.

Nous allons à présent développer les conditions actuelles d'une pérennisation de ces correspondances d'écrivain du XIX<sup>ème</sup> siècle. Cette fois, la centralisation de ces fonds de lettres sera facilitée, à travers le prisme d'un nouvel espace de circulation, au sein duquel l'envoi des lettres n'aura jamais été aussi rapide qu'à l'heure du courrier électronique : le Web 2.0. Car nous avons jusqu'ici dressé les limites pratiques et intellectuelles auxquelles ce fonds d'archives du XIX<sup>ème</sup> siècle pouvait faire face, hier. Mais ces questions ne se posent plus dans les mêmes termes ? avec l'arrivée des infrastructures numériques, qui rebalace le jeu des acteurs littéraires et patrimoniaux autour des ces corpus épistolaires. Sur le plan purement intellectuel, n'y aurait t-il pas un risque d'égarement à vouloir poursuivre ce qui pouvait naître de l'intersubjectivité, induit par l'échange épistolaire, à un instant précis, d'ores et déjà passé ? Cet indéfini, qui naissait d'un contexte particulier entourant la lettre au XIX<sup>ème</sup> siècle, « comment peut-il trouver une place, ou se faire entendre dans la cacophonie des nouveaux modes de communication ? ». La multi-correspondance contemporaine, marquée par l'usage dominant des réseaux sociaux, n'est t-elle pas l'ultime forme d'un désaveu culturel de fond pour la correspondance originelle, axée sur un nombre de restreint de personnes ? L'archive épistolaire numérique est-elle susceptible de passer outre ces problématiques énoncées ?

Nous verrons les défis à surmonter, dans ce nouvel espace du Web, visant à redonner à la figure de l'écrivain du XIX<sup>ème</sup> siècle son aura, quand bien même, par essence, « le numérique s'origine dans l'absence de prise en considération d'un sujet incarné »<sup>82</sup> ? A contrario de l'archiviste du XX<sup>ème</sup> siècle qui pouvait orienter ses efforts documentaires vers la seule personne de l'écrivain mais qui achoppait souvent à la structuration efficiente des fonds de correspondances. Cela, par manque de temps, d'espace et parfois même d'initiative culturelle forte dans les milieux institutionnels, comme nous l'avons évoqué... Dans le même temps, nous verrons qu'aujourd'hui, l'expression personnelle s'épanouit dans les espaces virtuels, et que le Web pourrait tendre à « une concrétisation du vécu via une phénoménologie du numérique ». Ces tendances s'inscrivent dans la tradition de l'introspection épistolaire des auteurs du XIX<sup>ème</sup> siècle tels que Gustave Flaubert et Émile Zola.

Nous avons donc essayé de souligner en quoi les correspondances de ces écrivains étaient aujourd'hui d'intérêt général. Ce qui va nous intéresser à présent sera de démontrer comment des équipes scientifiques parviennent, aujourd'hui à les archiver grâce au Web, afin d'élargir considérablement leur champ d'exploration.

<sup>82</sup> BACHIMONT, Bruno. *Arts et sciences du numérique : Ingénierie des connaissances et critique de la raison computationnelle*, Université de Technologie de Compiègne, 2004, p.146.

## **Partie II : LES ARCHIVES NUMÉRIQUES FACE AUX CORPUS DES CORRESPONDANCES DE GUSTAVE FLAUBERT ET D'ÉMILE ZOLA : ÉTUDE DE L'EXISTANT SUR LE WEB**

### **II.1. Les premières approches des équipes scientifiques face aux corpus de lettres numérisées**

L'idée de circulation fait partie intégrante de l'œuvre hyper-médiatique, dans son véhicule et dans son contenu. Cette œuvre en réseau, basée sur l'échange, le dialogue et le partage d'informations, s'inscrit totalement dans la tradition épistolaire, terme qui, rappelons-le, signifie étymologiquement « faire circuler ». L'épistolaire peut ainsi être considéré comme un terreau propice à la mobilité et à l'interactivité de ses données sur le Web. Au cours de cette partie de notre travail, nous allons voir dans quelle mesure le matériau de la lettre d'écrivain pourra prendre sa place au sein des infrastructures numériques. Ce processus comprend diverses étapes que nous étudierons, à la lumière de deux projets d'édition numérique des correspondances d'écrivains de Gustave Flaubert et d'Émile Zola. Ces projets ont été portés par deux équipes scientifiques, dont les approches ont varié sensiblement, à chaque moment clé de leur démarche de valorisation de ces corpus de lettres, sur le Web. Nous verrons comment la numérisation, la transposition et l'encodage numérique des données épistolaires résultent d'une appropriation collective et normalisée de cet objet d'étude tout à fait fascinant qu'est la lettre d'écrivain. Nous décrirons enfin la mise en œuvre, par ces équipes, d'une plateforme d'archivage numérique de ces corpus, en problématisant les conditions techniques, juridiques et intellectuelles de ces réalisations

#### **II.1.1. Contexte de ces projets**

Nous nous appuyerons donc respectivement sur les projets numériques menés par le Centre Flaubert de Rouen ainsi que sur celui du Centre d'Étude sur Zola et le naturalisme et des chercheurs de la Sorbonne nouvelle – Paris 3, intitulé CorrELEZ. Ces entreprises présentent certaines similitudes, au-delà des caractéristiques chronologiques et disciplinaires de leur objet de recherche commun, que nous avons décrit dans notre première partie de travail. Toutes deux ont poursuivi et continuent de poursuivre la réalisation d'une plateforme d'archivage numérique des correspondances d'un écrivain du XIX<sup>ème</sup> siècle. D'un autre côté, ce seront davantage leurs points de divergences, quant au traitement, à l'enrichissement et à l'archivage numérique des lettres, qu'il nous importera de mettre en exergue, afin de cerner toutes les facettes de l'existant. Soulignons également, avant de présenter le contexte de ces projets, que l'équipe du Centre Flaubert a d'ores et déjà mis en ligne et archivé, depuis l'année 2017 sur leur site web<sup>83</sup>, les lettres de Gustave Flaubert qu'ils avaient pris en charge. Nous parlons d'un corpus de 5601 lettres<sup>84</sup> écrites et reçues par Gustave

<sup>83</sup> Cf : <https://flaubert.univ-rouen.fr/correspondance/edition/>

<sup>84</sup> Nombre enregistré le 28 août 2019. Le nombre de lettres mis en ligne par l'équipe du Centre Flaubert est en constante évolution, avec les actualisations récurrentes des corpus épistolaires numériques.

Flaubert, qui sont non seulement mises en ligne, mais également archivées selon les standards internationales de l'archivage numérique des fonds patrimoniaux. Tandis que du côté de l'équipe Zola, la plateforme est aujourd'hui encore en cours de réalisation et sera prochainement opérative sur le Web<sup>85</sup>. Avant de remonter à l'origine de ces projets, nous pouvons esquisser la définition d'une plateforme, étant donné que ce sont ces structures, définies sur le Web, qui permettront aux équipes scientifiques d'archiver les lettres de leurs corpus respectifs.

Pour le Conseil national du numérique « une plate-forme est un service occupant une fonction d'intermédiaire dans l'accès aux informations, contenus, services ou biens édités ou fournis par des tiers. Au-delà de sa seule interface technique, elle organise et hiérarchise les contenus en vue de leur présentation et leur mise en relation aux utilisateurs finaux. A cette caractéristique commune s'ajoute parfois une dimension écosystémique caractérisée par des relations entre services convergents »<sup>86</sup>.

Il est tout d'abord intéressant de noter que ces deux équipes scientifiques font partie d'un groupe de recherche commun à l'échelle nationale. Le Consortium Cahier<sup>87</sup> est une fédération interdisciplinaire regroupant des acteurs qui entreprennent des projets d'édition numérique dans les domaines des « corpus d'auteurs ». Ceux-ci peuvent aussi bien relever de la littérature, de la philosophie ou d'une thématique liée à une école ou à une pratique disciplinaire. Les projets portés par ce consortium se basent sur l'existence d'une œuvre artistique, dont la numérisation et l'enrichissement documentaire en ligne sont donc susceptibles de donner lieu à de nouvelles recherches en SHS. Ce groupe est lui-même issu de la TGIR nommée Humanum, dédiée aux Humanités Numériques, dont nous analyserons dans notre ultime phase de recherche les perspectives technologiques en terme de valorisation pour ces corpus de lettres. Ce consortium a pour objectif de faciliter la transition numérique pour la recherche dans les sciences humaines et, à cette fin, de mettre à disposition divers outils et services visant à traiter, stocker, publier et diffuser les données des groupes de travail affiliés.

Les équipes Flaubert et Zola avaient plusieurs motivations pour intégrer ce groupe de recherche. Le partage d'expérience de chaque groupe en son sein devait les aider à maximiser la diffusion de leurs données épistolaires sur le Web, et ainsi à mieux les pérenniser à travers leur adaptation à des formats numériques standardisés internationalement. L'autre enjeu était de permettre aux chercheurs de ces équipes un libre accès à tous ces corpus, ce qui encourageait à la valorisation de ces productions, en leur donnant un cadre commun et des objectifs sur le long terme.

Le projet d'édition numérique des correspondances de Gustave Flaubert s'inscrit au sein d'un cadre encore plus vaste sur le Web, né en 2001, à savoir le Centre Flaubert. Cette plateforme se propose de présenter à l'internaute un « Tout Flaubert » en ligne. Cet idéal, qui tend à dévoiler le maximum de facettes de l'écrivain, ne pouvait donc qu'émerger qu'en prenant en compte son statut d'épistolier de l'écrivain, parmi les plus prodigieux de son siècle. Le projet est également rattaché au laboratoire de recherche de l'Université de Rouen, intitulé le CÉREDI<sup>88</sup> (Centre d'Études et de Recherche Éditer / Interpréter), qui offre aux spécialistes de l'écrivain un accès optimisé à des textes d'écrivains normands numérisés, parmi lesquels donc se trouve Gustave Flaubert.

<sup>85</sup> Elle sera accessible via cette adresse Web : <https://www.correspondance-zola.fr/>

<sup>86</sup> CNNum, Rapport « Ambition numérique : pour une politique française et européenne de la transition numérique ». 2015, p.59. En ligne : <https://cnumerique.fr/files/2018-04/CNNum--rapport-ambition-numerique.pdf>

<sup>87</sup> <https://cahier.hypotheses.org/>

<sup>88</sup> <http://ceredi.labos.univ-rouen.fr/main/>

Le projet est encadré par Yvan Leclerc, professeur émérite à l'université de Rouen, mais également directeur du Centre Flaubert et du CÉRÉDI, ainsi que par Danielle Girard, coordinatrice de la mise en ligne de la plateforme et des transcriptions. L'équipe scientifique se compose de 19 transpositeurs et de 24 thématiques. Une dizaine d'étudiants ont également pris part à ces opérations génétiques et intellectuelles autour des lettres de l'écrivain. Nous citerons également Jean-Eudes Trouslard, l'informatique du projet, avec qui l'équipe de chercheurs a rapidement trouvé un langage commun, pour la structuration numérique des lettres sur leur plateforme. Nous verrons que ce rapport exclusif, entre une équipe et un informaticien, aura toute son importance concernant les modalités d'archivage des lettres et la création d'une base de données épistolaire sur le Web.

Nous allons maintenant présenter le projet concernant l'édition numérique des correspondances d'Émile Zola sur le Web, intitulé CorrELEZ. Ce projet est actuellement réalisé par des chercheurs du Centre d'Etude sur Zola et le naturalisme (ITEM/CNRS-ENS) ainsi que de la Sorbonne nouvelle – Paris 3. L'Équipe Zola est coordonné par Jean-Sébastien Macke, ingénieur d'études à l'Institut des Textes et Manuscrits Modernes, et rassemble une vingtaine de chercheurs issus de divers pays (France, Brésil, Hongrie, Belgique, Portugal, Canada, Grande-Bretagne). L'équipe Zola entendait rassembler le plus large corpus d'archives épistolaires concernant l'écrivain du XIX<sup>ème</sup> siècle et créer des outils d'exploitation numérique de ce fonds de lettres, via l'encodage de leurs données sur le Web. L' édition numérique de cette correspondance intégrera la plateforme des « Archives Zoliennes »<sup>89</sup> (Archiz), qui recense, depuis 2012, toutes les archives entourant le père du naturalisme (œuvres romanesques, mais également dessins, photographies au sein d'une iconothèque...). La place dédiée aux manuscrits de l'écrivain au sein de cette plateforme est conséquente. Y sont présents notamment les dossiers préparatoires des romans de Zola, numérisés par la Bibliothèque nationale de France, ainsi qu'une transcription des ébauches de 20 volumes des *Rougon-Macquart*. Ces deux derniers exemples vont particulièrement nous intéresser dans notre étude de l'archivage des correspondances de l'écrivain, puisque la dimension manuscrite des lettres sera révélée à la lumière de ces anciens travaux. La présence d'Archiz a permis à l'équipe Zola de mieux structurer ce projets d'édition numérique. Cette plateforme initiale a permis à l'équipe d'obtenir les outils et les compétences nécessaires, pour l'archivage numérique de cette correspondance, selon les standards internationaux en vigueur actuellement.

Ce nouveau projet de l'équipe Zola consiste donc à éditer la correspondance active et passive de Zola et à archiver ce corpus de 25000 lettres sur un site web, développé en Wordpress et affilié à la plateforme d'Archiz. Concernant la correspondance active, les 5000 lettres ont déjà fait l'objet, comme nous l'avons vu, d'une édition en 11 volumes, mais ces derniers sont aujourd'hui difficilement accessibles. L'idée était donc de proposer un accès direct de ce fonds aux chercheurs, avec tous les outils de recherche propres au numérique, que nous détaillerons plus tard. Il s'agissait également de faire découvrir au public les lettres inédites qui n'ont jamais fait l'objet d'une édition scientifique. Enfin, l'articulation de la correspondance active à la correspondance passive de l'écrivain était l'ultime plus-value de ce projet, matérialisant le souhait de visibilité et de pérennité de cet immense fonds documentaire, pour les spécialistes de l'écrivain, depuis le XX<sup>ème</sup> siècle. Une volonté d'enrichissement des données des lettres de l'écrivain a sous-tendu cette démarche. Elle s'est notamment concrétisé avec l'apparition d'un nouvel appareil critique affilié à la correspondance active de Zola, que nous décrirons, et qui devait remettre les échanges épistolaires de l'écrivain dans leur contexte de l' époque.

Comme nous le voyons, ces deux projets d'édition numérique des correspondances de Gustave Flaubert et d'Émile Zola s'intègrent à un environnement, qui est d'ores et déjà

<sup>89</sup> Disponible à cette adresse Web : <http://www.archives-zoliennes.fr/>

pleinement consacré à la figure de l'un de ces auteurs du XIX<sup>ème</sup> siècle, sur le Web. Nous allons montrer que les équipes scientifiques ne souhaitent, néanmoins, pas réduire les lettres qu'elle traiteront à de simples entités, venant se greffer à ces plateformes. Ces documents feront l'objet d'une appropriation spécifique, qui commencera avec la numérisation des manuscrits de lettres et la mise à disposition de leur fac-similé en ligne.. La visée d'une telle opération est en premier lieu scientifique, afin de prouver que les lettres du corpus sont authentiquement autographés de la part de l'écrivain, et en accord avec leur transcription textuelle qui sera la leur, dans l'édition numérique. Nous allons voir comment s'articulent les deux étapes de cette entreprise qualitative, avec par ailleurs la gestion du volume des lettres à numériser qui ne devra pas être occultée par ces équipes. Nous décrirons les outils adoptés pour ces opérations, qui font de ces éditions de correspondances des véritables odyssées technologiques, avant d'être des projets éditoriaux et archivistiques.

## II.1.2. La numérisation des manuscrits de lettres

Concernant les deux projets du Centre Flaubert et de l'équipe Zola, l'étape de numérisation des manuscrits s'est révélé inévitable notamment pour les lettres inédites des auteurs, recueillies par les équipes au cours de leur entreprise. Il est à souligner que la correspondance passive de Gustave Flaubert et d'Émile Zola leur fournissait en ce sens davantage de lettres inédites que la correspondance active. Comme nous l'avons vu, les lettres écrites par ces écrivains ont déjà fait l'objet, au XX<sup>ème</sup> siècle, d'éditions et donc de fouilles archivistiques poussées, dans les divers lieux de conservation de ces manuscrits. Celles-ci ont en grande partie facilité à ce titre les opérations de numérisation menées entre autres par la BnF, à l'orée du XXI<sup>ème</sup> siècle. Les correspondances d'Émile Zola et d'Alexandrine Zola avec également celles Alfred et de Philippine Bruneau, sont par exemple entrées sur le catalogue BnF Archives et manuscrits, dès leur arrivée à la BnF en 2014. Ces nouveaux fonds venaient alors compléter l'effort de mise en valeur des manuscrits entrepris par la BnF à l'occasion du centenaire de l'arrivée du fonds Zola dans les collections du département des Manuscrits. Cette opération symbolique s'est traduite par une grande opération de numérisation des manuscrits, à laquelle les lettres de l'écrivain n'ont pas échappé, doublée d'une série d'événements scientifiques consacrés à ces documents. C'est donc dans ce cadre précis que les deux équipes scientifiques ont formulé, chacune à des échelles différentes, leur demande d'emprunt des versions numérisées des lettres qui existaient alors déjà, comme ce fût le cas pour le Centre Flaubert auprès de la BnF. Ils ont pu obtenir en conséquence une grande partie des manuscrits de lettres de leur corpus sous forme de fichiers numériques et travailler à leur contact, avant leur mise en ligne sur la plateforme. L'enjeu pour eux était en effet de venir confronter le texte de la lettre, ainsi que l'image du manuscrit qui lui est liée, en fac-similé, sur un même plan de visualisation.

Il nous faut garder à l'esprit que « les lettres numériques seront toujours les substituts d'un document par ailleurs inaccessible, sous une forme permettant l'interrogation, l'agrégation et la conservation »<sup>90</sup>. Les équipes souhaitent tendre avant tout vers une représentation du matériau de la lettre et de son contenu, tout en la pérennisant. La dimension patrimoniale est donc sous-jacente à cette entreprise, qui doit gérer des documents anciens et en restituer une vue nette et authentique... Dans les deux projets que nous analysons, « aucune

<sup>90</sup> HANKINS, Gabriel, « Correspondence : Theory, Practice and Horizons », *Literary Studies in the Digital Age: An Evolving Anthology*, 2015. En ligne : <https://dlsanthology.mla.hcommons.org/correspondence-theory-practice-and-horizons/>

édition ne présente de manière transparente l'artefact d'origine »<sup>91</sup> dans le sens où seul un contact physique avec le manuscrit pourrait le rétablir dans toute sa véracité.

C'est pourquoi les caractéristiques de l'objet épistolaire à numériser étaient à appréhender progressivement par ces équipes scientifiques, car elles s'inscrivaient dans des contextes de production et de réception multiples. Ces deux éditions numériques liées aux correspondances de Gustave Flaubert et d'Émile Zola ont ainsi, chacune, une double tradition à respecter. La première est historique avec la pérennisation d'un fonds de documents anciens. La seconde est critique avec la restitution d'un texte d'écrivain dans l'environnement d'étude littéraire et social qui est d'emblée le sien.

Nous avons cherché à montrer, dans notre première partie, dans quelle mesure il y' avait une archivé épistolaire à construire, en partant de la conservation matérielle, effective du fonds de lettres d'un écrivain. Or, ces grandes phases d'exploitation physique, alors esquissées, ne conviennent plus totalement pour définir, aujourd'hui, la démarche à adopter pour de tels projets au sein du Web. Les lettres de ces corpus vont entrer dans une autre dimension. Nous parlerons ici d'éditorialisation et de patrimonialisation des correspondances d'écrivain, pour évoquer la double tension perpétuelle auquel le corpus épistolaire numérique sera confronté sur le Web. Ces notions apparaissent à partir de la numérisation des lettres de la correspondance active et passive des écrivains, car le dédoublement du matériau de la lettre est d'ores et déjà une manipulation documentaire et archivistique, en soi ; la première pour ces équipes scientifiques.

Cette mise en visibilité et en accessibilité des contenus épistolaires sur le Web s'inscrit au fondement de l'éditorialisation de ces corpus que nous étudions. Nous relatons ici d'« un processus consistant à enrôler des ressources pour les intégrer dans une nouvelle publication », sans pour autant « rompre le lien existant entre la ressource et son document d'origine, introduisant une rupture entre la nouvelle production et les documents sur lesquels elle se fonde »<sup>92</sup>. Le premier degré de l'appropriation des lettres par ces équipes nous apparaît ici. Elle émane de la volonté des équipes à dématérialiser ces documents, afin de les détacher de l'unicité de leur support, qui freinait jusqu'alors leur exploitation sous une forme structurante. La numérisation des lettres s'inscrit dans la quête de sens originelle qu'ont rencontré les archivistes du XX<sup>ème</sup> siècle, face à la voluminosité et à l'indifférenciation de l'information de tout fonds épistolaire.

Mais dans le même temps, le statut de ces lettres, en tant que documents rares et anciens, a pu par certains aspects freiné l'usage de la lettre à numériser, chez ces équipes. scientifiques Soulignons que, souvent, dans les recherches menées en SHS, « la logique patrimoniale, du point de vue de l'exigence scientifique, outrepassse ses prérogatives en érigeant en dogme la préservation matérielle de l'objet au détriment de son étude »<sup>93</sup>. Il y' a donc bien, si ce n'est une dichotomie, du moins une mutuelle exaspération, entre ces deux directions auxquelles se résoudre face à une lettre d'écrivain manuscrite non encore numérisée. Afin de pérenniser et de valoriser le fonds de correspondance d'un écrivain sur le Web, les chercheurs devaient néanmoins les concilier toutes deux. On peut relever en cela une préséance du patrimonial sur le scientifique dans le rapport à ces sources, qui pourrait tendre à s'accroître si l'on prend en compte l'éphémérité du support de la lettre. C'est l'enrichissement documentaire des correspondances qui donnera aux lettres numérisées leur pleine dimension, n'étant plus dès

<sup>91</sup> Ibid.

<sup>92</sup> BACHIMONT, Bruno. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Hermes Science Publications, 2007, p.21.

<sup>93</sup> Ibid. p.190.

lors de simples documents de substitution. Les équipes scientifiques avaient à prendre conscience de ces rapports étroits dès l'étape de numérisation des manuscrits.

«La numérisation de la surface écrite ne fournit pas tous les éléments nécessaires à l'identification matérielle d'une lettre. Le traitement des données de l'analyse matérielle suppose un protocole spécifique, dont les résultats quantitatifs ou descriptifs peuvent en revanche être associés à l'image numérique du document »<sup>94</sup>.

Cela ne se produira pas, dans la plupart des cas, sans remettre en cas le bien-fondé de leur entreprise. Il faut souligner ici que « la numérisation bouleverse également notre vision de la propriété des matériels et, par conséquent, de celui qui a le droit et la responsabilité de les préserver. »<sup>95</sup>. Nous voyons que la numérisation d'une lettre manuscrite s'inscrit dans une démarche collective et sur le long terme. C'est pourquoi les équipes scientifiques du Centre Flaubert et de l'ITEM avaient besoin d'une intervention, d'un appui extérieur pour la mise en œuvre de cette étape. Ainsi, c'est Huma-Num qui a financé la numérisation des manuscrits des lettres inédites écrites par Gustave Flaubert. Du reste, le Département des manuscrits de la BNF a mis à disposition du Centre Flaubert ceux qui étaient d'ores et déjà numérisés sur la bibliothèque numérique Gallica. Une fois la lettre dématérialisée se posait alors la question de sa gestion, de sa préservation, et de son éventuelle mobilisation sur le Web. Auquel cas, cette exploitation devait l'entraîner à franchir des frontières poreuses et à enchevêtrer plusieurs sphères de circulation sur le Web, que nous nous attacherons à décrire à présent.

Il y' avait tout d'abord sphère publique qui englobe les documents ayant fait l'objet d'une publication. Les correspondances de Flaubert et de Zola que nous étudions dans notre travail ont comme nous l'avons vu chacun fait l'objet d'une édition et ont de fait intégrées cette sphère. Ces documents sont amenés à se confondre dans l'espace infini du Web car ils font partie d'un patrimoine à sauvegarder.

Venait ensuite la sphère collective, qui va inclure les fonds de documents réservés précisément à un groupe de travail, à une communauté de chercheurs prenant part à un projet culturel, telles que les équipes Flaubert et Zola. Avec leurs demandes de consultation et d'emprunt des manuscrits de lettres numérisés à la BNF, ces deux groupes de chercheurs se posent dans une certaine posture d'exclusivité documentaire. Le Centre Flaubert a ainsi été en relation avec Gallica, la bibliothèque numérique de la BNF, qui donne l'accès aux chercheurs à ses collections libres de droits ou dont les droits ont été négociés avec les ayants droit. Ces documents sont indexés dans Gallica, mais restent accessibles sur le site du partenaire, via le protocole OAI-PMH, que nous décrivons plus en détails ultérieurement. Cependant, la BNF n'avait numérisé aucune des lettres de Flaubert datant de 1830 à 1917. C'est pourquoi le Centre Flaubert a du également se charger de l'envoi en numérisation des lettres de jeunesse de l'écrivain que les chercheurs s'étaient procuré, pour la plupart, à la bibliothèque de l'Institut. Huma-Num a intégralement financé et s'est chargé de cette opération. De même, lorsque l'équipe recevait des lettres inédites, provenant de la correspondance passive de l'écrivain, elle décidait si oui ou non celles-ci pouvaient faire l'objet d'une numérisation. Ce n'est pas la pertinence de ces lettres qui devaient être évaluées, mais le fait qu'elles avaient bien été envoyées initialement à l'écrivain, à l'aide des formule d'adresse et des informations concomitantes de la lettre. Ces configurations diverses ont rendu la qualité des versions numérisées des lettres assez inégale,

<sup>94</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>95</sup> ROSENWEIG, Roy. « Pénurie ou abondance ? Préserver le passé à l'ère du numérique ». *Read/Write Book 2 : une introduction aux humanités numériques*. Marseille : OpenEdition Press, 2016. En ligne : <https://books-openedition-org.docelec.enssib.fr/oepl/226>

en fonction de la source et de l'ancienneté de la lettre. Nous repérons ainsi rapidement, sur la plateforme du Centre Flaubert, quels manuscrits sont issus des mauvaises photocopies anciennes, d'images basse résolution ou des excellentes numérisations actuelles. Concernant l'équipe Zola, nous releverons la collaboration menée avec le D<sup>r</sup> Brigitte Émile-Zola, arrière-petite-fille du romancier, qui conservé jusqu'à récemment une grande partie des lettres reçues par l'écrivain. Ce pan de la correspondance d'Émile Zola demeurerait donc inconnu du grand public. Le D<sup>r</sup> Brigitte Émile-Zola a donné son accord à l'équipe Zola pour la numérisation et la mise en ligne de ces lettres sur une plateforme publique. Ces discussions devraient prochainement mener à une convention écrite. Il y a à travers cette collaboration l'idée d'assumer le statut de Zola, en tant qu'écrivain et personnalité publique, qui considérerait que la vérité ne pouvait que se répandre qu'à travers une transparence totale de l'information. La reproductibilité de ces morceaux de vérité que sont les lettres de Zola, à travers leur numérisation, devait ainsi être encadrée. C'est donc dans de telles situations transitoires, au cœur des interactions interpersonnelles, scientifiques et institutionnelles, que la lettre d'écrivain s'inscrit proprement dans une sphère collective.

Nous devons enfin évoquer la sphère privée, qui correspond à l'espace de documentation propre au chercheur. Au sein de ces équipes scientifiques, cette sphère s'apparente au monde de l'immédiate proximité entre la lettre et le chercheur qui la parcourt, qu'elle soit manuscrite ou numérisée.

Ces deux projets nous enseignent que pour que l'édition numérique d'une correspondance d'écrivain arrive à maturation, il doit nécessairement s'articuler autour de ces espaces évoqués, à la fois publics et privés, physiques et virtuels... Aujourd'hui, l'interpénétration de ces sphères est « rendue possible par la facilité matérielle donnée à chacun de constituer de manière autonome des collections de documents en gérant, grâce à la numérisation (...) le passage de ces collections entre les trois contextes de médiations notés »<sup>96</sup>.

En définitive, la numérisation d'une lettre d'écrivain aura pour objectif de l'objectiver, en offrant une représentation à la fois distanciée et intime des traces écrites laissées par l'écrivain. Plus la restitution de sa matérialité sera efficiente, et plus le chercheur bénéficiera d'un accès privilégié à la phase d'écriture épistolaire, propre à chaque auteur. Comme nous l'avons souligné pour Émile Zola et Gustave Flaubert, chacun de ces documents a une histoire à raconter, au-delà de son texte, à condition de l'extraire dans un premier temps du flux historique et social auquel elle était liée. La numérisation permet ce détachement et cette réhabilitation, grâce à la maniabilité et l'agencement des fac-similé de chaque lettre, que les chercheurs pourront manier au besoin.

« Numériser et éditer un manuscrit, dans notre domaine, suppose par conséquent la prise en compte des différentes dimensions de l'objet (documentaire, archivistique, génétique), dont le contenu comporte des informations de multiples natures qui devront être structurées (usage de descripteurs, de normes et de règles différentes. »<sup>97</sup>

Nous allons étudier à présent comment les chercheurs se livrent à des études séquentielles du texte de la lettre. Cette analyse génétique et littéraire se donne pour rôle d'attester de la justesse des opérations de numérisations de manuscrits épistolaires. La prise en compte de la linéarité du texte et de la chronologie de l'écriture épistolaire seront des points fondamentaux dans l'approche des chercheurs.

<sup>96</sup> PEDAUQUE T. Roger, *Le document à l'ère du numérique*, Caen : C&f Éditions. 2006, p.170.

<sup>97</sup> IDMHAND, Fatiha. RIFFARD, Claire. WALTER, Richard. « L'édition électronique de manuscrits modernes ». *Expérimenter les humanités numériques : des outils individuels aux projets collectifs*. Montréal : Les presses de l'université de Montréal, 2017, 166 p.

### II.1.3. Enjeux de l'« auctorialité » avec la transcription et de l'annotation numérique des correspondances

Comme nous l'avons vu avec l'apparat critique dans les éditions de correspondances en volumes, il est essentiel d'inscrire le texte épistolaire au sein d'un réseau de sources et de réflexions, avant de le mettre à la disposition du chercheur. Cela se vérifie d'autant plus pour une période aussi référencée que peut l'être le XIX<sup>ème</sup> siècle, comme nous l'avons entrevu dans notre première partie. Les infrastructures du numérique permettront un déploiement de ces transcriptions et annotations textuelles liées à l'épistolaire. Nous verrons dans quelle mesure les équipes Flaubert et Zola ont opté pour une démarche caractéristique concernant cette étape, en appliquant deux types de traitement technique adaptés à chaque fois aux lettres numérisées. L'équipe Zola a choisi de procéder à leur *OCRisation* tandis que le Centre Flaubert s'est tourné vers le *crowdsourcing*.

L'équipe Zola s'est donc de son côté appuyé sur l'édition moderne des correspondances de l'écrivain, d'ores et déjà numérisées par le Centre d'étude sur Zola et le naturalisme et disponible en format PDF. Les chercheurs souhaitaient donc extraire ces données standardisées de chaque lettre et en obtenir des documents modifiables de type Word ou Excel. Pour cela, il leur était nécessaire d'utiliser un OCR, c'est-à-dire un logiciel de reconnaissance optique de caractères, qui traduit les images de textes imprimés ou dactylographiés en fichiers numérique de texte. Cette technique est réservée aux textes dont la police de caractères est uniforme et n'est donc pas adaptée à la nature cursive des caractères manuscrits. Quelques réussites ont toutefois été constatées à ce niveau, à condition de traiter peu de caractères à la fois et que ceux-ci aient rédigé par une seule personne. L'équipe Zola s'est alors renseignée sur l'utilisation du logiciel OCR d'Huma-Num, intitulé ABBYY, qui était la solution la plus optimisée dans cette configuration. Avec une plateforme de stockage adaptée pour accueillir les fichiers texte avant leur *OCRisation*, nommée ShareDocs, l'opération a pu être menée à bien, avant que ces nouveaux documents numériques ne puissent transiter vers leur plateforme d'archivage numérique attitrée, comme nous le verrons par la suite.

Il était néanmoins compliqué d'apprendre à ces logiciels à reconnaître des connexions intellectuelles spécifiques, pour traduire notamment certains traits d'esprit de la part des écrivains dans leurs lettres, que seuls des lecteurs avertis de ces écrivains seraient capables de déceler... C'est pourquoi une chercheuse de l'équipe Zola a procédé à la relecture des lettres ayant été soumises à cette opération, afin de rédiger un petit guide de recommandation pour la relecture des volumes suivants. Ces pratiques nous montrent qu'au fil de l'avancée de ces projets, la différence entre l'expert et l'amateur d'un sujet tel que l'épistolaire sur le Web tendra à s'amenuiser, au profit d'une vision collective et d'une méthodologie commune de la diffusion intellectuelle.

Cette tendance devient d'autant plus apparente à la lumière du projet du Centre Flaubert. Les responsables du projet ont souhaité reprendre la transcription des lettres de l'écrivain à leur point initial, à savoir en partant des manuscrits de lettres à leur disposition. Ce qui a justifié cette décision était le souhait de proposer une édition qui soit en majorité nativement numérique. Cela n'était pas un manque de confiance adressé aux éditions papier antérieures de la correspondance de Flaubert. Soulignons que les lettres présentes dans les éditions en volume de Jean Bruneau n'avaient pas toutes établies à partir des autographes de l'écrivain,

recueillies et archivées entre temps au Centre Flaubert. Cela a poussé l'équipe scientifique Flaubert à quelques corrections de principe et à des redatations concernant ces documents. En outre, ce ne sont pas moins de 172 lettres inédites ( 140 lettres de Flaubert, 32 lettres destinées à Flaubert) qui ont été retrouvées durant la réalisation de ce projet et retranscrites manuellement, en format texte. Ces lettres ont également été relues plusieurs fois, avec l'appui visuel des autographes de l'écrivain, qui permettaient de vérifier l'authenticité du document étudié.

Les responsables du projet ont pour ce travail fait appel à une vingtaine de transpositeurs. Ceux-ci pouvaient s'appuyer sur l'image numérisée du manuscrit et sur les outils de référencement numérique en cours d'élaboration au sein de la plateforme du Centre Flaubert. C'est ainsi que l'index des noms propres s'est enrichi au fil de la transcription des lettres, de même que le moteur de recherche plein texte de la plateforme. Chaque transpositeur progressait par lot de 10 lettres avant une mise en commun collaborative de leurs résultats. Au sein de ce projet, l'objectif du « chercheur collectif » est de mettre à la disposition du public les matériaux de travail de l'écrivain entourant la production d'une lettre au XIX<sup>ème</sup> siècle. La taille des manuscrits, la texture de l'encre et les diverses traces écrites font partie de cet ensemble génétique sous-jacent à tous les fonds de correspondances d'écrivain. Mais pour cela, il faudra « mettre l'écriture en acte »<sup>98</sup>, en respectant les diverses phases de la production et de la circulation épistolaire. Car si l'auteur ne peut que s'effacer derrière son fonds de lettres constitué au fil des années, l'émergence du texte épistolaire sur le Web marque, quant à lui, l'émergence du principe d'auctorialité. Nous parlons ici d'une instance textuelle se surajoutant à lettre numérique et qui sera en évolution constante... Ce processus ne s'obtient qu'au prix du maintien du contenu des lettres dans leur altérité et d'une reformulation propre au transpositeur, nécessaire afin d'accéder à toute l'intelligibilité du texte. C'est ainsi que « l'appropriation réussie est celle qui parvient à dégager le pôle objectif du texte dans son ipsité du pôle subjectif de la compréhension »<sup>99</sup>.

De tels projets culturels au long souffle recourent donc au *crowdsourcing* afin de maximiser leur chances d'une restitution contextualisée de la lettre. Le *crowdsourcing* consiste à « déléguer les contenus aux individus ou aux collectivités grâce à la technologie, et de penser qu'ils feront progresser ainsi l'activité et les contenus du Web en répondant, en groupe, au défi proposé »<sup>100</sup>. Cette logique avait d'ores et déjà été adoptée par les chercheurs du Centre Flaubert pour la transcription des brouillons du roman de l'écrivain intitulé « Madame Bovary ». Le roman était en effet au programme du Baccalauréat des Terminales Littéraires au cours de l'année scolaire 2014-2015. Une dizaine de professeurs de l'éducation Nationale et leurs élèves avaient répondu présents, afin de réaliser ce travail génétique, en s'appuyant notamment sur la correspondance Gustave Flaubert à Louise Colet dont de nombreux éléments mettent en lumière l'entreprise littéraire de l'écrivain comme nous l'avons vu. Il serait compliqué de recueillir les motivations propres à chacun de ces transpositeurs, au moment de se lancer dans ces opérations individuelles et complexes. Ces bénévoles nous enseignent que les correspondances de Flaubert et de Zola continuent, à travers le temps, à solliciter l'échange humain, au-delà donc de leur acte communicatif originel. Nous nous intéresserons pour notre

<sup>98</sup> LECLERC, Yvan. « Éditer en ligne les manuscrits de Flaubert ». Université de Rouen. 29 septembre 2015. En ligne : <https://webtv.univ-rouen.fr/videos/04-editer-en-ligne-les-manuscrits-de-flaubert-yvan-leclerc-ceredi-2/>

<sup>99</sup> BACHIMONT, Bruno. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Hermes Science Publications, 2007, p.172.

<sup>100</sup> NOIRET, Serge. « La *digital history* : histoire et mémoire à la portée de tous » *Read/Write Book 2 : une introduction aux humanités numériques*. Marseille : OpenEdition Press, 2016, p.152. En ligne :

part à la façon de reconnaître et de valoriser chacune de ces participation à l'échelle d'une plateforme numérique sur le Web.

Comme nous l'avons précédemment vu avec les études génétiques, on remarque beaucoup de ratures dans les lettres de Gustave Flaubert. L'équipe scientifique s'est donc interrogé, dès l'origine, s'il fallait ou non procéder à des interventions « scribales » au sein de la transcription, selon le degré de lisibilité de l'écriture de l'auteur. Apparaissait-t-il nécessaire de rendre compte de chaque rature de l'écrivain, d'archiver numériquement toutes ces lettres « en l'état » ? À partir de quel moment doit-on ajouter, supprimer ou corriger une information sur une lettre jugée incomplète ? De même, Flaubert avait pour habitude de mettre des majuscules à certains substantifs sans explication apparente... Tous les chercheurs ont alors développé une approche pragmatique de la transcription et ont finalement choisi de s'efforcer de reproduire au plus près toutes les caractéristiques de la page manuscrite, d'une façon pour ainsi dire mimétique. C'est le principe de la transcription littérale du texte, également appelée *diplomatique*. Concernant ce projet, le Centre Flaubert a opté pour cette transcription symbolique, représentant chaque élément significatif de la page manuscrite par des signes diacritiques ou d'autres conventions.

« Il faut penser la transcription aussi bien en fonction du public et des disciplines envisagées pour l'exploitation du corpus (grammairien et linguiste, historiens, littéraires, etc.) mais aussi en fonction d'un traitement homogène (et donc réaliste) de tout le corpus : les principes doivent s'appliquer sur toutes les lettres »<sup>101</sup>. Le premier principe de la transcription est génétique ; il vise à garantir l'exhaustivité des données des lettres pour leur interprétation. Le second est linguistique : il consiste à « repérer les unités linguistiques sur lesquelles s'appuient les transformations et repérage de marques de style. »<sup>102</sup>. Car si l'unicité de l'observable canonique des éditions papiers assurait la stabilité du texte épistolaire publié, les lettres numériques, que nous considérons ici sur le Web, font face à une pluralité syntagmatique et contrainte à ce niveau-là.

Concernant à présent la correspondance passive de ces écrivains, autrement dit les lettres qu'ils ont reçues, l'approche de chaque équipe a une nouvelle fois différé. L'équipe Zola, de Paris, a souhaité faire participer d'autres spécialistes de l'écrivain, qu'ils soient français ou étrangers, autour de ce corpus spécifique de lettres. La finalité de cette collaboration était un partage d'expertise, relatif à l'ensemble des aires géographiques d'où les lettres à Zola étaient expédiées (Europe de l'Est, Europe du Nord, Amérique du Nord, Amérique du Sud, Orient, etc...). Pour cela, la seule condition était la sélection des lettres par pays, leur répartition puis leur mise à disposition des chercheurs, sous un format numérisé. Par ailleurs, une grille d'analyse commune était fournie à tous les chercheurs afin qu'ils puissent ensuite confronter leurs résultats à la lumière d'un même prisme. Les transcriptions de ce corpus restent toutefois minoritaires (10% du total), car ce qui intéressait en premier lieu les chercheurs était la mise à disposition visuelle et publique de ce fonds inédit. La numérisation de ces manuscrits va donc primer dans un premier temps, afin de laisser le soin aux chercheurs de s'approprier progressivement ces documents inédits et de les enrichir à leur tour.

<sup>101</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.15. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>102</sup> CRASSON, Aurèle. MASPERO, Janeta. « Annotation collaborative en ligne de l'archive manuscrite ». Genesis, 27 | 2006, pp.165-166. En ligne : [https://www.persee.fr/doc/item\\_1167-5101\\_2006\\_num\\_27\\_1\\_1398](https://www.persee.fr/doc/item_1167-5101_2006_num_27_1_1398)

Pour ce qui est de la correspondance passive de Gustave Flaubert, ce sont 1111 lettres<sup>103</sup> rédigées à l'intention de Gustave Flaubert qui ont été exclusivement retranscrites pour leur mise en ligne sur le Web. Celles-ci ont d'ores et déjà intégrées l'index thématique de la plateforme. Attardons nous à présent sur cet outil élaboré par Danielle Girard, qui se propose de cerner les différents sujets abordés par l'écrivain dans ses lettres, qu'ils soient métaphysiques, sociétaux ou relatifs à la composition littéraire. Sa réalisation est le fruit direct du travail des transcripteurs. Près de 120 thèmes ont été recensés, qui abordent les grandes facettes de Flaubert, de l'homme comme de l'écrivain. On trouvera par exemple tout ce qui concerne sa santé, sa conception du style ou la genèse de ses œuvres. Dans ce contexte numérique, une étude thématique ne peut être productive que si elle redistribue ces unités de savoir selon une granularité approfondie grâce à l'encodage numérique des lettres, que nous décrirons par la suite. Il pourrait par exemple apparaître qu'un sujet ait été traité la plupart du temps avec un même destinataire, ce qui est susceptible de recontextualiser la correspondance toute entière de l'écrivain... Cette création entièrement numérique est un prolongement interactif de l'appareil critique présent dans les éditions papier de la correspondance de l'écrivain. L'index thématique est l'une des raisons expliquant le souhait de l'équipe Flaubert de se priver des annotations collaboratives sur leur plateforme.

Nous allons développer cette nouvelle approche de la transcription textuelle d'un fonds épistolaire sur le Web, en partant de ce choix du Centre Flaubert. Cela leur permettait en effet d'éviter d'intégrer toute représentation fautive au travail de l'équipe scientifique, qui se garantit en cela du sérieux de sa retranscription en maîtrisant les tenants et les aboutissants du processus. D'un autre côté, elle se prive de suggestions d'internautes qui pourraient mettre davantage en lumière une lettre, avec un apport d'éléments de contextualisation neufs et originaux... Un lien de suggestions est toutefois disponible, afin de spécifier par mail à l'équipe du projet une éventuelle erreur de transcription ou un mot noté « illisible » parmi les lettres archivées. L'annotation est donc ce qui se rajoute au texte de la lettre. Elle peut être de nature explicative ou bien critique, comme pouvait l'être l'index dans les éditions papiers des correspondances. Dans les deux cas, elle commente et ne décrit pas.

« Le lecteur peut comparer une annotation avec l'information contenue dans l'image du document, dans la mesure où l'annotation ne se substitue pas à l'image. (...) Elles [les annotations] sont traitées par le logiciel interface comme des métadonnées qui indexent l'image en question et permettront aux utilisateurs suivants de localiser par mots-clés les pages indexées. »<sup>104</sup>.

Les équipes scientifiques se sont posés la question de l'utilisation éventuelle de ces annotations, avant même de choisir pour quel format opter, parmi celles dites d'identification, circonstancielle ou encore thématiques. Leur interrogation était la suivante : jusqu'où la production d'une somme de savoirs autour des correspondances d'un écrivain est susceptible de mener les chercheurs ?

Dans le cas de l'édition numérique CorrELEZ, les annotations consisteront dans un premier temps à retranscrire l'apparat critique des 11 volumes papier de la correspondance générale. Sa précision et sa conception savante aidant, il serait dommage de s'en priver, d'autant que les spécialistes de Zola ont coutume à l'utiliser dans leurs études depuis toutes ces années. Mais dans une autre mesure, l'équipe scientifique se voit dans l'obligation de proposer un nouvel appareil critique (caractérisées en tant que « notes secondaires ») afin de compléter éventuellement les notes existantes et surtout d'en ajouter de nouvelles pour

<sup>103</sup> Résultat enregistré le 28 août 2019. Ce nombre évoluera au fil des nouvelles transcriptions, dans le temps.

<sup>104</sup> ESSEVAZ-ROULET Baptiste, *La numérisation d'archives : des fondamentaux techniques aux programmes de numérisation*, Voiron : Territorial Éditions, 2010, p.74.

l'édition numérique des lettres inédites. En définitive, on ne peut se permettre de hiérarchiser les annotations entre elles, car toutes contribuent *in fine* à l'éclairage de l'œuvre de l'écrivain. Ce processus n'est pas pour l'instant en libre accès, c'est-à-dire collaboratif, mais réservée à l'équipe Zola. Les responsables du projet, songent toutefois d'ores et déjà à la possibilité offerte à l'utilisateur de proposer ce système. Là où l'édition papier d'une correspondance rendait toute annotation assez froide et figée, le numérique se distingue, avec tous les cheminements sémantiques induits par l'Hypertexte.

Nous voyons, avec ces deux projets, que la question de l'annotation collaborative est du côté de ces équipes scientifiques soit reléguée, soit en cours de réflexion pour une éventuelle adoption à moyen terme. Cela nous interroge sur la question de l'*Open Source* et du statut des outils collaboratifs en ligne, affiliés à une plateforme d'archivage sur le Web. Dans le cas des correspondances d'écrivains au sein desquelles, comme nous l'avons précédemment constaté, le lecteur peut se sentir crouler sous les détails et les concepts en tous genres, n'y aurait t-il pas effectivement le risque d'une surabondance d'éléments avec cette fonctionnalité supplémentaire ? Mais dans le même temps, nous avons aussi remarqué que dans le cas des correspondances passives, certaines personnes sont actuellement en possession de lettres et pourraient avoir leur mot à dire, quant à la numérisation et à l'enrichissement documentaire de celles qu'ils transmettront aux équipes scientifiques...

Les opérations de transcription des lettres que nous venons d'étudier consistent à séparer un corpus épistolaire, puis chaque lettre de ce corpus, en diverses unités d'informations. Celles-ci doivent dès lors être documentées pour ne pas se perdre dans la masse des données sur le Web, et cela se traduira la plupart du temps par une indexation numérique ; étape primordiale au sein de ces projets. L'émergence des relations entre chaque segment de la lettre donnera alors lieu à des associations dynamiques, au fondement du principe de l'Hypertexte... Nous allons aborder à ce titre l'indexation numérique de ces corpus de lettres d'écrivain.

## II.2. Indexation numérique des correspondances

Nous avons constaté que durant les premières étapes de ces projets, les chercheurs faisaient face à des choix de différentes natures selon les différents niveaux de l'édition envisagés. Ces interrogations vont redoubler une fois que les équipes auront en leur possession les manuscrits ainsi que les données textuelles extraites des lettres de leur corpus, qu'il adviendra de renseigner. Le processus d'indexation numérique des correspondances d'écrivain aura pour enjeu principal d'assurer l'interopérabilité des données des lettres, nécessaire en vue de leur mutualisation et de leur valorisation sur le Web. Les deux équipes scientifiques que nous étudions ont à ce titre été confronté à des choix techniques et technologiques déterminants, en fonction des objectifs respectifs de leur projet. L'édition numérique d'une correspondance doit obéir aux mêmes exigences que celle des éditions imprimées que nous avons établi dans notre première partie. Elle devra par ailleurs intégrer des impératifs techniques supplémentaires, comme le respect des normes du domaine numérique. Les projets du Centre Flaubert et de l'équipe Zola se basent à ce titre sur des solutions *Open Source*, comprenant des langage informatiques et des formats de données numériques normalisés. En effet, la « possibilité d'exploration libre de données produits selon un protocole validé scientifiquement constitue la première source de création de nouvelles connaissances »<sup>105</sup>. Nous nous intéresserons tout d'abord aux conditions primordiales d'une indexation numérique

<sup>105</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.11. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

pour ces deux correspondances du XIX<sup>ème</sup> siècle à la lumière des spécifications Dublin Core que nous détaillerons. Nous soulignerons ensuite l'intérêt d'amener ce procédé d'encodage numérique à un degré supplémentaire, grâce à l'encodage des lettres d'écrivains en TEI.

## II.1. Description minimale du contenu des lettres numériques à l'aide des spécifications de Dublin Core

Le protocole d'indexation numérique des correspondances peut rapidement s'avérer complexes pour les équipes de chercheurs en SHS n'étant pas habituées à ces opérations de descriptions des documents sur le Web. Ce protocole devra sans cesse s'adapter et s'appliquer tant aux pratiques éditoriales qu'archivistiques, en vigueur pour les infrastructures numériques, tout en considérant la coordination des usages patrimoniaux, scientifiques, pédagogiques et culturels des corpus numérisés. Cette réflexion collective quant à la description des données textuelles et multimédias, au sein de leur corpus épistolaire, a été une étape obligatoire pour les équipes Flaubert et Zola. Avant de considérer l'aboutissement d'un outil *ad hoc* d'archivage des lettres et à son intégration au sein d'un environnement de recherche, fallait-il encore pouvoir l'approvisionner... Il advenait donc dès l'origine de considérer l'hypertexte, nouvel espace de travail à leur disposition, comme une construction intellectuelle à maîtriser afin mettre leur corpus à son épreuve.

L'équipe du Centre Flaubert a privilégié l'encodage des données de chaque lettre en HTML. Tandis que le XML permettra de décrire, structurer et échanger des données encodées sur le Web, le HTML ne fera qu'afficher ces données. Le XML est extensible et permet de créer ses propres balises en fonction des données traitées. En HTML, les balises sont prédéfinies et donc figées, bien qu'elles permettent par ailleurs de développer une structure cohérente de l'information. En outre, il y' aura toujours la possibilité d'inclure des ressources multimédias dont les images des manuscrits de lettres, ou encore des formulaires de saisie et des programmes informatiques. Retenons que le HTML permet de créer des documents interopérables avec des équipements très variés de manière conforme aux exigences de l'accessibilité du web. De rendre visible, en résumé, un contenu proposé par une page web sur un navigateur web. Toutefois, l'écueil rencontré peut alors être un enfermement dans un format propriétaire des données épistolaires, qui freineraient dans une moindre mesure les perspectives d'interopérabilité et de pérennisation, autour de la plateforme d'archivage numérique qui les accueillera.

Mais derrière l'affichage de cette forme originale transcrite du texte épistolaire, encore brute, il apparaîtra toutefois essentiel de rendre le texte des lettres exploitable » et/ou visualisable pour un autre objectif que la lecture « en version originale ». Cela renforce « l'intérêt de faire une transcription en utilisant des systèmes d'encodage comme le XML et la TEI. »<sup>106</sup>, comme nous le constaterons par la suite à la lumière du projet CorrELEZ, qui développe actuellement un schéma TEI à cet effet.

Il est néanmoins possible d'agréger finement et de diffuser, dans un cadre numérique, des informations sur les lettres à partir d'une description minimale mais exigible en la matière. Le Centre Flaubert a ratifié, dans son entreprise éditoriale et archivistique, l'encodage des champs de leurs lettres selon les spécifications DublinCore. Celles-ci ont été prescrites, depuis 2003 à

<sup>106</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.15. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

l'échelle internationale, avec la norme ISO 15836<sup>107</sup>. Il s'agit là des métadonnées permettant une description minimale pour toute ressource électronique disponible sur la toile. Concernant le Centre Flaubert, cette opération a été réalisé par l'informaticien attiré du Centre, en accord avec les responsables du projet. Car dans le cadre de ces projet collaboratifs, les chercheurs ne sont pas toujours rompus à ces pratiques purement informatiques. Au sein de l'équipe Zola, aucun membre de l'équipe ne s'est dédié spécifiquement à ces opérations d'encodage. Soulignons à ce propos le « trait particulier de l'archive numérique : elle est aussi souvent, voire davantage, la trace d'un travail de recherche, plutôt qu'une ressource autonome et exploitée par les chercheurs »<sup>108</sup>. En cela, le nombre restreint de champs d'une lettre d'écrivain, à encoder grâce au Dublin Core, a l'avantage de ne pas décourager les collaborateurs du projet, souvent peu férus d'informatique.

Ce sont les composantes de la plateforme de travail des équipes sur le Web qui induiront dans un second temps les modalités de l'encodage numérique des lettres. Dans le cadre d'EMAN pour l'équipe Zola, les chercheurs ont pu non seulement créer leur collection des métadonnées Dublin Core, mais également d'autres comprenant des métadonnées personnalisées. L'équipe a ainsi pu ajuster ces métadonnées descriptives supplémentaires, selon les spécificités des lettres d'Émile Zola. Ainsi, ils ont pu créer des balises complémentaires pour la presse ou encore pour les documents pédagogiques qui entouraient la figure de Zola dans ses propres lettres. Ce standard permettait aux chercheurs d'extraire automatiquement de nombreuses informations sur le document, ainsi que reconstituer les relations qui existent entre documents... En d'autres termes, enrichir un fonds de correspondance et le valoriser revient à associer à chacune de ses unités documentaires publiées des informations descriptives, administratives et structurelles....

« Cela étant, on note aussi une volonté de répondre à l'enjeu de l'accès technique, toujours via la documentation. Effectivement, les métadonnées ne sont pas uniquement descriptives, elles visent aussi à décrire le dispositif, le format, le logiciel. On appelle ces métadonnées, métadonnées de préservation, et ces informations sont précieuses pour pouvoir maintenir un accès technique au texte. »<sup>109</sup>.

Ce type de métadonnées nous intéresse particulièrement concernant ces deux projets d'édition numérique des correspondances, car elles préparent le terrain de l'archive numérique épistolaire, dont nous souhaitons dresser l'émergence et le maintien dans le temps, au sein de ce travail. Ces métadonnées engloberont tout d'abord les métadonnées structurelles, présentant par exemple des modèles de données et des données de référence dans le domaine de l'épistolaire. Viendront ensuite les métadonnées administratives, qui renseignent explicitement la lettre numérique afin d'aider à la gestion de cette ressource.

Intéressons nous plus en détail à ces champs Dublin Core, à leur souplesse au contact des corpus de lettres d'écrivain, qui viendront révéler ces grandes familles de métadonnées. L'élément <dc:date> va permettre de poser les jalons chronologiques entourant le cycle de vie de la lettre, d'abord manuscrite puis numérisée. Les chercheurs pourront renseigner la date de rédaction de la lettre, de sa réception, de son acquisition par une bibliothèque, jusqu'à celle de sa dématérialisation, et de ses enrichissements numériques successifs. Le rôle des transcripteurs et des

<sup>107</sup> Pour plus de renseignements quant à cette norme ISO 15836 : <https://www.iso.org/fr/standard/71339.html>

<sup>108</sup> BOUHAI, Nasreddine. SZONIECKY, Samuel. *Intelligence collective et archives numériques*. Londres : ISTE Editions. 2017, p.40.

<sup>109</sup> HENRY, Clara. *L'archivage de la littérature numérique en ligne*. Enssib : Mémoire ARN, 2018, p.61. En ligne : <https://www.enssib.fr/bibliotheque-numerique/documents/68377-l-archivage-de-la-litterature-numerique-en-ligne.pdf>

annotateurs des lettres pourra par exemple être valorisé, grâce à cette balise.

L'élément <dc:subject> permettra d'opérer la classification des thématiques abordées au sein d'une lettre. Ce processus de caractérisation de l'information est susceptible de faciliter, comme nous l'avons vu avec le Centre Flaubert, la création d'un index thématique, à partir des mots-clés encodées au sein de cette balise.

L'élément « dc:format » va poser quant à lui la question de la description du format physique et électronique du document épistolaire. Ces projets, qui entendent présenter un fac-similé des lettres manuscrites numérisées, ont tout intérêt à décrire ainsi la matérialité de la lettre et à renseigner les conditions de sa numérisation. Cela impliquera parfois, notamment pour la correspondance active d'un écrivain, de remonter jusqu'aux institutions de conservation des lettres. Car, comme nous le verrons par la suite, le format d'une lettre numérisée implique des obligations descriptives légales, tels que les crédits photographiques... En attendant, les lettres d'écrivain à la disposition des équipes se prêtent à ces descriptions numériques élémentaires, en étant un matériau dont le format est unique en son genre.

Sans avoir à lister les 15 balises prescrites dans le format Dublin Core, nous nous apercevons toutefois que les chercheurs pourraient vite se sentir limités dans leurs descriptions. Leur connaissance sur le sujet ne pourront en effet pas toutes se fondre dans l'établissement de ces nouvelles notices documentaires numériques. Car « lorsque nous utilisons les éléments qualifiés de *Dublin Core*, nous avons encore besoin des sémantiques plus riches surtout dans un domaine aussi complexe que le patrimoine culturel »<sup>110</sup>.

C'est cette couche supplémentaire d'information du contenu des lettres qui va particulièrement nous intéresser ici. Car comme nous le verrons par la suite pour l'équipe Zola, la plateforme Omeka va permettre aux chercheurs, qui lui fourniront les données extraites de chaque lettre, de se concentrer sur le contenu de celles-ci. Les étapes de programmation n'entreront presque pas en compte au cours de cette étape, grâce aux plugins de la plateforme. Ces simplifications procédurales vont de pair avec l'utilisation progressive du Dublin Core. Dans le cas de l'équipe Zola, les chercheurs ont décidé de se concentrer sur un corpus d'environ 2000 lettres envoyées à Zola, depuis les pays étrangers, au moment de l'Affaire Dreyfus. Cela nous permet de mieux comprendre la réception de l'Affaire à l'étranger et dans différents milieux sociaux. La correspondance passive fait l'objet d'une description via les champs DublinCore avec numérisation de la lettre depuis une photocopie ou d'après la lettre originale. La correspondance active donne les métadonnées DublinCore et la transcription de la lettre grâce à l'*OCRisation* des volumes de publication papier, avec la numérisation du document quand celle-ci est disponible. La transcription s'impose ici comme un effort pour donner à lire, et notamment aux chercheurs, qui pourront ainsi s'appropriier le texte et les dires de l'écrivain plus facilement.

L'objectif final est la mise en ligne du corpus numérisé sur une plateforme Oméka, respectant des normes d'inter-opérabilité OAI-PMH. Ce sigle désigne le "protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes" et s'adresse aux acteurs impliqués dans la valorisation des collections culturelles sur le Web, leur permettant d'échanger des métadonnées, pour multiplier l'accès aux documents numériques. Nous voyons qu'il importe de problématiser chaque champ de la lettre afin de mieux le décrire, en les segmentant et en les traitant séparément. Dans une perspective d'étude textuelle, « le segment est en quelque sorte l'unité syntagmatique et sémantique mobile, emboîtable et pertinente que manipule l'« auteur ». De ce point de vue, l'« auteur » ne travaille pas avec des données

<sup>110</sup> BOUHAI, Nasreddine. SZONIECKY, Samuel. *Intelligence collective et archives numériques*. Londres : ISTE Editions. 2017, p.163.

mais des segments qui, dès l'origine, font tous l'objet des comportements de recherche »<sup>111</sup>. L'optique est de se retrouver en face, non plus de pages manuscrites numérisés, mais bien de paragraphes encodés, accessibles en recherche plein-texte sur le web.

La lettre bénéficie de composantes irrévocables, tant dans sa forme que matérielle que dans celle de son contenu ; l'écrivain se résolvant à la linéarité de l'écriture épistolaire. En cela, son encodage numérique nous enseigne d'autant plus clairement que la signification et la normalisation des éléments sont plus importantes que l'instanciation spécifique en XML ou autre.

Ainsi, l'utilisation des balises HTML, pour laquelle le Centre Flaubert a opté dans son projet, n'est pas préjudiciable, même si elle limite considérablement le champ de l'encodage numérique des données de la lettre. Toutefois, leurs relations aux autres lettres du corpus sera parfois plus ambiguë ou il sera difficile de la rattacher à son contexte spatio-temporel. Cela est d'autant plus vérifiable dans la cadre de l'archivage des manuscrits d'écrivains...

« Plus on précise l'information sur des concepts non génériques, plus on referme le manuscrit sur sa complexité. Décider en revanche des objets qui font consensus permet de parler de la même chose mais à travers des présupposés différents et des descriptions fouillées. La terminologie est donc primordiale même si elle donne lieu à des variantes ou des déclinaisons. »<sup>112</sup>.

C'est pourquoi le choix d'associer l'encodage des informations en *Dublin Core* à des informations encodées, dans des champs personnalisés, en XML, a tout lieu de se révéler pertinent dans la plupart des cas. Comme nous le verrons par la suite, des plateformes telles qu'Omeka autorisent plusieurs types de métadonnées numériques, afin de rendre la tâche d'indexation plus souple pour les chercheurs, comme l'ont constaté ceux de l'équipe Zola.

Les balises «auteur», «destinataire», «lieu», «date», étant considérées comme balises de «structure» selon le format Dublin Core, on peut ainsi étendre la nomenclature de ces balises de structure à l'étiquetage d'autres constituants génériques des lettres. Citons à ce titre les formules d'adresse et d'adieu, la signature, les Post-scriptum de la lettre. Mais ces descripteurs pourraient encore être plus pointus en informant sur le « genre » de lettre dont il est question : lettre de condoléances, de félicitations, accusé de réception, ordre de ventes à la Bourse, lettre de vœux, dédicaces, lettres administratives... Comme nous l'avons vu dans notre première partie avec les analyses codicologiques, ces structures épistolaires sont généralement très codifiés pour tout œil averti. Se cantonner exclusivement à un encodage en Dublin Core revient à limiter le niveau de structuration des métadonnées et les informations des lettres pouvant être récupérées. De plus, « l'usage d'identifiants "standardisés" en particulier dans les notices en Dublin Core n'est pas stabilisé, ce qui rend difficile la possibilité d'identifier des doublons et d'enrichir les notices en conséquence. »<sup>113</sup>

Pour l'équipe Zola, d'autres questions pratiques se sont posées, telles que la conservation des notes scientifiques d'origine, liées aux éditions papier de la correspondances. La création de nouvelles notes, sous la forme d'un appareil critique innovant, aurait été un *challenge* de taille pour les chercheurs. En outre, l'édition papier fait des renvois de page en page ; modèle

<sup>111</sup> BERT, Jean-François. RATCLIFF, J. Marc. *Frontières d'archives : recherches, mémoires, savoirs*, Paris : Archives contemporaines, 2015, p.20.

<sup>112</sup> CRASSON, Aurèle. « Archives manuscrites littéraires : l'apport du numérique pour l'édition et la recherche scientifique », *Genesis*, 30 | 2010, p.45. En ligne : <http://journals.openedition.org/genesis/112>

<sup>113</sup> MAHE, Annaïg. Prime-CLAVERIE, Camille. « Le défi de l'interopérabilité entre plates-formes pour la construction de savoirs augmentés en sciences humaines et sociales ». ISTE éditions, 2017, p.15. En ligne : [https://archivesic.ccsd.cnrs.fr/sic\\_01511618/document](https://archivesic.ccsd.cnrs.fr/sic_01511618/document)

qui devient obsolète avec le numérique. Il a alors fallu transposer les numéros de page en URL afin d'entrer progressivement dans le Web Sémantique. Les éditions papiers pouvaient en effet présenter une lecture documentaire de la correspondance d'un écrivain, là où la dimension éminemment relationnelle du texte épistolaire se perdait. L'encodage numérique des lettres leur insuffle en quelque sorte une nouvelle fois leur densité mnésique.

Les éléments d'une description minimale des correspondances d'écrivain sur le Web permettent donc de « faire émerger des « saillances perceptives » comme autant de « diégèse mémorielle » qui serviront de base pour l'éditorialisation de l'archive »<sup>114</sup>. Nous en revenons à cette notion clé pour la phase post-numérisation des lettres d'écrivains.

Dans ce chapitre, nous avons esquissée une liste argumentée des descripteurs nécessaires pour les lettres à indexer sur le Web, et montré la systématisation des traitements éditoriaux et techniques à adopter avec les corpus épistolaires. Il nous reste à présent à poser les questions de pérennité et d'interopérabilité de ces corpus à travers un encodage plus détaillé de leurs métadonnées. La TEI nous semble être à ce titre un système pertinent pour développer les axes de l'exploration scientifique informatisée, afin de stimuler à terme la mise en réseau des informations des lettres archivées sur le Web... Car derrière ces opérations éditoriales, caractérisées par la mise en lignes de lettres d'écrivains du XIX<sup>ème</sup> siècle, une autre optique se dessine pour ces équipes scientifiques. Celle de pérenniser leur corpus de lettre et de lui attribuer une nouvelle dimension historique, par l'entremise des plateformes d'archivage sur le Web. C'est ainsi qu'« éditer à l'ère du numérique et du multimédia n'a rien d'une activité « positiviste » allant de soi, mais recouvre des enjeux politiques que les possibilités de l'édition hypertextuelle en ligne font singulièrement apparaître. »<sup>115</sup>.

## II.2.2. Perspectives d'un processus d'encodage numérique plus détaillé des correspondances d'écrivain

Comme nous l'avons précédemment défini, l'encodage est l'action de structuration d'un texte numérique avec des balises de différents formats de type HTML, TEI, ou encore EAD comme nous le verrons prochainement. Chaque format possède son propre langage mais également ses propres finalités. Avec ces corpus, l'encodage concernera aussi bien des spécifications de mise en forme du texte épistolaire que des indications de structure matérielle de la lettre, ou encore des interprétations sémantiques de son contenu.

L'apparition de la TEI et du langage XML décuplent les possibilités d'annotation et rendent pensable la production d'éditions web-centrées et facilement accessibles d'une correspondance d'un écrivain. L'émergence de ce format est très récente, et le caractère expérimental, dispersé, de leurs application ne permet pas toujours d'avoir une vue globale sur les corpus numériques en cours de constitution, notamment dans le domaine fécond et volumineux de l'épistolaire. Toutefois, il présente des avantages sur le long terme, en rapport au balisage HTML d'un corpus de lettres. Les chercheurs qui déploieront des schémas en TEI pourront identifier les blocs textuels et les illustrations de chaque lettre, pour ensuite les nommer dans la terminologie de spécialité qui conviendra à leur besoin. Ils pourront par

<sup>114</sup> BOUHAI, Nasreddine. SZONIECKY, Samuel. *Intelligence collective et archives numériques*. Londres : ISTE Editions. 2017, p.19.

<sup>115</sup> LERICHE Françoise, « Quelle édition pour quel public ? Les avatars de l'édition électronique de la correspondance de Proust », *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008, p.60. En ligne : <https://journals.openedition.org/recherchestravaux/89>

ailleurs délimiter les blocs textuels identifiés, et les caractériser en fonction de leurs propriétés. C'est ainsi que la plateforme d'archivage des correspondances pourra être constituée, de façon interactive et concertée entre les chercheurs.

Les responsables du Centre Flaubert ont rencontré des spécialistes en TEI mais n'ont pas estimé que ce type d'encodage numérique des données était judicieux pour ce travail. Comme entrevu dans notre dernier chapitre, le fil rouge de leur entreprise consistait en une retranscription nouvelle des lettres de l'écrivain, plus lisible et rigoureuse, que n'avaient pu l'être les publications en volumes. L'éclairage de leur contenu devait passer avant tout par la mise à disposition du fac-similé des lettres et de leur transcription écrite, scientifique, sur le même plan de visualisation. En cela, leur souhait s'est matérialisé et l'ergonomie de leur plateforme est un modèle du genre, élaboré par l'informaticien du Centre Flaubert.

« La lettre n'est jamais autoréférentielle (...) Pour la mise en place d'un système de référents, il faut savoir raison garder : ne pas multiplier les encodages pour des indexations infaisables, ne pas saturer le texte d'annotations référentielles au risque de perdre un des objectifs premiers de tout projet d'édition de corpus : donner à lire » (p.11)

Nous allons toutefois montrer, au regard du projet CorrELEZ, que d'autres pratiques numériques peuvent faire foi en la matière, en assurant l'accès, la diffusion et la conservation précise des lettres à l'aide des directives TEI. Ce standard international d'encodage numérique des données offre des possibilités pour ainsi dire indéfinies d'indexation et d'enrichissement numérique des lettres, avec des balises personnalisables. Celles-ci permettront par exemple de représenter les diverses formes de l'écrit que peut contenir une lettre et de faire participer l'utilisateur à cette dynamique proprement génétique. Nous sommes dans la tradition des annotations multiples des corpus de lettres, avec les écritures numériques qui viendront compléter celles des écrivains en quelque sorte. Nous allons relater plus concrètement les possibilités offertes par cet ensemble d'enchâssements documentaires, en décrivant tout d'abord sa structure initiale...

La balise <surface> permet de définir une surface d'écriture à transcrire, la balise <zone> délimite une parcelle de la page à encoder, tandis que la balise <line> identifiera les différentes lignes d'écritures de la lettre... Dès le principe, le cheminement de la lecture d'une lettre par l'internaute est d'ores et déjà préétabli au sein de ces trois balises. Chaque séquence devra être pensée et interprétée assez finement pour assurer non seulement la représentation visuelle mais également la construction sémantique inhérente à toute la correspondance. Le projet CorrELEZ n'a pas encore commencé l'encodage en TEI des lettres de Zola, mais un schéma a été développé par les responsables du projet, impliquant un encodage numérique simple mais suffisamment détaillé de la correspondance de Zola. Cette étape interviendra prochainement, l'équipe devant encore songer à l'outil à adopter pour ces indexations numériques. Il y avait une possibilité au sein de la plateforme de travail Omeka de coller le texte extrait de la lettre « océrisée », depuis Word, (bouton "Paste from Word") grâce à l'éditeur TinyMCE. Celui-ci permet en outre de conserver la mise en page propre au texte épistolaire de la lettre retranscrite. Cependant, cela supposait de copier-coller manuellement le texte dans le champ correspondant de chaque notice. Dans le cas des plusieurs milliers de notices comprises dans ce projet, cela pourrait s'avérer extrêmement chronophage pour le nombre restreint de chercheurs impliqués. Rappelons que le corpus de la correspondance active de l'écrivain comprend plus de 5000 lettres à encoder. Les chercheurs ont actuellement le choix entre deux méthodes d'indexation des lettres sous le modèle TEI. La première option est l'utilisation du logiciel Oxygen XML Editor. C'est une des solutions que proposait la plateforme Omeka, consistant à transformer le document Word de la lettre en un fichier XML-

TEI, ce qu'il est possible de faire très simplement avec le logiciel Oxygen. Omeka ne lisant pas le langage XML-TEI, les chercheurs ont mis en place pour les besoins de ce projet sur la plateforme EMAN, un *plugin* appelé « Transcript ». Celui-ci permet de transcrire directement dans le logiciel les fichiers numérisés et de les encoder ensuite en TEI sur la plateforme. Ces opérations spécifiques ont nécessité l'intervention d'un développeur, Vincent Buard, ce qui nous montre que comme pour le Centre Flaubert, l'aspect infrastructurel du numérique n'est pas à négliger... La deuxième option est l'utilisation d'une plateforme collaborative de transcription. L'équipe a notamment en vue une plateforme multi projets de transcription contributive et d'exploitation des données de masse intitulée TACT, conçue par le groupe ELAN de l'université de Grenoble. Les chercheurs de l'équipe ZOLA auraient alors la possibilité de déléguer la transcription et l'annotation des manuscrits numérisés des lettres de l'écrivain à des non-spécialistes, rejoignant par là même les initiatives en *crowdsourcing* du Centre Flaubert.

Maintenant que le contexte entourant le projet CorrELEZ est posé, revenons aux spécificités du langage XML-TEI et aux perspectives offertes par entremise, pour tout corpus épistolaire, que nous tenterons d'ébaucher. Ce langage a été développé par le Consortium TEI dans le but, d'une part, de structurer les métadonnées liées au texte numérique et, de l'autre, d'enrichir le texte publié avec des informations scientifiques (...) L'encodage des transcriptions des textes en TEI paraît alors évident lorsqu'il s'agira d'enrichir l'édition électronique des textes, dans la mesure où l'immense bibliothèque de balises proposées par la TEI complète l'édition scientifique électronique. Cela nécessite un temps d'apprentissage assez conséquent pour les chercheurs des équipes scientifiques, bien qu'il y ait également une possibilité de traduire d'autres formats de données dans ce standard à condition d'avoir normalisé les vocabulaires et contrôlé les formats utilisés au préalable. D'où l'intérêt d'une base en Dublin Core pour les indexations numériques plus détaillées des lettres...

### II.2.3. Le protocole TEI-XML à la lumière de la lettre d'écrivain à indexer : détails et développements

Il existe donc un balisage spécifique dédié à l'encodage numérique des lettres en TEI, que nous décrirons à présent, avec l'élément <correspDesc>, mis en place par le groupe nommé « Correspondance SIG » du consortium TEI. Cet élément TEI « sépare la description d'une lettre en deux divisions, l'événement d'envoi d'une lettre (<correspAction>) et le contexte social autour de cette lettre, y compris les autres correspondances connexes (<correspContext>). »<sup>116</sup> L'élément <correspAction> est consacré à la description structurée des noms, lieux et moments de la correspondance d'un écrivain, ainsi qu'à toute autre action l'ayant entourée. Cet élément englobe entre autres l'attribut @type décrivant la nature de l'action, avec plusieurs valeurs proposées : envoyé, reçu, transmis, redirigé, transmis... L'élément <correspContext> va donc faire quant à lui le lien avec d'autres éléments extérieurs à la production de la lettre à encoder, à commencer par une redirection vers la lettre qui lui répond, de la part du destinataire... Nous voyons ici que le principe de l'indexation numérique d'une lettre est de restituer les liens factuels, tout autant qu'intellectuels, la rattachant aux autres lettres du corpus. Ces possibilités numériques de contextualisation de la lettre ont intéressé à plus d'un titre l'équipe Zola, qui souhaite à travers un schéma TEI<sup>117</sup> somme

<sup>116</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>117</sup> Cf : Annexe 1 (p.90)

tout assez basique, donner aucune signification supplémentaire à la correspondance de Zola. Ces balises permettent en effet de comprendre une lettre comme un acte de communication tout autant que comme une source d'information, et de coder les données relatives à cet acte unique. La signification des données numériques à encoder primera en conséquence sur leur forme. « Ainsi le recueil et la constitution des corpus doivent-ils prendre en compte les spécificités textuelles, génériques et discursives des matériaux considérés, car ces spécificités en déterminent les régimes d'interprétation »<sup>118</sup>

Ce type de description idéal permet à l'éditeur de coder non seulement les éléments décrivant factuellement l'objet-lettre, mais également le contexte biographique, social et artistique ayant conduit à l'écriture de la lettre. Sans oublier l'indication des nouvelles productions épistolaires qu'elle aura engendré, historiquement, si nous souhaitons aller plus loin. Ces modalités de l'encodage numérique présuppose dès lors des constructions intellectuelles nouvelles à établir par ces équipes scientifiques. L'intégration des lettres à un environnement archivistique qui les dépasse a été, depuis le XX<sup>ème</sup> siècle une façon de les pérenniser via une problématisation constante de leur contenu. Cela se traduira, selon le format TEI, par une structuration spécifique de l'information de la lettre, pouvant donc « adopter de multiples styles de présentation, pour des publics multiples, sur la base d'un seul document racine sous-jacent. »<sup>119</sup>

Il est à noter que d'autres éléments de la TEI peuvent être mis en œuvre dans la balise <CorrespDesc>. Il y a par exemple, l'élément <ParticDesc>, qui décrit les participants de la correspondance (scripteurs, destinataires, etc.) et leurs liens avec l'écrivain (avec l'élément <relation> : parenté, amitié, etc). À ce titre, l'équipe scientifique du projet CorrELEZ s'est attachée à compléter les métadonnées des lettres en fonction des compétences linguistiques de chaque chercheur, en fonction des pays représentés au sein du corpus. Les lettres conservées au Centre Zola sont classées par pays, ce qui facilite la répartition des tâches pour les linguistes, en faisant des lots de photocopies pour chaque langue à traduire. Concernant les destinataires appartenant au cercle intime de l'écrivain, il a fallu remonter aux sources familiales, en contactant les descendants d'Émile Zola, afin de confirmer l'identité de chaque destinataire. Ces balises ont donc été pensées afin de structurer les champs de la correspondance de la façon la plus concentrée possible, quant au nombre de lignes d'encodage, et rigoureuse, quant à leur enchaînement lors de l'affichage sur le site Web. Ainsi, la disposition des caractères sur la lettre manuscrite peut être reproduite numériquement avec assez de justesse dans un forme texte.

Mais l'indexation numérique des lettres de ces projets ne serait pas complète si les chercheurs ne pouvaient pas décrire également l'image numérisée des manuscrits de lettres à leur disposition. Il existe les métadonnées de l'*International Press Telecommunications Council*<sup>120</sup> (IPTC). Ces jeux d'attributs de métadonnées sont en effet applicables à des fichiers texte, mais plus particulièrement à des images et d'autres sources multimédias. Il s'agit d'informations sur l'auteur d'une image et sur le sujet de la photographie (coordonnées, mots clés, URL...) à encoder selon le protocole TEI-XML. Pour le projet CorrELEZ, quelques 2 000 lettres manuscrites, de 1 à 5 pages environ, feront l'objet

<sup>118</sup> LONGHI, Julien. « Humanités, numérique : des corpus au sens, du sens aux corpus ». Questions de communication, 1 / 2017, p.14. En ligne : <https://www.cairn.info/revue-questions-de-communication-2017-1-page-7.htm>

<sup>119</sup> DUMONT, Stephan. « Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF) », Digiversity, 29 octobre 2015. En ligne : <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata->

<sup>120</sup> Pour plus d'informations au sujet de l'IPTC : <https://imatag.com/blog/2018/05/11/metadonnees-dimages-etat-lieux-inedit-inquietant/>

de cette description spécifique de leur version numérisée. Ce serait une erreur de penser que les images des lettres viennent se rajouter au texte, à seule fin de l'illustrer... Les deux se complètent dans l'archivage numérique de leur contenu, car pour extraire les données textuelles d'une lettre grâce à la TEI, il faudra penser au préalable leur alignement avec l'image de la lettre qui s'affichera devant l'internaute... Une balise des métadonnées IPTC permet d'afficher, à côté de l'image, la transcription écrite de la lettre selon le degré d'alignement souhaité par les chercheurs, à des fins de lisibilité. Le format IPTC est le format le plus souvent utilisé à l'international, pour informer quant aux crédits photographiques, toute présence d'une version manuscrite numérisée sur le Web. Ce qui s'avèrera notable et pertinent dans le cas des correspondances d'écrivains du XIX<sup>ème</sup> siècle, étant donné la diversité des lieux de conservation qu'ont connu par lesquelles ont transité les lettres de ces corpus, dans le passé. C'est un format de données, qui par ailleurs, permet de protéger les médias ainsi encodées. La sécurisation des données pour chaque matériau visuel de la lettre d'écrivain s'impose de plus en plus pour ces équipes scientifiques, car si « jusqu'alors la localisation et l'indication de la côte [de ces correspondances] faisaient foi, désormais on peut et on doit montrer : l'image fait preuve »<sup>121</sup>. C'est davantage sa représentation, plutôt que l'objet-lettre lui-même qu'il adviendra de considérer dans ces processus d'indexation numérique, où les chercheurs doivent indiquer en définitive une disposition spatiale et virtualisée de l'écriture épistolaire. De ces effets d'accumulation de l'information autour des lettres numérisées pourra alors surgir une sorte de systématisation dans l'approche des chercheurs. Car sur le Web, « afin de pouvoir travailler sur l'image, les chercheurs doivent souvent la créer aux-mêmes »<sup>122</sup>. Le protocole TEI-XML nous révèle donc, cette façon effective, les enjeux de la patrimonialisation des correspondances d'écrivains sur le Web.

Dans cette optique, il existe une autre façon de prolonger la matérialité des lettres via leur encodage numérique. Il s'agit de la description des diverses sources d'archivage de la lettre manuscrite et de ses exploitations sous format papier, dans le passé. Toutes les éditions des correspondances de Gustave Flaubert et d'Emile Zola pourraient alors être identifiées, et faire définitivement s'inscrire ce travail d'édition numérique dans une tradition documentaire, retraçable dans l'immédiat. Les chercheurs pourront à ce titre indiquer la nature des données de la lettre à encoder : si elles sont issues d'une retranscription manuelle ou technologique, si elle incluent des commentaires explicatifs au sujet des lettres... Cela sera permis grâce à la TEI et l'élément <sourceDesc>, suivi de son attribut @type. Celui-ci sera amené, avec l'évolution des éditions numériques de correspondances, à prendre en compte de plus en plus de valeurs. Celles qu'il ratifie déjà sont « print », « online » et « hybrid », afin d'indiquer la nature nativement numérique ou non des données manipulées par les chercheurs de ces projets. L'interopérabilité des lettres sur le Web prendrait alors sa pleine mesure, avec des possibilités équivalentes d'effectuer des liens entre diverses plateformes d'archivage de correspondances d'écrivains. Une ouverture des données épistolaires sur le Web qui serait à toute fin garant d'une valorisation par l'exploitation multi-contextuelle de ces informations capitales. C'est dans cette optique que de nos jours, de plus en plus d'archives numériques fournissent des enregistrements de métadonnées de leurs documents en ligne, ainsi qu'un identificateur de ressources uniforme (URI). Ces adresses URI ne sont que des séquences de caractères sans espaces, souvent sous la forme d'URL, qu'il sera possible de coder dans l'élément <correspDesc> du protocole TEI-XML, et d'amener ainsi l'épistolaire vers de nouvelles dimensions. La traçabilité des données des lettres numériques pourra alors être prises en compte. « Les identifications uniques de lettres dans une archive doivent être associées à un enregistrement des modifications et améliorations apportées à la transcription et à la représentation,

<sup>121</sup> ROUSTAN Mélanie, *La recherche dans les institutions patrimoniales : sources matérielles et ressources numériques*. Presses de l'Enssib, 2016, p.21.

<sup>122</sup> Ibid. p.185.

ainsi qu'à un enregistrement des modifications de format (par exemple, de SGML à XML). »<sup>123</sup>

Certes, les équipes Flaubert et Zola n'ont pas jugé utile, ou du moins tardivement, d'adopter ce protocole TEI-XML pour l'indexation numérique de leurs correspondances d'écrivains. Toutefois, leur présence dans le groupe de travail Consortium Cahier est un indicateur par mi d'autres que ces équipes restent ouvertes à ces propositions de description de l'information des lettres numériques. Nous avons dressé dans ce chapitre la nécessité de réfléchir aux modalités de tout processus d'indexation des lettres, à travers le respect des normes d'archivage numérique. Si celles-ci sont respectées, elles doivent amener les lettres à être interopérables, une fois leur présence validée par les équipes scientifiques, sur leur plateforme Web. Car « pour que ce qui est devenu une « archive » au sens propre puisse être utilisé par d'autres chercheurs, encore faut-il qu'elle soit accessible en ligne sur un serveur pérenne. »<sup>124</sup>. Ce point précis sera l'enjeu de l'ultime chapitre de cette partie où nous nous sommes proposés d'analyser l'existant des deux projets de correspondances numérique de Flaubert et de Zola.

## II.3. Exploitation des corpus épistolaires de correspondances d'écrivain sur les plateformes du Web

Face aux données numériques des lettres d'écrivains, la méthodologie des équipes Flaubert et Zola a évolué, une fois leur indexation réalisée. Les éléments saillants de la lettre ont pu être révélés grâce à l'encodage numérique et sont désormais forts d'une densité mnésique nouvelle. Celles-ci étaient désormais aptes à être non seulement classées et rendues visibles sur les pages Web, mais archivées de surcroît en respectant les standards du numérique. Nous allons donc dans ce chapitre faire un zoom sur cette problématique de l'accessibilité du corpus épistolaire sur le Web, à partir de ces portées d'entrées que sont les plateformes du Centre Flaubert et d'Archiz... L'évolutivité du corpus épistolaire dans les bases de données et les fouilles de données opérées par les chercheurs en son sein nous semblent être deux points primordiaux à aborder, en suivant cette perspective.

### II.3.1. Écueils et perspectives d'une base de données épistolaire

En décrivant les procédures d'indexation numérique de la lettre, il nous semblait essentiel de rappeler que les chercheurs ne devaient, à aucun moment de ce processus, perdre de vue le regard de l'utilisateur qui parcourra leur plateforme, in fine. Nous avons souligné les nouvelles perspectives mémorielles que devraient contenir ces documents sur le Web, car c'est désormais « un saut fonctionnel qui est mis en œuvre : on passe de la recherche d'images à la recherche d'objets dans les images »<sup>125</sup>. La problématique pour les chercheurs était donc de

<sup>123</sup> DUMONT, Stephan. « Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF) », Digiversity, 29 octobre 2015. En ligne : <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif/>

<sup>124</sup> BOUHAI, Nasreddine. SZONIECKY, Samuel. *Intelligence collective et archives numériques*. Londres : ISTE Editions. 2017, p.40.

<sup>125</sup> DELPIERRE, Nicolas. HIRAUX, François. MIRGUET, François. *Les chantiers du numérique : Dématérialisation des archives et métiers de l'archiviste*, Paris : L'Harmattan, 2012, p.135.

constituer pour leurs lettres numériques une base de données, sur leur plateforme, qui suivait une approche davantage coordonnée que sectorielle... Il nous faut ici éclairer le sens de ces mots. La coordination des opérations ayant traits à l'organisation des lettres, sur la plateforme de ces projets, ne signifiait pas un changement d'échelle par rapport au maniement fondamental de la lettre. Nous sous-entendons ici que les chercheurs doivent agencer désormais des grands ensembles de documents qu'ils ont déjà renseignés. C'est à l'aide de ces descriptions numériques que les regroupements à opérer pourront prendre une signification réelle pour les chercheurs. Pour eux, l'enjeu à ce stade du processus va être de rechercher et d'accéder à l'information contenue dans les lettres d'une manière à la fois pseudolocale mais aussi sémantique. Le volume des correspondances d'écrivain est une variable qu'il faut continuer de prendre en compte, par souci de lisibilité de la lettre, et même lorsque celles-ci peuvent être archivées dans leur ensemble sur le web... La présence d'une lettre d'écrivain dans un corpus s'éclairera tout d'abord à sa propre échelle, en tant qu'objet ayant pénétré les sphères du Web, mais également à celle du réseau de lettres du même écrivain sur la plateforme. Donc, «il est important d'ouvrir ces grands champs conjugués de la connaissance où les chercheurs pourront travailler à des reconstructions déparcellisées»<sup>126</sup>.

Le second écueil qui se pose aux chercheurs concernant l'accès à la lettre d'écrivain sur ces deux plateformes va être pour les chercheurs et informaticiens de ces projets de retarder au maximum l'arrivée du gap sémantique. Pour définir brièvement ce problème rencontrée par les bases de données numériques, il s'agit de « ce qui empêche que les machines puissent elles-mêmes décider de leur cheminement dans les ressources du web »<sup>127</sup>. Le système d'une base de données, quel qu'elle soit, effectue en effet des tâches de détection et d'agrégation de l'information en son sein. Il y a donc une nécessité pour les chercheurs d'objectiver chaque relation entre les lettres et leurs constituants dans la base de données et de rendre pour cela multimodale leur plateforme d'archivage de ces lettres. Nous allons voir les équipes Flaubert et Zola ont répondu et tentent aujourd'hui de répondre sciemment à ces défis systémiques.

Car malgré leurs approches différentes, ces deux projets se situent initialement au sein du même espace d'exploitation de leur travail, à savoir l'Hypertexte arborescent. Leur plateforme tendent toutes deux vers le même objectif, qui est d'adopter une « une forme plus libre, une structure en étoile dont l'œuvre inachevée serait le point névralgique associant le combinatoire, l'hypermédia et le Web Sémantique»<sup>128</sup>.

Mais derrière ces contraintes se dévoilent également des perspectives pour toute base de données épistolaire sur le Web, à exploiter. Le contenu d'une lettre, de Flaubert ou de Zola, est la plupart du temps restreint, ce qui facilitera son intégration à cette structure rhizomatique, donnant tout loisir aux chercheurs de valoriser les liens intrinsèques et extrinsèques à ces documents. Ceci à l'inverse des œuvres purement littéraires qui, même après avoir été rendues numérique par leur indexation, sont dotées d'une ossature rendant difficile toute plongée au sein du texte, bien souvent du fait de sa densité.

« L'indétermination – que les bases de données ont du mal à tolérer – indique une autre différence entre le récit et la base de données. Le récit fait signe vers l'inexplicable,

<sup>126</sup> Ibid. p.160.

<sup>127</sup> FAURÉ, Christian. « Le gap sémantique et l'architecture orientée message ». Hypomnemata : supports de mémoire, mis en ligne le 5 octobre 2013. Accessible en ligne : <http://www.christian-faure.net/2013/10/05/le-gap-semantique-et-larchitecture-orientee-message/>

<sup>128</sup> LINKES, Serge. « Éditer le manuscrit inachevé » *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008, p.198. En ligne : <https://journals.openedition.org/recherchestravaux/106>

l'indicible, l'ineffable, alors que la base de données repose sur l'énumération, exigeant la formulation explicite des attributs et des valeurs de données »<sup>129</sup>

Mais la correspondance d'un écrivain s'émancipe de la fiction et des stratifications implicites du récit, lui préférant un cadre d'expression normatif, comme nous l'avons souligné. En ce sens, la lettre de ces écrivains peut trouver dans les bases de données un allié de taille, car celles-ci tendront de facto vers l'inclusivité de leur contenu. Les fonds de correspondance, qu'il advient de mobiliser sans cesse pour mieux les pénétrer, se déploient au fond d'une façon équivalente à celle des pages Web. Tous deux ne s'identifient pas au sens du *codex* originel, a contrario des livres et du procédé d'écriture traditionnelle d'un auteur. Ce principe du palimpseste numérique fait prévaloir une nouvelle spatialisation de l'information archivée sur le Web, qui aura tout lieu de valoriser les correspondances d'écrivains, à condition de problématiser leur mobilité nouvelle.

« L'essor des bases de données numériques, même s'il élargit radicalement les possibilités d'exploitation et d'interprétation des sources, ne signifie pas l'avènement d'une méthode de travail unique et d'un outil exclusif ; tout outil informatique ne saurait se substituer à la réflexion et à l'histoire elle-même. »<sup>130</sup>

Nous voyons qu'avec ces deux projets d'éditions électroniques de correspondances d'écrivains, l'intention des équipes Flaubert et Zola va au-delà d'une simple transposition et mise en ligne des lettres... Elle tend à proposer une nouvelle interface d'exploration de la lettre, susceptible d'ailler au-delà la recherche d'occurrences. Nous dévoilerons dans notre dernière partie les fonctionnalités de visualisation accentuée des corpus numériques de lettres, susceptibles dévoiler de nouveaux enjeux historiographiques entourant Gustave Flaubert et Émile Zola... Mais cette accessibilité inédite de la lettre dans la base de données fait resurgir une dernière problématique, concernant la diffusion de l'information dans la lettre, cette fois-ci sur le plan juridique...

« Les bases de données sont protégées juridiquement par les articles L 112-1 à L 112-4 du Code de la propriété intellectuelle, mais le statut de l'utilisateur-historien, producteur de données, varie selon la période étudiée ; ce qui modifie en conséquence les dispositions légales concernant les données entrées. Le droit d'auteur protège le producteur uniquement si les données entrées sont originales. Or, le sont-elles vraiment lorsqu'elles sont issues de documents anciens et théoriquement libres de droits ? »<sup>131</sup>.

La question des Ayants-droits dans les lettres réapparaît donc ici en substance, d'où la nécessité pour les équipes scientifiques de passer une convention écrite avec la famille de l'écrivain. Une fois ces questions considérées, les chercheurs pourront de nouveau songer à l'agencement de leurs lettres sur la plateforme, en répondant à l'un des derniers enjeux de leur entreprise, à savoir l'évolutivité de leur corpus. Et nous verrons que là encore, les approches adoptées par nos deux équipes scientifiques diffèrent, ce qui donne leur confère toute leur richesse dans notre problématique de recherche.

<sup>129</sup> KATHERINE-HAYLES Nancy. *Lire et penser en milieux numériques : attention, récits, technogénèse*. Grenoble : Éditions littéraires et linguistiques de l'université de Grenoble, 2016, p.286.

<sup>130</sup> MOREAU, Yves. «Épistolarité et base de données : l'étude d'une correspondance savante à l'aide des TIC ». *L'historien face au manuscrit*. Presses universitaires de Louvain, 2012, pp.255-266. En ligne : <https://books.openedition.org/pucl/1278?lang=fr>

<sup>131</sup> Ibid.

### II.3.2. Délimitation du corpus de lettres numériques et mise en accessibilité sur la plateforme du Web

Nous avons vu que la création et la gestion d'une base de données consacrée à la correspondance d'un écrivain était facilitée grâce au profil fragmentaire de ces documents. Mais si son alimentation ne posera aucun soucis sur la plan volumétrique, une délimitation du corpus des lettres d'écrivain s'impose toutefois aux chercheurs qui souhaitent l'archiver et le rendre accessible sur le Web.

Pour le Centre Flaubert, cette problématique a consisté dans un premier temps à «concilier les exigences d'utilisateurs très divers (...) dans un «panier» pour les exploiter ensuite hors connexion, jusqu'aux chercheurs «pointus» qui voudront faire porter leur requête sur des secteurs bien délimités du corpus»<sup>132</sup>. Nous avons précédemment évoqué l'aspiration de la plateforme du centre Flaubert à archiver tous les documents, littéraires ou non, rattachées à l'écrivain normand. En ce sens, l'archivage numérique de la correspondance de Flaubert ne serait qu'un jalon posé au sein d'un espace encore plus vaste, dédié à la figure centrale de l'écrivain et où « se dessine un ToutFlaubert ». Derrière cet idéal, la première contrainte qui apparaît alors est temporelle, car « si le corpus [de lettres] est extensible , le temps imparti à chacun, concepteur, réalisateur ou utilisateur, ne l'est pas»<sup>133</sup>.

Comment les chercheurs de ces deux équipes pourraient t-elles alors mettre en place des mécanismes de régulation documentaire au sein de leur plateforme ? L'équipe scientifique du Centre Flaubert a d'ores et déjà rendu accessible sur le Web toutes les lettres de la correspondance active de Gustave Flaubert à sa disposition. Cela en ne se posant aucune contrainte pour le nombre de lettres à inclure, dans un souci d'exhaustivité. Les chercheurs s'appuyaient sur les 4700 lettres déjà éditées en version papier, ajoutant à ce corpus les lettres inédites, pour arriver à un volume d'environ 5000 lettres. Les nouvelles lettres de l'écrivain retrouvées viendront progressivement se greffer à ce fonds, ce qui signifie que le corpus ne connaîtra jamais de limitation, tout comme celui son pendant passif. Les chercheurs travaillent actuellement sur environ 3000 lettres reçues par Flaubert, et continuent à recevoir quotidiennement des nouvelles lettres de la part de collectionneurs privés qui viendront compléter leur correspondance générale de l'écrivain d'ici à 2021. Ce projet de mise en visibilité des lettres s'articule donc progressivement ; l'important étant pour les chercheurs de partir des lettres écrites par Flaubert avant de dévoiler progressivement ensuite leurs répercussions...

Concernant l'équipe Zola, le volume encore plus conséquent des lettres à traiter au sein de cette correspondance passive les a obligé à circonscrire différemment leur corpus de lettres. Un premier corpus a été délimité, en rapport avec la période de l'affaire Dreyfus, qui va des années 1897 à 1899 et inclut des lettres envoyées à l'écrivain depuis l'étranger. Ce fonds est composé de 2000 lettres, équivalent donc à environ 10% du corpus général estimé. Nous voyons que cette équipe a régulé dans le temps l'arrivée de nouvelles lettres sur leur plateforme, à travers des délimitations événementielles, thématiques mais aussi chronologiques. Par exemple, la délimitation d'un corpus dédié aux lettres envoyées à Zola durant sa période d'écriture de la saga des «Rougon-Macquart » est en cours de réflexion, chez les chercheurs, pour étudier la réception effective du naturalisme dans le monde.

<sup>132</sup> LECLERC, Yvan. « L'édition intégrale en ligne des manuscrits de *Madame Bovary* ». *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008, p.235.

<sup>133</sup> Ibid. p.239-240.

Mais l'évolutivité des corpus de lettres que nous présentons est à considérer non seulement selon le nombre de lettres mais également selon le nombre de termes d'une lettre à archiver. Les interrogations pour les chercheurs redeviennent alors génétiques. Prenons l'exemple de Flaubert qui reprend constamment les mêmes phrases et tournures de phrases dans ses lettres. Chaque mot commun pourra ainsi avoir des centaines, voire des milliers d'occurrences, après une recherche effectuée dans la base de données des ces plateformes. C'est ainsi qu'une base de recherche multi-critères a été mise en place par l'informaticien du Centre Flaubert, afin de faire porter sa requête sur des points bien précis du corpus. Comme nous l'avons vu, les entrées sont multiples au sein des correspondances et sont surtout vouées à augmenter. Le nom du correspondant, le lieu de rédaction et le lieu de réception de la lettre sont les trois principaux critères de recherche de la base de données des lettres de Flaubert. Cela n'empêche pas que 272 correspondants sont actuellement recensés, chacun faisant l'objet d'une notice dans l'index des noms propres de la plateforme. De même pour les 69 lieux de rédaction où furent rédigées les lettres écrites par Flaubert ; chiffre qui pourrait tripler au regard de la correspondance passive de l'écrivain... La barre de recherche temporelle couvre les années 1828 à 1880. C'est pourquoi les répertoires et index de la correspondance numérique doivent être sans cesse ouverts pour les chercheurs, afin de faciliter toute actualisation de corpus.

« Un corpus de correspondance ne doit jamais être considéré comme définitif et complet. Le numérique permet un enrichissement du corpus, une actualisation des données et des outils, tout en apportant des solutions pour l'organisation et la constitution du corpus de lettres. »<sup>134</sup>.

Prenons à présent l'exemple de la correspondance de Zola. Les derniers volumes de la correspondance, publiés chez Gallimard en 2014 avec les « Lettres à Jeanne » et les « Lettres à Alexandrine », a dérouté les spécialistes de l'écrivain d'un point de vue stylistique. Le changement de tonalité qu'elles instaurent, dans un registre intime où Zola ne s'aventura pas souvent par rapport à l'ensemble de ses lettres de l'écrivain, était alors une nouveauté indéniable pour les études menées à son sujet. Leur présence sur la plateforme du projet CorrELEZ présentera alors un gisement fascinant de la recherche pour les passionnés de cette œuvre qui n'avaient pas pu se procurer les volumes de publication de ces lettres. La nécessité de valoriser la correspondance passive grâce au Web émerge alors, d'autant plus que le numérique permet de traiter de forts volumes évolutifs de documents, et concourir à leur réalisation, à leur diffusion.. Le souhait des archivistes du XX<sup>ème</sup> siècle se trouve dans l'usage des outils numériques par les équipes scientifiques porteurs de ces projets. «Les outils numériques permettent d'explorer le libre accès, le travail en commun, le remix et la recherche itérative ; ils ont le potentiel pour offrir une vision alternative à la communication savante».<sup>135</sup>

Mais cela requiert dès l'origine du projet, comme nous l'avons montré, une souplesse dans le choix des logiciels de base de données et également au niveau des métadonnées des lettres, qui doivent pouvoir être exportées librement...C'est tout « le principe (...) de l'accès comme usage et de l'usage comme accès »<sup>136</sup> des données dans le domaine des SHS. Dans cette optique, les

<sup>134</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.5. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>135</sup> ADEMA, Janneke. « Mettre en pratique ce que l'on prêche. La recherche en sciences humaines et sa praxis critique ». *Read/Write Book 2*. Marseille : Open Editions Press, 2012, p.102. En ligne : <https://books-openedition-org.docelec.enssib.fr/oep/226>

<sup>136</sup> SAJUS, Bertrand, Web 2.0 et mémoire : de la conversation à la conservation. *Documentaliste-Sciences de l'Information* [en ligne], 2009, vol. 46, n.1, p. 55

équipes scientifiques pourront se démarquer seulement grâce à l'architecture de leur plateforme, devant suivre les normes numériques pour toute infrastructure présente sur le Web.

Les correspondances de Zola seront disponibles sur un site développé en WordPress développé par un prestataire extérieur, avec un thème dérivé de la plateforme Archiz auquel il sera rattaché. Cette filiation ne nous intéresse qu'en dernier lieu, concernant notre objet d'étude. C'est pourquoi nous nous concentrerons à présent sur des notions clés de l'évolutivité des corpus épistolaires numériques, à savoir leur moissonabilité et à leur exportabilité<sup>137</sup>. C'est le plan de travail numérique des chercheurs du projet CorrELEZ qui va nous intéresser ici, intitulé Oméka, qui ratifie le moissonnage des données numériques des lettres selon le protocole OAI-PMH. Rappelons que cet outil se trouve sur la plateforme Eman<sup>138</sup> d'édition de manuscrits et d'archives numériques, où seules seront exposées et récupérées les métadonnées encodées en Dublin Core. Les champs personnalisés, que nous avons précédemment évoqués, n'aideront les chercheurs qu'à structurer l'information des lettres avec plus de granularité... La création de corpus de lettres signifiants, déterminants, ne se fera qu'au prix de la postulation, par les chercheurs, de l'archivage numérique de la correspondance d'écrivain par son exploitation scientifique. Faut-il encore inclure des paramètres d'exportabilité des données de la lettre, en les rendant "moissonnables", c'est à dire accessible sur un serveur web que l'on nommera « entrepôt », et interopérables. C'est pourquoi chaque description de lettre doit inclure une URL pointant vers le document décrit, dans sa version numérique. Cette opération a pu être effectuée par les chercheurs de CorrELEZ grâce à Oméka. Sur cette interface<sup>139</sup>, un tableau de bord indique le nombre de contenus, de collections, de mots-clés et d'extensions en cours d'enrichissement et de transaction.

La création d'un entrepôt OAI-PMH nécessite plusieurs composants. Il est donc primordial que les métadonnées affiliées aux ressources de la lettre soient interopérables, avant de transiter vers leur base de données. Dans le cadre du projet CorrELEZ, il s'agit en l'occurrence de la base de données MySQL Oméka. Un serveur web, accessible par Internet est également requis, de même qu'une interface de programmation, appelée API (Perl, PHP, Java-Servlet...), un identifiant pour l'entrepôt et l'URL de base. Sans oublier une application qui soit capable de répondre aux 6 requêtes OAI (Identify, ListSets, ListIdentifiers, ListMetadataFormats, ListRecords, GetRecord) et une gestion du flux permettant de renvoyer les notices par paquets, pour les bases de données ayant à traiter des milliers de documents numériques.

### **II.3.3. Maintien de l'archivage numérique d'une correspondance d'écrivain : à quelles conditions ?**

Dans ce contexte de création de ces nouveaux objets éditoriaux que sont les correspondances numériques d'écrivains, nous étudierons les lettres de Gustave Flaubert et d'Émile Zola à la lumière d'un nouveau concept : la « philologie numérique ». Cette façon d'appréhender le texte épistolaire sur le Web permettra d'en créer de nouveaux modes d'accès,

<sup>137</sup> DUPOND, Marie. « De l'édition à l'éditorialisation. Pour une approche historique des éditions de correspondances ». Université Lille 3. 22 novembre 2016. En ligne : <https://live3.univ-lille3.fr/video-etudes/dhnord-2016-de-ledition-a-leditorialisation-pour-une-approche-historique-des-editions-de-correspondances.html>

<sup>138</sup> <http://eman-archives.org/EMAN/>

<sup>139</sup> Voir Annexe 2.

qui pérenniseront les correspondances d'écrivains. Cette opération se produira à partir de la diversité des formes de balisage, qui « facilite de manière inédite la lecture des textes et, mutation essentielle, révolutionne leur appréhension par la constitution de ces textes en «corpus»raisonnés »<sup>140</sup>.

La mutualisation des couches documentaire autour des lettres s'impose pour ces plateformes d'archivage sur le Web que nous étudions. Nous avons tracé les divers degrés de l'enrichissement numérique d'une lettre, avec pour commencer un sujet et ses champs descripteurs induits dans les spécifications Dublin Core. Puis, sa transcription approfondi sous forme d'un document textuel enrichi, et enfin l'image numérisée de la lettre sous forme d'un document image éventuellement enrichi. C'est la présence simultanée de toutes ces versions numériques de la lettre qui justifiera une accessibilité optimale de celle-ci sur le Web... Cette architecture archivistique nouvelle pourraient s'apparenter à une « une instrumentation idoine [qui] doit permettre d'articuler et de mettre en réseau documents et vues en une hiérarchie d'autorité et de qualité : maintenant les versions exemplaires pour en faciliter la consultation systématique, le retour pour authentifier ou certifier les vues construites pour les appropriations diverses. »<sup>141</sup>. Il advient dès lors pour les chercheurs de concilier les «problématiques des réseaux de corpus, qui permet à l'auteur ou au lecteur de l'un de ces mondes de connaissances d'interroger dynamiquement les autres»<sup>142</sup> ;

Une problématique que nous retrouvons dans les modalités-mêmes du protocole OAI-PMH : « le protocole OAI-PMH n'interroge pas directement la base de données de l'organisme, mais un entrepôt obtenu par reformatage des données de la base. L'entrepôt OAI est donc une extension de la base d'origine, une image à un moment donnée de la base (...) On parle alors de recherche d'information asynchrone. »<sup>143</sup>. Nous voyons que d'après ce protocole, en cours d'étude du côté de l'équipe Zola, les correspondances d'écrivain feraient alors partie intégrante des infrastructures Web en étant soumises à leur propre système itératif.

Nous avons dans la première partie de notre travail évoqué les spécificités de l'épistolaire au XIX<sup>ème</sup> siècle, parmi lesquelles l'épuisement du détail enfoui dans les lettres ou les nombreuses ratures manuscrites empêchant parfois l'interprétation d'une phrase précise... Cette mise en accessibilités des contenus de la lettre sur la plateforme doit permettre de pallier à ces limites inhérente au genre épistolaire. Cela doit à la fois « permettre d'articuler l'objectivation documentaire et l'appropriation lectoriale, en opposant une référence permanente à des vues dynamiques »<sup>144</sup>

La linéarité de la lecture, qui primait au sein des éditions physiques avec les renvois aux index en fin d'ouvrage, commencera alors à être supplantée par les spécificités du Web de données, à

<sup>140</sup> LERICHE, Françoise. « Quel balisage pour les corpus épistolaires numériques ? De l'annotation traditionnelle du « document » à une analyse générique et pragmatique », dans Rastier, F. et Ballabriga, M., *Corpus en lettres et sciences sociales*. Paris, Texto, 2006; En ligne : <http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Sommaire.html>

<sup>141</sup> BACHIMONT, Bruno. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Hermes Science Publications, 2007.

<sup>142</sup> LOCHARD, Éric-Olivier. « L'édition de manuscrits au prisme du paradigme instrumental Arcane » dans LERICHE, F. MEYNARD C. *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008, p.95. En ligne : <https://journals.openedition.org/recherchestravaux/92>

<sup>143</sup> MAHE, Annaïg. Prime-CLAVERIE, Camille. « Le défi de l'interopérabilité entre plates-formes pour la construction de savoirs augmentés en sciences humaines et sociales ». ISTE éditions, 2017, p.3. En ligne : [https://archivesic.ccsd.cnrs.fr/sic\\_01511618/document](https://archivesic.ccsd.cnrs.fr/sic_01511618/document)

<sup>144</sup> BACHIMONT, Bruno. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Hermes Science Publications, 2007, p.186.

commencer par les liens hypertextuels. L'utilisateur au sein de la base de données pourra « construire son parcours à travers les données devenues hypertextuelles » et ainsi obéir plutôt à « une lecture exploratoire qu'à une lecture en continu. »<sup>145</sup> de la correspondance d'un écrivain. L'enjeu pour les équipes scientifiques consiste donc à transformer les anciennes notes de bas page, en hyperlien vers la lettre suggérée, grâce à une table de correspondances qui se voudra interactive. Ce travail représente plusieurs milliers de notes à modifier dans leur format, sachant qu'elles doivent bien sûr être interopérables pour intégrer cette nouvelle base de données.... Les notices documentaires, doivent être considérées individuellement, en tant qu'entrées multiples au sein de la correspondance, mais elles ne suffiront plus dès lors à elles-mêmes. Elles doivent se fondre au sein de ce système d'information... « Avec le corpus numérique de correspondances, on dépasserait même le feuilletage hypertextuel pour être dans le domaine de la navigation par interrogations »<sup>146</sup>. (p.18).

Pour ces deux projets, l'idée était d'obtenir sur l'écran plusieurs fenêtres ouvertes en même temps et de pouvoir choisir celles que l'on veut ouvrir ou fermer, ou déplacer horizontalement : fac-simile, transcription, notes d'origine, notes secondes (nouvelles notes ajoutées à l'édition papier, afin d'apporter une précision nouvelle, une correction, etc...). Ces fenêtres peuvent être considérés, selon la terminologie de Michel Foucault, comme des cadres énonciatifs, qui permettront au chercheur de remettre à tout moment la lettre étudiée dans son contexte opératif. Nous allons à présent nous intéresser aux perspectives d'ajustement et d'extension de ces corpus épistolaires numériques qui s'offriront à ces chercheurs, si ces derniers souhaitent aller plus loin dans leur entreprise d'archivage numérique.

<sup>145</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.18. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>146</sup> Ibid.

### III. PERSPECTIVES DE L'ARCHIVAGE NUMÉRIQUE DES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE SUR LES PLATEFORMES DU WEB

Nous approchons à ce stade de notre recherche d'un enseignement essentiel concernant l'établissement d'une archive numérique épistolaire. Une quête de nouveaux horizons concernant les données des lettres d'écrivains a pu émerger, grâce de leur indexation numérique et leur intégration à des base de données épistolaires, sur le Web. Nous avons retracé le cheminement des équipes scientifiques porteuses de ces projets, sur plusieurs années, qui tendait à amener les lettres de Gustave Flaubert et Émile Zola vers une extensibilité d'elles-mêmes absolument novatrice. Cela marquera l'une des perspectives de l'archivage numérique des correspondances d'écrivain, avec également l'ouverture des données des lettres vers d'autres collections numériques sur le Web. Les lettres ne seront plus simplement numérisées, au regard de ces perspectives, mais seront rendues proprement numériques. Les questions de leur usage et de leur médiation, à partir de leur plateformes d'archivage découleront, de ces points entrevus.

« La distinction entre patrimoine numérique et patrimoine numérisé ne sont pas à confondre. Le patrimoine numérique est tout à la fois un objet numérique et l'outil qui sert à le faire fonctionner [...] Le patrimoine numérisé renvoie à l'existence d'un objet original existant dans une matérialité différente et dont la numérisation permet d'en sauvegarder une image ». Pourtant ils mettent en jeu tous deux la question de l'usage de la mémoire sociale dans la qualification du patrimoine et le statut même de ce qui fait ce patrimoine »<sup>147</sup>.

#### III.1. Extension des corpus épistolaires sur le Web aux Humanités Numériques

##### III.1.1. Le projet Mapping the Republic of Letters et les fonctionnalités de représentations visuelle de corpus

Nous allons tout d'abord nous intéresser à l'évolution intrinsèque d'une plateforme d'archivage numérique des correspondances sur, au regard des outils développés en Humanités Numérique pour ce type de corpus. Des fonctionnalités pourraient en effet faciliter l'appréhension de la correspondance d'un écrivain sur le Web.

L'idée est de constituer un véritable espace numérique, où la correspondance pourrait s'étudier quasi exhaustivement, avec des outils de recherche et d'analyse, communs, et à la pointe. Pour dévoiler des perspectives en la matière, l'exemple du projet intitulé « Mapping the Republic of Letters » (MLR) nous semble être d'une grande importance. Ce projet a été développé par des chercheurs de l'université de Stanford, en collaboration avec l'Electronic Enlightenment Project. Il s'agit de la plus vaste collection en ligne de correspondances éditées du début de la période moderne. Cette archive épistolaire numérique réunit les lettres de personnes de chaque continent du monde, à partir du début du XVII<sup>e</sup> siècle jusqu'au milieu du XIX<sup>e</sup> siècle. Elle fournit ainsi une matière première abondante pour reconstituer l'une des plus grandes

<sup>147</sup> DOBEDEI, Vera. TARDY, Cécile. *Mémoire et nouveaux patrimoines*, Marseille : OpenEdition Press. 2015.

«conversations» historiques du monde, à travers les siècle. L'université de Stanford avait donc à sa disposition un immense fonds de lettres numérisées, écrites et reçues par les figures de Lumières en Europe. Les chercheurs n'avaient donc pas à indexer les données numériques de ces lettres et se sont concentré sur la création de leur base de données épistolaire. Chaque lettre du corpus a été répertorié, avec ses caractéristiques propres : noms de l'expéditeur et du destinataire, lieux d'expédition et de destination, date, sujets... Ils ont pu dès lors développer des outils d'analyse et d'interprétation de leur corpus textuel, issus des Humanités Numériques. Ces derniers devaient présenter toutes sortes de représentations quantitatives des données de leurs lettres, d'une façon ludique (camemberts, courbes et histogrammes présentant divers niveaux fréquences et de répartition numérique, selon différents critères). La phase la plus élaborée de leurs représentations visuelles de corpus devait finalement prendre la forme de cartes heuristiques et de cartographiques interactives...

« On voit donc que cette visualisation produit une sorte de réification de la notion de réseau de correspondance et suivant un modèle complètement anachronique. Tout se passe comme si les lettres voyageaient en ligne droite d'un point à l'autre sans considération des détours, de toutes sortes d'événements et de difficultés qui produisent une expérience particulière de la réalité du continuum spatio-temporel à l'intérieur du cadre historique donné. »<sup>148</sup>.

Avec ce projet, ce n'est pas tant l'affichage du contenu des correspondances qui prévaut, mais plutôt les mécanismes d'accessibilité aux lettres numériques, qui préparent déjà pour les chercheurs un terrain d'interprétation de la donnée épistolaire. Ces fonctionnalités ont été envisageables grâce à des logiciels d'exploration de données tels que *Palladio*, permettant de produire de riches visualisations axées notamment sur l'aspect social et géographique du corpus, à partir des métadonnées des lettre structurées. Cette opération s'est déroulée par l'intégration d'un système d'information géographique et de métadonnées – c'est-à-dire des données descriptives et informationnelles, distinctes du contenu textuel des lettres – tirées de la collection « Electronic Enlightenment »<sup>149</sup>.

La plateforme MLR offre par ailleurs un outil de visualisation qui examine la circulation spatio-temporelle des lettres à travers la « République des Lettres » européenne. Outre les textes sources des lettres, la ressource fournit des informations chronologiques, biographiques et géographiques pour plus de 50 000 lettres. Cela a permis aux chercheurs d'interroger le corpus de lettres et d'en explorer tous ses recoins. Ils ont ainsi pu définir si la République des lettres avait ou non des frontières. La frise chronologique reste toutefois un outil primordial d'analyse d'un fonds de correspondance, et seul l'analyse humaine est susceptible, grâce à elle, de déceler les variations du nombre de lettres par années. De surcroît, l'utilisation des bibliothèques graphiques (dites d'« interfaces utilisateurs ») à source ouverte, a permis aux chercheurs de visualiser les données récoltées avec un degré de granularité optimal.

Nous voyons, à partir de cet exemple, que la première étape pour les chercheurs souhaitant à étendre leur corpus épistolaire, en créant de nouvelles fonctionnalités numériques, consistera à comprendre la « forme » de l'archive. Avec ces outils avancés des Humanités Numériques, la notion de « réseau épistolaire », comme on pouvait l'entendre au XVIII<sup>ème</sup> siècle, devenait plus lisible à travers ce projet. Sur cette plateforme, il est par exemple possible de suivre une controverse ou la naissance et la diffusion d'un concept. Les unités de représentation graphique peuvent alors apparaître en bleu ou en rouge, pour désigner visuellement les résultats les plus proches ou les plus éloignés de la requête initiale par mots-clés. Cela peut s'avérer pertinent si l'on constate par exemple une densification de points rouges au sein d'un même lieu géographique, susceptibles de révéler un

<sup>148</sup> MOUNIER, Pierre. *Les humanités numériques*. Paris : Maison des Sciences de l'Homme, 2018, p.74.  
En ligne : <https://books.openedition.org/editionsmsmh/12039?lang=fr>

<sup>149</sup> En ligne : <https://www.e-enlightenment.com/>

pic d'activité épistolaire localisé au cours d'une période donnée. Quant à l'isolement de certains points bleus sur la carte, ils indiquent aux chercheurs qu'ils ont atteints dans leur recherche une granularité maximale. Ces considérations nous renvoient à un concept réputé dans toute théorie des graphes, qui est celle d'« intermédiarité ». Celui-ci permet aux chercheurs d'interpréter les nœuds au sein de leur représentation graphique du corpus de lettres. Si un nœud possède une grande intermédiarité, cela signifie qu'il a une grande influence sur les transferts de données dans le réseau épistolaire, sous l'hypothèse que ces transferts se font uniquement par les chemins les plus courts. Nous voyons donc ici une façon de fluidifier notre perception des données d'un grand volume de lettres. Il est ensuite possible de valider notre première impression en passant le curseur sur ces indicateurs colorés. La constitution d'un point nous apparaît alors à travers le nombre de lettres reçues ou envoyées, et surtout, le nombre de relations entre les points s'affiche, confirmant d'un point de vue factuel ces liens visuels. D'où l'importance de bien choisir ses descripteurs du côté des chercheurs, au moment de l'indexation numérique de leurs lettres, ce qui pouvait leur permettre alors de bien cerner les grandes idées que les savants véhiculaient. Suite à cela, les représentations visuelles de corpus pouvaient indiquer le plus précisément possible leurs interactions dans un champ donné, à l'aide des outils d'Humanités Numériques. « Ainsi, on met en lumière d'autres facettes des données, on diversifie la publication des résultats scientifiques, et on peut s'associer à d'autres corpus afin d'ouvrir d'autres pistes de recherche. »<sup>150</sup>. Nous voyons en effet que ces liens visuels vont au-delà de l'hypertextualité des plateformes sur le Web, tendant vers une mise en virtualité interactive du contenu épistolaire. Nous allons montrer comment les correspondances de Gustave Flaubert et d'Emile Zola pourraient par exemple enclencher ce processus de virtualisation de leur corpus, à l'avenir, en se concentrant dès l'indexation des données sur les éléments saillants de leur matériau de recherche. « Quand on consulte un corpus numérique de correspondance, on s'attend à pouvoir avoir accès à des visualisations spécifiques : géolocalisation des relations, réseaux sociaux, etc. Les choix d'encodage et d'édition sont fortement reliés à la question de la visualisation »<sup>151</sup>.

### III.1.2. Gestion de l'incertitude des données épistolaires numériques et perspectives humanistiques

Concernant les correspondances d'écrivains du XIX<sup>ème</sup> siècle que nous étudions au sein de notre travail, les Humanités Numériques pourraient en effet venir modeler cet objet très spécifique, à partir des caractéristiques les plus impactantes qu'auront préalablement révélé les chercheurs. Nous avons vu à travers la plateforme MLR que la République des Lettres au XVIII<sup>ème</sup> siècle était à étudier avant tout dans un cadre spatio-temporel restreint, pour obtenir des résultats significatifs. C'est pourquoi cette dimension est essentielle à prendre en compte dès l'indexation numérique d'un corpus de lettres. Toutefois, ces données peuvent être sujettes à interprétations diverses. Cela ne signifie pas leur incohérence, mais plus souvent leur incomplétude. C'est pourquoi les chercheurs doivent toujours considérer un corpus épistolaire comme un grand ensemble de données à modéliser, selon les variables d'ajustements qu'ils considéreront comme les plus probantes. Il apparaît alors comme une perspective de faire apparaître les lacunes textuelles d'une lettre ou

<sup>150</sup> MASSOT, Marie-Laure. « Humanités numérique : Transformation des savoirs et des métiers ». La vie de la recherche scientifique, Sncs Fsu, 2014, p.2. En ligne : <https://hal.archives-ouvertes.fr/hal-01794171/document>

<sup>151</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.8. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

encore l'ambiguïté des enregistrements historiques à leur sujet. À l'échelle correspondance qui serait représentée visuellement in fine, ces données spécifiques pourraient alors apparaître sous la forme de traits en pointillés ou d'un code couleur défini. Ces partis pris ont lieu, de la part des chercheurs, au moment de l'indexation des données numériques liées aux lettres, et marqueraient une trajectoire sans cesse évolutive de l'archive épistolaire sur le Web.

En suivant, citons l'exemple de « la date [qui] doit souvent être indiquée comme une hypothèse. C'est une habitude à prendre. L'encodage des dates est celui qui est le plus segmenté ; les principes doivent être fixes et connus pour pouvoir travailler le corpus. Enfin, il ne faut pas oublier qu'éditer son corpus de correspondances c'est aussi apprendre la gestion de l'incertitude »<sup>152</sup>.

Pour cela, il existe l'élément <date> dans la balise TEI <correspDesc>, que des projets tels que CorrELEZ seront amenés à modéliser. Cet élément fait partie des plus complexes à encoder, en raison des suites de chiffres qu'il faut inscrire en respectant un ordre précis, que les chercheurs devront maîtriser. L'incertitude que nous évoquons ici portera avant tout sur l'exactitude des dates liées aux lettres ; point sur lequel il est complexe de revenir une fois établie l'architecture du code d'établie. D'où la variété des attributs qui sont liés à cet élément, à savoir : @when, @notBefore et @notAfter, sans oublier @from et @to. Ces attributs permettent de circonscrire une ligne temporelle le plus précisément possible tout en distinguant les actes d'envois et de réception de la lettre en question... En conséquence, les zones d'incertitudes sont fortement réduites lors de l'encodage des données temporelles d'une lettre mais continuent à exister. Les écrivains ou leurs destinataires peuvent en effet avoir oublié de faire figurer la date de la rédaction de leur lettre. De plus, si l'écrivain a rédigé plusieurs lettres destinées à plusieurs destinataires au cours de la même journée, les noms de destinataire peuvent être facilement confondus au moment de leur encodage numérique par les chercheurs... C'est pourquoi il existe également, liés à l'élément <date> de la TEI, l'attribut @calender ainsi que l'attribut @cert, qui vont permettre d'objectiver ce traitement de l'incertitude, qui sera pourra être retracé par les chercheurs sur le Web. L'attribut @calender permettra de valider la date d'écriture ou de réception de la correspondance à l'aide d'un véritable calendrier numérique, afin d'éviter les fautes de frappes liées chiffres ou les espacements mal encodés entre les jours, les mois et les années. Le second attribut, @cert, va permettre quant à lui aux chercheurs de faire part des éventuelles méconnaissances de l'équipe scientifique au sujet de la moindre information qu'ils avanceront au sein d'une de leur balise...

Les stratégies de représentation de données absentes ou incertaines n'ont pas encore été explorées de manière approfondie dans l'analyse visuelle, même au sein de projets tels que MLR, qui font pourtant foi en la matière. Les collaborateurs de ce projet songent actuellement à cette nouvelle fonctionnalité, qui pourrait naître et se déployer au sein de leur plateforme d'archivage numérique des correspondance, selon le modèle de l'«ampliation». Ce terme désigne l'extension des données induite par leur interaction visuelle seule, avec les autres données du corpus, sur la plateforme. Développer et mettre en place un environnement Web «ampliable» serait un pas considérable en avant pour les chercheurs qui mènent ces projets d'archivage numérique des correspondances, dont la tâche première est de combler les lacunes dans les connaissances épistolaires, grâce à l'interprétation de leur contenu. Ces zones « grises » de l'incertitude d'un corpus de lettres deviendraient de facto des gisements de la recherche, exploitables directement par les chercheurs, ce qui pourrait par exemple amener ces projets à devenir inter-disciplinaires, dans la perspective des Humanités Numériques. Ces zones peuvent porter sur des dates comme nous l'avons montré, mais également sur des noms de destinataire ou d'expéditeur, sur des lieux de conservation

<sup>152</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.12. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

des lettres... Le fait de laisser une porte d'ouverte, à travers le balisage numérique d'une correspondance, peut inciter les chercheurs à ne pas reléguer même la plus petite information d'une correspondance, mais à la déléguer, à court ou moyen terme. En passant ensuite un appel à chercheurs, ces points d'achoppements multiples, dans la constitution numérique d'un corpus épistolaire, gagneraient à être soumis, par un accès ouvert mais retraçable encore une fois, à l'œil d'un historien, d'un chercheur en généalogie... L'intégration de ces nouvelles métadonnées serait déjà le signe d'une effervescence autour de l'archive numérique de la correspondance d'écrivain en constitution. Il ne faut pas oublier que ces plateformes sont consacrées à la production de nouvelles connaissances en sciences humaines et à leur archivage numérique. C'est pourquoi un champ d'exploration si vaste que peut l'être celui d'une correspondance d'écrivain doit s'accorder ces marges de progression collaborative, pour certifier le cheminement le plus scientifique qui soit au sein de leur plateforme.

Partant de cette quête d'authenticité des données numériques liées à l'épistolaire, la dimension temporelle nous entraînera de fait vers la dimension spatiale des lettres, avec l'identification et la certification des lieux mentionnés dans les lettres d'écrivains. Nous allons rester dans le cadre de l'encodage en TEI des lettres mais en découvrant cette fois un nouveau format, s'intitulant CMIF, avec le service <correspSearch> qui lui est lié. Ce format nous intéresse car il « permet de créer des fichiers XML rassemblant des éléments <CorrespDesc> » dans lesquels « chaque élément (...) est utilisé de manière plus restrictive que les recommandations TEI »<sup>153</sup>. De telles restrictions nous intéressent tout particulièrement dans notre étude concernant les marges d'erreurs dans l'encodage spatio-temporel d'une correspondance, car elles appellent la plupart du temps à l'emploi de fichiers d'autorité. Ces derniers seront la seule façon d'identifier de façon unique chaque champ, chaque entité de la correspondance d'un écrivain, marquant un peu plus sa justification historique et son établissement en tant qu'archive numérique.

« Les fichiers d'autorité sont des bases de données généralement créées et toujours gérées par des consortiums coopératifs de bibliothèques, souvent dirigées par une bibliothèque majeure ou nationale (...) conçus pour «établir des formes de noms» et d'autres métadonnées «utilisées dans les notices bibliographiques... afin de fournir un accès uniforme aux documents des catalogues de bibliothèque et d'identifier clairement les auteurs et les vedettes-matières. »<sup>154</sup>.

Certaines de ces données obligatoires codées dans l'en-tête TEI du fichier CMIF feront l'objet d'un contrôle d'autorité systématique. Il s'agira du nom de l'expéditeur, du nom du destinataire ainsi que du lieu d'écriture et de réception de la lettre, qui seront affiliés à une ID unique. Mais si l'on se place à l'échelle spatiale, l'incertitude resurgit alors très rapidement pour le chercheur. C'est pourquoi une base de données à licence gratuite et ouverte, telle que peut l'être GeoNames, nous apparaît comme une source pertinente pour vérifier chaque nom de lieu récoltés par les équipes scientifiques dans les lettres d'écrivains. Certains lieux pouvant être moins réputés que d'autres, le chercheur pourra tout de même encoder à son gré des adresses géographiques très spécifiques dans les balises. Les écrivains du XIX<sup>ème</sup> siècle étaient mobiles, comme nous l'avons évoqué avec Émile Zola, et rédigeaient parfois leurs lettres dans des coins reculés du monde... Ces identifications pourraient ensuite concerner les personnes mentionnées dans les lettres, comme nous le verrons dans notre prochain chapitre avec le site Web de la Bibliothèque Nationale de France qui met à

<sup>153</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.25. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>154</sup> DUMONT, Stephan. « correspSearch – Connecting Scholarly Editions of Letters », *Journal of the Text Encoding Initiative*, Décembre 2016-Juillet 2019, p.7. En ligne : <https://journals.openedition.org/jtei/1742#tocto1n2>

disposition les fiches de références de l'ensemble de ses catalogues numérisés ou non. Il existe aussi le VIAF à l'échelle internationale, utile pour retrouver non seulement le nom d'un écrivain français mais également celui de ses relations et donc de potentiels destinataires de lettres. Les ID qui seront ainsi nouvellement accolées aux personnes et aux lieux décrits dans les lettres favorisera la traçabilité des ces données sur le Web, et faciliteront leur interopérabilité. Nous voyons donc le que traitement de l'incertitude dans les lettres doit être systématisé avant d'intégrer de nouvelles sphères éditoriales et patrimoniales sur le Web.

Le service <CorrespSearch> permet en définitive de « fournir des métadonnées permettant de partager des corpus de lettres quelque soit leur format »<sup>155</sup>. Cette interface (API) est destinée à être encapsulée dans une page Web et permettre de retrouver un ensemble de métadonnées identifiants des projets de correspondance. Il ne s'agit pas seulement d'un moteur de recherche. Trois ans après son lancement, 23 000 enregistrements de lettres de plus de 90 éditions et publications avaient été intégrés à la base de données... Cela nous semble être une base tangible pour ensuite créer un gisement de fichiers XML basés sur les éléments fondamentaux de la balise <correspDesc>, que chaque lettre éditée devra contenir a minima. L'ouverture des bases de données épistolaires que les équipes scientifiques étudiées institueront sur le Web sera ainsi optimisée, grâce au format libre des données numériques qui facilite leur agrégation et leur partage entre infrastructures numériques. « Par conséquent, avec des données suffisamment étendues et un logiciel approprié, il sera possible d'effectuer des recherches, par exemple, sur des réseaux sociaux ou des réseaux de correspondance basés sur la correspSearch. »<sup>156</sup>

### III.1.3. Perspectives des Humanités Numériques pour les correspondances de Gustave Flaubert et d'Émile Zola sur le Web

Tout cela nous amène à étudier les perspectives concrètes que les projets portés par l'équipe Flaubert et par l'équipe Zola pourraient étudier et mettre en œuvre à partir de leur plateforme, prochainement. Nous devons pour cela délaissier la dimension kaléidoscopique dans la visualisation du corpus épistolaire entrevu avec MLR et la représentation des flux d'échange épistolaire au XVIII<sup>ème</sup> siècle. Pourquoi les caractéristiques propres aux correspondances d'écrivains du XIX<sup>ème</sup> siècle entrevues dans notre première partie ne pourraient-elles pas s'appliquer à une perspective spécifiquement issue des Humanités Numériques à leur tour ? Nous allons montrer que l'affect des épistoliers, tels que Flaubert et Zola, pourrait ainsi être matérialisé graphiquement, à l'aide des outils de visualisation que nous avons évoqué dans le chapitre précédent. Jean-Sébastien Macke, responsable du projet CorrELEZ, songe aujourd'hui à la possibilité de cartographier les émotions épistolaires suscitées par Émile Zola lors de l'affaire Dreyfus, en représentant graphiquement le degré d'émotion suscitée par chacune des lettres. En s'appuyant donc sur la correspondance passive de l'écrivain, en provenance de plusieurs pays à travers le monde, les chercheurs pourront étudier les variables d'une réceptivité épistolaire entourant cet événement délimité sur le plan spatio-temporel. Cette exploitation pourrait devenir envisageable via des fouilles de données soumises à ces grands fonds de lettres numériques. Une fois encore, c'est l'indexation numérique des lettres qui

<sup>155</sup> Ibid.

<sup>156</sup> DUMONT, Stephan. « Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF) », Digiversity, 29 octobre 2015, p.10. En ligne : <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif/>

induira l'agrégation de termes significatifs lors de ces recherches automatisées et ciblées. Certains termes pourront être rattachés par les chercheurs d'un point de vue sémantique à l'indignation, ou au contraire au soutien inconditionnel porté à la suite de l'écrivain durant les années 1897 à 1899. Représenter visuellement l'étendue spatio-temporelle de l'expression des subjectivités pourrait s'avérer tout à fait fascinant dans le cadre des correspondances d'écrivains du XIX<sup>ème</sup> siècle, où l'intime prévalait alors dans les lettres. En cela, la plateforme d'archivage numérique des correspondances de Zola pourrait tendre vers la représentation scientifique de la caractérisation de l'intime au XIX<sup>ème</sup> siècle, que nous avons évoqué dans notre première partie. Ces entreprises sont toutefois envisageables en respectant toutes les conditions de l'archive épistolaire numérique que nous avons développé dans notre travail, via une indexation et un agencement des données des lettres qui soit le plus « aléatoire » possible.

Un domaine de la recherche humaniste s'intéresse à ces perspectives spécifiques que nous venons d'énoncer, à savoir la linguistique computationnelle. Cette frange des SHS se propose de traiter le langage naturel et d'enrichir notre représentation de connaissances avec ces informations initialement subjectives. Les lettres, dans leur immédiateté d'exécution, pouvaient en effet amener les écrivains à délaisser leur structure linguistique conventionnelle et à adopter de nouveaux termes dans leur expression. L'intégration d'un thésaurus aux plateformes numériques que nous étudions peut alors être une plus-value certaine, dans cette entreprise d'interprétation du contenu des lettres, pour valoriser le champ sémantique à première vue indécélable au sein correspondance. Dans le cadre du projet du Centre Flaubert, cette idée a déjà émergé notamment à travers l'index thématique que nous avons évoqué. Néanmoins, l'idée d'un thésaurus qui ne serait pas axé sur la figure centrale de l'écrivain, mais sur les relations proprement sémantiques entre chaque mot pourrait amener à un degré supérieur l'interprétation du contenu des lettres de l'écrivain. L'apport de ces nouveaux savoirs interviendra à travers la mise en réseau des données épistolaires sur le web, elle-même rattachée à l'enjeu de leur interopérabilité numérique.

Cette mise en relation des éléments constitutifs de la lettre peut enfin s'opérer au sein de sa dimension génétique, avec la surface de la lettre manuscrite qui aura été numérisé et indexé. Pour illustrer les perspectives en Humanités Numériques qui seraient ainsi envisageables dans ce domaine de la lettre numérique, le « Langage d'encodage génétique » (LEG, anciennement nommé « HyperNietzsche Markup Language », HNML) nous paraît important à étudier. Conçu par Paolo D'Iorio et par son équipe en 2003 dans le cadre du projet HyperNietzsche, ce langage de structuration de la donnée numérique, est une déclinaison du langage XML qui permet d'encoder les manuscrits anciens, parmi lesquels donc la lettre d'écrivain. Il permet d'encoder les caractéristiques matérielles de l'écrit, comme la couleur, le type d'instrument d'écriture utilisé, le type d'alphabet, mais aussi les processus génétiques de l'écriture, comme l'ajout, la rature, la surcharge. Tous ces éléments sont constitutifs de la surface matérielle d'une lettre d'écrivain, comme nous l'avons montré précédemment. Ce langage dispose en outre d'un ensemble de balises pour marquer les interventions de l'équipe scientifique, dans le cadre des annotations explicatives des lettres précédemment évoqués. Nous voyons ici une nouvelle façon d'encourager et de valoriser la participation des « auteurs » à des projets d'édition numérique des correspondances d'écrivains. Ce langage rend palpable toute intervention sur la lettre sans pour autant dénaturer son contenu profond. L'utilisation de ce langage ou de l'un de ses dérivés pourrait s'avérer tout à fait intéressant concernant le projet d'archivage numérique des lettres de Gustave Flaubert, dont le processus d'écriture épistolaire était très marqué comme nous l'avons montré. Cela étant, l'encodage des lettres de l'écrivain au format HTML pourrait une nouvelle fois freiner ces possibilités d'accès directe à la donnée numérique de la lettre. De plus ce langage permettrait un texte épistolaire selon son

intertextualité, à savoir la structure régissant chaque champ de la lettre et qui en fait un objet textuel unique en son genre. La citation ou la reformulation des paroles du destinataire auxquels l'épistolier répond – trait spécifiquement dialogique – pourrait alors être intégré lors de l'encodage de la correspondance et se matérialiser visuellement à l'étude génétique de la lettre. Nous pourrions par exemple avoir un lien hypertextuel relié à une phrase de Gustave Flaubert nous renvoyant vers un message de sa maîtresse Louise Colet, qui a motivé l'écriture de cette phrase à ce moment de la lettre....

### **III.2. Enjeux de l'ouverture des corpus épistolaires aux collections numériques des institutions patrimoniales sur le Web**

Nous allons à présent analyser les perspectives d'une évolution extrinsèque qu'une plateforme d'archivage numérique des correspondances sur le Web pourrait mettre en œuvre. Cette question se posera à travers l'enjeu d'ouverture des données numériques des lettres d'écrivain, vers des collections numériques d'autres lettres sur le Web. Le cadre institutionnel des bibliothèques nous intéressera en premier lieu.

Nous allons montrer les perspectives des collaborations sur le Web entre les équipes scientifiques que nous avons étudié dans notre seconde partie et les institutions culturelles et patrimoniales. L'objectif initial est de combler les lacunes documentaires et technologiques pour ces deux approches différentes des correspondances numériques d'un écrivain... Les spécialistes de l'écrivain peuvent apporter, de par leur connaissance biographique et linguistique une réflexion approfondie quant à l'organisation du corpus de lettre sur la plateforme. Leur appropriation des manuscrits littéraires fait sens. Tandis que du point de vue d'un usager, connaisseur d'un fonds spécialisé comme peut l'être un fonds de lettres d'écrivain, « la politique de numérisation des établissements patrimoniaux se révèle dans bien des cas peu explicite et son résultat en ligne peu compréhensible »<sup>157</sup>.

Les institutions qui pratiquent cette politique à une grande échelle comme la BnF sont vite limitées dans leurs descriptions des données numériques. Le Dublin Core fera foi. Mais les bibliothécaires et les archivistes qui officient sur ces plateformes institutionnelles seront quant à eux habilités à encadrer la pérennisation du corpus numérisé de lettres et son intégration dans un environnement de recherche normatif. Les métadonnées qu'ils utilisent sont certes circonscrites, mais leur utilisation sera toujours au fait des standards archivistiques internationaux. Concernant les évolutions permanentes de tels corpus, comme ceux que nous étudions, cette dimension est à prendre à compte par les équipes scientifiques. Pour cela il y'a nécessité« d'assister les chercheurs dans l'expression de leurs points de vue afin que ceux-ci soient interopérables avec les standard de l'*Open Linked Data* »<sup>158</sup> afin que le format des données des lettres encodées soit découplée de la technologie du diffuseur, autrement dit du serveur de la plateforme d'archivage des correspondances ...

Le site web <https://data.bnf.fr/> regroupe les données de qualité de la BnF dans des pages « Auteurs », « Œuvres », « Thèmes », « Lieux » et « Dates », et les lie sur le web à d'autres ressources. Ce projet s'inscrit dans une démarche d'ouverture des données, définie par le

<sup>157</sup> ROUSTAN Mélanie, *La recherche dans les institutions patrimoniales : sources matérielles et ressources numériques*. Presses de l'Enssib, 2016, p.52.

<sup>158</sup> Ibid. p.55.

Consortium W3C dans la perspective du Web sémantique ou Web de données. La plateforme d'archivage de ces données, CubicWeb, est une plateforme libre de développement d'applications web sémantique, publiée sous la licence LGPL. Elle permet de exporter « des identifiants pérennes pour les documents, les ressources (DOI, Ark, ISSN etc.) et pour les acteurs de la communauté (ORCID, IdRef) afin de contribuer à la construction du Linked Science sur lequel les fournisseurs de service pourront s'appuyer. »<sup>159</sup> Nous retrouvons ici les enjeux entourant les deux projets d'édition numériques de correspondances d'écrivain, précédemment analysés.

La BnF a, depuis le 1<sup>er</sup> janvier 2014, placé l'ensemble de ses métadonnées descriptives (données bibliographiques et d'autorité) sous la « Licence Ouverte /Open Licence » de l'État élaborée par la mission Etalab. L'utilisation de ces métadonnées est libre et gratuite sous réserve du maintien de la mention de leur source et de l'indication de leur date de récupération. Les notices documentaires récupérées le sont au format UNIMARC.

Il est possible de trier les notices par type de notice, auteur, titre, année, date d'ajout dans le panier, date de transfert sur le serveur ftp. Ces dates peuvent être différentes. Le tri par date de transfert peut vous permettre par exemple de transférer à nouveau sur le serveur ftp un même ensemble de notices transféré une première fois et qui n'aurait pas pu être récupéré. Avec le tri par dossiers, toutes les notices s'affichent regroupées par dossier. Nous retrouvons une procédure qui se rapproche sensiblement de celle que l'équipe Zola a adopté sur la plateforme de travail Oméka.

La collaboration avec ces institutions pourrait permettre aux équipes scientifiques d'accéder à un ensemble nouveau de fonds documentaires, visant à élargir leur propre vision des correspondances. Les descriptions des lettres opérées par les bibliothèques sont susceptibles de venir enrichir celles initialement encodées sur la plateforme.

Toutefois, des problèmes d'interopérabilité des données vont assez rapidement apparaître. L'emploi des langage hypertextuels tels que le HTML et le XML n'entre a priori pas en corrélation celui des bibliothèques dans leurs collections spéciales, comportant notamment les manuscrits de lettres. L'encodage dans ces institutions a lieu la plupart du temps sous la forme de données MARC (catalogage lisible par machine) et de descriptions archivistiques codées (EAD), ou dans des collections en ligne sous une variété d'autres formes. Les métadonnées MARC sont spécifiées dans la norme ISO 2709, qui définit un format d'échange informatique de notices bibliographiques entre bibliothèques classiques. Concernant nos corpus épistolaires numériques, le champ de tels métadonnées s'avère bien vite restreint, au-delà de certaines descriptions matérielles pouvant intéresser les équipes scientifiques. Nous pouvons en outre évoquer le modèle RDF, qui décrit les relations entre bases de données graphiques dans lesquelles aucune hiérarchie ou clé primaire n'est nécessaire. C'est le principe de l'ontologie..... Cela pourrait s'avérer être un frein à certaines collaborations institutionnelles...

Mais des perspectives s'ouvrent du côté de l'EAD. Ce standard d'encodage des fonds d'archives est utilisé à une échelle internationale, notamment par les bibliothèques et musées pour décrire leurs collections numériques, et plus particulièrement celles composées de manuscrits. Ce standard d'encodage des données est également basé sur le XML, ce qui est en mesure de faciliter les échanges de données éventuels avec les plateformes d'archivage numérique des correspondances. Les modalités de ce format respectent donc la double tradition historique et critique, précédemment développée, que les équipes scientifiques

<sup>159</sup> MAHE, Annaïg. Prime-CLAVERIE, Camille. « Le défi de l'interopérabilité entre plates-formes pour la construction de savoirs augmentés en sciences humaines et sociales ». ISTE éditions, 2017, p.16. En ligne : [https://archivesic.ccsd.cnrs.fr/sic\\_01511618/document](https://archivesic.ccsd.cnrs.fr/sic_01511618/document)

évoquées dans notre travail se doivent de poursuivre. L'usage de l'EAD va faire naître des collaborations entre les spécialistes des écrivains, qui fourniront leurs données extraites des lettres, et les archivistes spécialisés dans la bibliothéconomie. Ces derniers seront chargés de produire des catalogues de métadonnées, respectant les standards numériques de normalisation et qui seront spécifiques au format de la lettre.

«La mise en ligne des inventaires encodés (XML-EAD) permet des recherches et des navigations non seulement dans chaque inventaire, mais dans l'ensemble des inventaires ainsi encodés et, ce faisant, donne accès (...) à l'ensemble des **références** utiles à chaque recherche»<sup>160</sup>

Ces objets numériques que sont les lettres encodées sur le Web pourraient alors être perçues comme existant, de fait, au cœur de diverses relations sémantiques. Il importe alors de réaliser leurs descriptions avec des prédicats, d'où l'importance d'un vocabulaire standard, commun entre les équipes scientifiques et les milieux patrimoniaux, nécessaire pour enregistrer, conserver et diffuser en permanence les lettres numériques. Cela marquerait également, du côté des bibliothèques prenant part à ces projets, un changement de perspective quant à leur façon de documenter leurs fonds d'archives. « Il y a là une spécificité de l'édition numérique : on éclate le processus éditorial d'une représentation fixe par liste vers une représentation personnalisable par réseau »<sup>161</sup>. Nous sommes alors en droit de nous interroger sur l'ouverture d'un tel champ de données numériques à une plus large communauté de chercheurs, celles-ci étant pour la plupart inédites et de fait décontextualisées. Qui fixerait alors les droits de manipulation, d'enrichissement et d'exécution documentaire lors de la rencontre entre plusieurs plateformes sur le Web ?

Il pourrait s'agir, la plupart du temps, des fournisseurs de service affiliés aux plateformes, qui collectent et moissonnent les données numériques des lettres. Le moissonnage s'effectue à partir de requêtes formalisées à l'adresse de l'entrepôt, les résultats sont alors intégrés dans l'index ou le répertoire du site moissonneur<sup>162</sup>. Ce protocole est utilisé notamment par les Archives Ouvertes et les entrepôts institutionnels, il s'est aujourd'hui largement répandu dans les institutions patrimoniales et notamment les bibliothèques. Il permet entre autre de construire des sites portails thématiques avec uniquement le résultat de requêtes sur les entrepôts repérés sur cette thématique. Reprenons l'exemple de « Mapping the Republic of Letters », dont les outils de visualisation expérimentale des données se sont basées sur les publications extraites des catalogues de la BNF. Ces documents renseignaient notamment sur les lieux de publication des lettres de Voltaire, qui après avoir été recensés par les chercheurs du projet, ont donné lieu à une présentation expérimentale de la diffusion épistolaire des lettres du philosophe des Lumières. Nous voyons donc que l'ouverte et le partage des données numériques de la lettre aux institutions se déroule sur la plan non seulement textuel mais visuel.

« Il y a même maintenant un changement d'optique avec la possibilité de publier facilement et systématiquement de gros corpus d'images numériques. Cela est facilité par l'existence de formats image standards, manipulables et éditables en grande masse dans des bibliothèques numériques fondées sur l'image de l'objet et ses métadonnées associés. »<sup>163</sup>.

<sup>160</sup> DELPIERRE, Nicolas. HIRAUX, François. MIRGUET, François. *Les chantiers du numérique : Dématérialisation des archives et métiers de l'archiviste*, Paris : L'Harmattan, 2012, p.192.

<sup>161</sup> Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017, p.19. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)

<sup>162</sup> Pour accéder à la liste des requêtes : [http://www.bnf.fr/documents/intro\\_oaipmh.pdf](http://www.bnf.fr/documents/intro_oaipmh.pdf).

Ainsi, « Le numérique augmente enfin la dimension référentielle de la correspondance : elle peut alors être véritablement considérée comme un « hub » qui amène à toutes les dimensions de l'étude d'un auteur, en multipliant les liens vers le contexte, la biographie, les œuvres et leurs réceptions »<sup>164</sup>. Seules des collaborations institutionnelles et patrimoniales seraient susceptibles d'obtenir ces résultats d'ouverture. Pour citer Pierre-Marc de Biasi, ce serait l'occasion de tendre vers « le rêve fou de la connaissance intégrale de la vie créative d'un écrivain. »<sup>165</sup> avec ce postulat de la correspondance numérique en libre accès. Nous voyons une nouvelle fois l'impact réel de la multiplier des points d'accès aux lettres numériques. La recréation virtuel d'un ensemble cohérent est à ce prix, quitte à provenir de contributions documentaires dispersées, mais dont le regroupement fait sens. Pour la BnF, l'offre documentaire de la bibliothèque numérique Gallica doit servir à la création de tels ensembles, en concertation avec d'autres établissements. De même, les bibliothèques disposant de ressources numériques complémentaires de Gallica sont invitées à y présenter leurs documents. Ces opérations bibliothéconomiques visent à « la conservation, la diffusion, la mise en réseau de corpus, pour développer les axes de leur exploitation de l'histoire des sciences au champ plus étendu des études historiques et pour informatiser l'enquête scientifique. »<sup>166</sup>. Mais la communication de ces corpus spécialisés devra ratifier les exigences des utilisateurs sur le Web pour faciliter leur prise en charge de ces fonds complexes. Le dernier point à résoudre pour les chercheurs va être justement de définir les usages de ces corpus épistolaires et le public susceptible d'être interpellé non seulement par leur présence, mais également par leur archivage sur le Web. Assurer la permanence d'un texte épistolaire grâce au numérique relève donc d'une coordination non seulement à l'échelle de l'équipe scientifique qui porte le projet, mais à l'échelle d'un environnement encore plus vaste, via la mise en relation des plateformes.

Nous avons par exemple mentionné l'évolution des mouvements littéraires au fil de ce siècle. Pourquoi la rencontre des plateforme d'archivage des correspondances d'écrivains de cette époque n'amènerait-elle pas à repenser ces objets patrimoniaux sous un angle trans-historique ? Pour préciser notre propos, il s'agirait de considérer les lettres d'écrivains en tant qu'œuvres de l'esprit à part entière, dépassant par la même leur statut de documents acquis dans les institutions patrimoniales qui les conservent, sous forme manuscrite, aujourd'hui encore. C'est ainsi que « les correspondances peuvent désormais constituer des corpus textuels, si un mode éditorial spécifique parvient à désancrer les lettres de leur concaténation chronologique – donc de la logique historico-biographique des éditions papier qui, jusqu'à présent, a imposé ce mode de lecture essentiellement documentaire des correspondances »<sup>167</sup>.

L'archivage numérique de ces textes anciens n'a pas encore fait l'objet d'une politique de conservation d'ensemble ni de coordination patrimoniale au niveau international. Ces projets

<sup>163</sup> Ibid. p.14.

<sup>164</sup> Ibid. p.20.

<sup>165</sup> Cf : <http://www.savoirs.essonne.fr/dossiers/les-technologies/communication-tic/tic-et-litterature-un-mariage-heureux/complement/resources/>)

<sup>166</sup> DUPOND, Marie. FERJOUX, Céline. *Usages des patrimoines numérisés : un réseau interdisciplinaire et interinstitutionnel*, Usages des patrimoines numérisés, Paris : Université Sorbonne Paris Cité, 2017, p.64. En ligne : [http://udpn.fr/IMG/pdf/rapport\\_de\\_synthe\\_se\\_udpn.pdf](http://udpn.fr/IMG/pdf/rapport_de_synthe_se_udpn.pdf)

<sup>167</sup> LERICHE, Françoise. *Quel balisage pour les corpus épistolaires numériques ? De l'annotation traditionnelle du "document" à une analyse générique et pragmatique*, Actes du colloque international d'Albi, juillet 2006. En ligne : <http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Leriche.pdf>

d' édition numérique de la correspondance d'un écrivain pourraient donc constituer les premiers jalons d'une mutualisations des recherches menées sur le sujet.

L'évolution des standards de métadonnées, rendant plus flexible l'indexation numérique des lettres, va dans ce sens. La sphère épistolaire pourrait ensuite s'élargir aux artistes du 19<sup>ème</sup> siècle (les premiers photographes de l'histoire, les peintres, les sculpteurs...). Nous pouvons d'ores et déjà relever l'édition numérique des correspondances de Vincent Van Gogh<sup>168</sup>. Et pourquoi pas envisager ensuite l'arrivée de projets consacrées aux lettres échangées par tous les contemporains de ce siècle, qui ont eux aussi pris leur plume et leur feuille afin de témoigner de leur condition de vie... L'attachement à l'objet-lettre traverse après tout les territoires, aussi bien que les siècles.

Nous évoquerons enfin l'ouverture de ces plateforme d'archivage numérique des correspondances d'écrivains aux dispositif *CollEx*. Ceux-ci ont pour objectif de faciliter, sur l'ensemble du territoire, l'accès aux documents nécessaires à la recherche, quelle que soit leur forme. développement de la conservation partagée des collections numériques. Cette problématique de portée internationale deviendra critique dans les années à venir, et les initiatives actuelles (conservation de fichiers numériques par le CINES et par la BNF ) devront donc être amplifiées dans un cadre d'action commun.

Notons par exemple qu'une bourse de numérisation a été octroyé à Sorbonne Université par ce dispositif, pour un projet concernant la correspondance de Frédéric Le Play, ingénieur français du XIX<sup>ème</sup> siècle. Les projets d'archivage numérique des correspondances pourraient ainsi avoir dans l'optique d'orienter le processus de consultation et de visualisation des métadonnées de leurs lettres en tant que forme d'analyse et de recherche savantes sur le Web.

### **III.3. Quels publics et quels usages pour l'archive épistolaire numérique sur le Web ?**

Puisque ces plateformes ne constituent, au final, que le cadre de la nouvelle expérimentation des correspondances d'écrivain sur le Web par l'utilisateur, il nous fallait conclure notre recherche par une approche lui étant toute entière dédiée. Ces perspectives, que nous allons dresser à l'échelle de l'internaute qui découvre et parcourt sur le Web les corpus de lettres numériques, sont avant tout cognitives et humanistes. Deux questions vont sous-tendre notre raisonnement : celle de la médiation du savoir scientifique et celle de la recherche en SHS. Nous verrons dans quelles mesures les corpus des correspondances de Flaubert et Zola offrent des recoupements spécifiquement numériques à exploiter en ce sens. Le partage numérique des informations comprises ces lettres a permis aux équipes scientifiques de confronter leur matériau épistolaire à celui d'autres personnes ou groupes de travail sur le Web. Nous soulèverons également alors les perspectives interdisciplinaires induites par ces ouvertures et ces transmissions de données

<sup>168</sup> Voir Annexe 2.

### III.3.1. Médiation scientifique par le dépouillement des données numériques des lettres au sein de nouvelles plateformes du Web

Dans quelle mesure une plateforme d'archivage des correspondances d'écrivain est-elle susceptible de constituer une source de la recherche et un objet d'étude à part entière sur le web ? Comme nous l'avons montré, l'environnement numérique devra sans cesse réorganiser les correspondances d'écrivain dans leur constitution, prenant de ce fait en compte les possibilités d'accéder au texte source de leur encodage. Les chercheurs qui réalisent ces projets doivent donc prendre conscience de la nécessité d'investir la question des publics et des usages d'un corpus épistolaire sur le Web. Le rôle des plateformes accueillant les correspondance d'un écrivain sur le Web est en effet d'archiver non seulement le corpus de lettres en question mais aussi de faciliter son relais avec l'environnement extérieur à la plateforme...

Nous allons illustrer ce dernier point avec le projet « Biographes »<sup>169</sup>, mis en œuvre par l'Agence Nationale de la Recherche. L'objet d'étude des chercheurs était de montrer l'intrication des savoirs biologiques dans la création littéraire au dix-neuvième siècle. Nous voyons donc que plus un objet d'étude est précis, plus un fonds vaste comme peuvent l'être ceux des correspondances du XIX<sup>ème</sup> siècle aura des chances de répondre à ces requêtes. La littérature française s'est très vite approprié les savoirs biologiques émergents au cours de ce siècle, les a utilisés de manière plus ou moins correcte, les a commentés, les a prolongés dans leurs développements, participant ainsi à leur promotion en un fait culturel global. Gustave Flaubert et Émile Zola ont tous deux cultivé une fascination pour les « modèles » biologiques de leur siècle, qui leur ont en retour servi d'objet d'étude, dans leur entreprise qui consistait à dépendre le réel. Mais les discours développés par ces deux écrivains ne sauraient transparaître qu'à travers leur œuvre romanesque, où, du reste, la fiction tendrait à gommer les traits scientifiques alors esquissés. C'est bien au sein de leur correspondance que l'on retrouvera leurs réflexions sur ces sujets, qui à leur tour, ont infléchi la réception et, parfois, la pensée des savants eux-mêmes. C'est cette circulation des savoirs que le projet « Biographes » aborde, en se focalisant sur les informations issues de la biologie. Ce terme s'incarne à titre de discipline distincte seulement à partir du XIX<sup>ème</sup> siècle et entraîne avec lui l'émergence d'une conscience disciplinaire. Mais pour connaître le rôle effectif de la littérature dans la métamorphose et la diffusion de représentations ou de modèles de pensée devenus prégnants dans la culture, il faut recueillir des informations tangibles, en la matière. Les outils de recherche et de représentation des données que nous avons entrevu au sein de cette partie permettent cette exhaustivité dans l'approche de ce savoir, pour en proposer par la suite une étude de référence. L'interdisciplinarité est au fondement de ce projet, destiné tant aux spécialistes de la littérature qu'aux historiens de la culture et des sciences. À une plus grande échelle, nous pourrions nous demander quels seraient les enjeux épistémologiques, idéologiques et esthétiques du partage libre de ces savoirs ?

Les chercheurs du projet « Biographes » ont donc extraits des lettres de Gustave Flaubert, allant des années 1830 à 1880, qui abordaient la thématique biologique. Ils se sont

<sup>169</sup> <https://biolog.hypotheses.org/>

pour cela appuyés sur la plateformes d'archivage numérique du Centre Flaubert, afin d'accéder par mots-clés successifs aux occurrences qui avaient traits à leur problématique de recherche. L'équipe scientifique du Centre Flaubert a donc donné son accord afin que les chercheurs de « Biographes » puissent accéder au fichier source issu de l'encodage en HTML des correspondances sur le Web. On relève ainsi, dans le corpus tiré de cette extraction de données numériques, près de 90 lettres de l'écrivain pour la période 1830-1844, que les chercheurs ont trouvé intéressant de diffuser. Cela constitue une sélection plutôt conséquente, étant donné l'infinité des sujets abordés par l'écrivain dans ses lettres, comme nous l'avons constaté. C'est pour ces questions de volumes que les chercheurs voulaient à leur tour ont donné l'accès à ce même code source, issu de cette collaboration entre équipes scientifiques sur le Web, aux internautes sur la plateforme. Le gisement de la recherche devient alors concret à partir de cette mise en accessibilité des données numériques, sujettes à réutilisation dans de nouveaux contextes.

« La Correspondance (de Flaubert) des années 1850 défend l'idée d'une science de l'homme inspirée des sciences naturelles. En quête d'une nouvelle forme d'intelligibilité libérée des préjugés moraux et religieux, Flaubert fait de la vie un nouveau transcendantal par rapport auquel il légitime l'invention d'une littérature exposante : elle reposerait sur les mêmes principes d'indépendance que la critique dont il prône l'usage dans la science de l'homme. »<sup>170</sup>.

Mais toutes ces notions ne sont arrivés à maturation dans l'esprit de l'écrivain qu'après des années de réflexion et surtout de confrontation d'idées sur le sujet, facilités par l'échange épistolaire... Aussi, cette sélection de lettres devait permettre de retracer un cheminement personnel et intellectuel de l'écrivain sur une question. De rendre compte en l'occurrence de « l'élaboration syncrétiste de la pensée de Flaubert, qui s'exprime d'ailleurs de manière fragmentée et discontinue dans la correspondance, et sur plus de trente ans »<sup>171</sup>.

Nous trouvons également ici les prémices de la vision naturaliste que Zola définira quelques années plus tard dans son œuvre. Flaubert a porté avant lui un « coup d'œil médical de la vie »<sup>172</sup>, qui transparait au sein de sa *Correspondance*. Il serait alors tout à fait étonnant de constater, rétrospectivement, dans quelle mesure ces deux écrivains, sur lesquels nous nous sommes penchés dans ce travail, se seraient influencés sur la question, d'une façon indirecte ou non. Cela nous ramène à la question de l'ouverture des données numérique de chaque lettre, pour ces deux plateformes. Une fois que la correspondance électronique de Zola sera en ligne, de nouveaux liens pourraient être ainsi à opérer entre ces deux écrivains, leurs lettres, et pourquoi pas celles de leurs contemporains... L'appartenance commune de ces deux projets au Consortium Cahier offre en ce sens des perspectives, en terme d'interopérabilité des données, pour montrer la richesse des échanges intellectuels dans la seconde moitié du 19<sup>ème</sup> siècle, entre personnalités publiques...

Il advenait donc pour les chercheurs du projet « Biographes » de déceler à travers ces lettres plus qu'une simple influence de la biologie dans la pensée de l'artiste. Cela a contraint les acteurs du projet à des approches plus sophistiquées que la description de simples relations entre les documents récoltés. Il advenait donc de créer des bases de type tableurs, permettant l'exploitation et devant encourager les chercheurs à effectuer des « dépouillements » de la donnée épistolaire à partir de ces outils.

<sup>170</sup> SEGINGER, Gisèle, « Éléments pour une biocritique. ». Flaubert : Revue critique et génétique, Institut des textes manuscrits modernes (ITEM). n°13, 2015, p.1. En ligne : <http://flaubert.revues.org/2439>

<sup>171</sup> Ibid. p.7.

<sup>172</sup> Lettre de Flaubert à Louise Colet, Croisset, 24 avril 1852.

Des logiciels libres de publication et de recherche plein texte, pour corpus XML/TEI, vont permettre de jongler entre les documents, les concepts ainsi que les différentes sources du corpus épistolaire. Prenons pour exemple PhyloLogic4, développé par le “projet” ARTFL au sein de l’Université de Chicago, et pierre angulaire de la plateforme numérique du projet « Biographes ». Les liens hypertextes archivés renvoient à des textes déjà numérisés et nettoyés, voire encodés au format TEI-XML dans le cadre d’éditions électroniques comme CorrELEZ mai aussi vers d’autres corpus encodés en HTML comme ce fût le cas pour celui du Centre Flaubert. Ces outils ont permis d’effectuer des recherches en texte intégral dans les lettres de l’écrivain, en ligne ou après téléchargement.

Nous évoquons ici un « système cognitif global qui génère une sorte de documentarisation, dans laquelle un document n’est plus qu’une représentation éphémère d’un besoin ponctuel de connaissance »<sup>173</sup> La lecture d’une lettre répond tout à fait à ce type de découverte culturelle et historique fragmentaire sur le Web. Les perspectives formelles consisteraient à faciliter les interactions disciplinaires afin de générer du sens, à partir d’un même matériau de recherche ; en l’occurrence la lettre d’écrivain. C’est une chaîne à perpétuer pour ces équipes scientifiques sur le Web : l’assimilation des pratiques éditoriales numériques, l’apprentissage des règles de conservation et de sécurisation des données s, la recherche de partenariats culturels cohérents autour de l’archive numérique.... C’est là le prix d’une diffusion extrêmement large, à distance et sans risque de détérioration, du matériau de la lettre d’écrivain et de ce qu’il peut représenter dans toutes les disciplines de la SHS. L’exploitation de ces textes enrichira alors les corpus de lettres à partir de nouvelles découvertes, construisant des corpus secondaires de lettres... Seules l’appropriation et le dépouillement textuel seront à maîtriser par les chercheurs durant ces échanges de données numériques soigneux.

Ces exemples esquissés, parmi lesquels le projet « Biographes », concourent donc à la valorisation des correspondances d’écrivain sur le Web. Ils relaient le contenu épistolaire vers des univers souvent éloignés, à première vue, du monde littéraire. Cette adaptation de la lettre numérique à de nouveaux environnements d’étude conduit les chercheurs qui l’archivent à considérer les rapports entre édition et recherche, tout en se confrontant à la question de la médiation et des usages des résultats de la recherche scientifique. Ces principes sont liés aux conditions de développement et de pérennité qui correspondent au temps long de la recherche scientifique. Ces nouveaux principes sont coordonnés à de nouveaux objets, les données numériques des lettres, en posant la question non seulement de leur conservation, de leur structuration et de leur exploitation, mais aussi de leur portée intellectuelle. L’enjeu pour les chercheurs est de placer le corpus épistolaire au centre de l’expérience usager sur le Web, à condition donc de favoriser les collaborations dans la recherche également à l’échelle interpersonnelles, comme nous allons à présent le démontrer.

### III.3.1. Médiation pour les recherches en SHS à travers l’ouverture de la plateforme d’archivage épistolaire numérique des correspondances

Notre seconde illustration de la médiation scientifique des correspondances sur le Web, portera cette fois sur les interactions entre équipes scientifiques et usager cognitif. L’édition

<sup>173</sup> PEDAUQUE T. Roger, *Le document à l’ère du numérique*, Caen : C&f Éditions. 2006, p.139.

numérique d'une correspondance conduit comme nous l'avons vu avec la biologie à éprouver au sein de ces pratiques la question de des rapports entre domaines de connaissances. Cette interdisciplinarité pourrait encore s'intensifier et se systématiser en prenant en compte la nature collaborative et collective de la pratique scientifique sur le Web. Cela passe concrètement par l'élaboration des outils de traitement commun de corpus de différentes natures pour des exploitations de domaines distincts.

Prenons un exemple tout à fait singulier : celui d'une doctorante en sociologie et agrégée de sciences économiques et sociales à l'ENS Lyon. Frédérique Giraud a réalisé une thèse<sup>174</sup> consacrée à la compréhension des manières d'être écrivain d'Émile Zola et s'est appuyée pour cela très largement sur la correspondance de l'écrivain. Le fil rouge de cette thèse était de reconstruire le point de vue d'Émile Zola au gré de son parcours en société, qui a façonné de fait son engagement dans le monde. Lorsque des objets de recherche sont très précis à ce point, une correspondance numérique s'avère être un formidable outil de fouille textuelle. Nous l'avons vu dans notre première partie de travail ; Émile Zola n'hésitait pas à faire part de ses ambitions en société à travers ses lettres, à partager unilatéralement et donc sans fioritures l'idée qu'il se faisait de sa propre trajectoire personnelle. Les occurrences relatives à chaque discipline du savoir, comme ici celles concernant purement la sociologie, peuvent rapidement être restituées par le truchement d'une base de données épistolaire numérique... L'émergence du projet CorrELEZ, d'où ont jailli des échanges entre Frédérique Giraud et l'équipe Zola, ont d'ailleurs coïncidé avec l'arrivée de la doctorante au sein de l'équipe « Écritures du XIX<sup>ème</sup> siècle » au sein de l'ITEM. Cet exemple est donc frappant sur la plan de la perpétuation des savoirs qu'un corpus épistolaire sur le Web est susceptible d'alimenter, alors même que la plateforme n'était pas constituée.

Cet exemple nous rappelle que derrière les infrastructures en Humanités Numériques et les collaborations institutionnelles, l'émulation face à l'archive numérique des correspondances d'un écrivain prend tout d'abord consistance à l'échelle humaine. Il ne faut pas oublier que c'est « l'utilisateur cognitif qui doit de fait être le pivot, le centre de tout système d'interprétation »<sup>175</sup> au sein d'une plateforme où lui seraient mis à disposition un volume conséquent de données. Les rendre hiérarchisables et modulables, comme le souhaitent les acteurs que nous avons évoqué jusqu'à ce stade de notre recherche, est un acte fort qui marque le décloisonnement du fait culturel tout entier. Cette médiation nouvelle, la correspondance l'incarne de par son genre à soi, qui est protéiforme. Le discours élaboré dans leurs lettres par des écrivains du XIX<sup>ème</sup> siècle tels que Flaubert ou Zola est adressé à un public déterminé qui peut varier du plus restreint au plus étendu. L'archivage des correspondances, qu'il soit physique ou numérique, va donc dès lors nous permettre de nous poser la question de l'usage épistolaire de demain. Certes, dans notre contexte de mondialisation, les lettres ont été remplacées largement par les mails. Mais toutes celles qui ont circulées avant notre ère n'en demeurent pas moins, aujourd'hui, être une formidable piqûre de rappel quant à l'art de communiquer et de circonscrire le champ de nos pensées. L'étude des lettres de ces écrivains a cela d'enthousiasmant que les manuscrits et le texte se complètent tout à fait. C'est le signe qu'aucun des projets d'édition numérique entrevus ne signifiera la fin de l'édition imprimée des lettres ou l'obsolescence des volumes de correspondance existants.

« Pendant longtemps, on continuera à lire les lettres de Flaubert sur papier ; l'édition électronique, complémentaire et non rivale, apportera ses avantages techniques :

<sup>174</sup> GIRAUD, Frédérique. *Ecrire pour résister au déclassement social : analyse sociologique de la carrière et des pratiques littéraires d'Emile Zola*. École Normale Supérieure de Lyon, 2014.

<sup>175</sup> HOLZEM, Maryvonne. LABICHE, Jacques. *Dessillement numérique : énonciation, interprétation, connaissances*. Bruxelles : P.I.E-Peter Lang S.A., 2017, p.156.

l'actualisation immédiate, l'exhaustivité des lettres et des documents explicatifs, et l'invention de chemins de lectures personnalisées. »<sup>176</sup>.

Bien au contraire, le développement d'éditions de lettres enrichies numériquement et archivées sur le Web, pourrait conduire à des éditions imprimées complémentaires de lettres sélectionnées, car le processus de création et d'impression de correspondances sélectionnées deviendrait plus facile. Mais le numérique sera la clé de voûte de ces projets. La question clé sera de savoir si, formellement, ces projets feront gage de fiabilité et lisibilité dès leur prise en charge des corpus de lettres. Dans le même temps, l'exploitation des données épistolaires nouvellement extraites devra respecter les normes d'archivage et d'interopérabilité numérique, afin que leur usage soit facilité par tous les publics qui se sentiront concernées par elles, sur le Web. Qu'elles soient numériques ou imprimées, les lettres représentent « un aspect essentiel de la vie civilisée », et d'autant plus à cette époque charnière qu'est le 19<sup>ème</sup> siècle, qui marque, à son niveau, les prémices des sciences de l'information et de la communication régissant toutes nos activités actuelles. La lettre est un matériau qui a traversé le temps. Le texte qu'elle renferme est porteur d'une mémoire sociale et ne peut être compris que si les conditions de circulation inhérentes à ce support sont restituées dynamiquement. Les outils des Humanités Numériques que nous avons développés dans cette partie en témoignent, avec toutes les réussites, les échecs et les marges d'évolution technique et technologique qui les caractérise. Le tout sans avoir à passer par d'autres intermédiaires que la plateforme sur le Web, établie dans un idéal d'accès ouvert à ses contenus, jusqu'à la source de leur encodage. La valorisation des correspondances est donc au prix d'une visibilité totale du cheminement numérique des lettres, dans la perspective de sa réappropriation toujours plus intime et personnel de l'utilisateur sur le Web.

« Le Web offre ainsi une prépondérance de « mémoire » conjuguée à la première personne, et les témoignages peuvent être transmis directement sans la médiation de l'historien, sans que son sens critique n'orchestre l'écriture de l'histoire, sans différencier les sources et sans les insérer dans un contexte signifiant »<sup>177</sup>. Un processus, à la fois cognitif et historique, qui a le mérite de s'accroître et d'être continuellement représenté sur le Web, dans ce qui est devenu presque « une ère postépistolaire. »<sup>178</sup>

<sup>176</sup> LECLERC, Yvan. JUNGERMANN, Nathalie. « Gustave Flaubert. Correspondance. », *FloriLettres*, 91 / 2008, p.5.

<sup>177</sup> NOIRET, Serge. « La *digital history* : histoire et mémoire à la portée de tous ». *Read/Write Book 2 : une introduction aux humanités numériques*. Marseille : OpenEdition Press, 2016. En ligne : <https://books-openedition-org.docelec.enssib.fr/oep/226>

<sup>178</sup> HANKINS, Gabriel, « Correspondence : Theory, Practice and Horizons », *Literary Studies in the Digital Age: An Evolving Anthology*, 2015.

En ligne : <https://dlsanthology.mla.hcommons.org/correspondence-theory-practice-and-horizons/>

# CONCLUSION

En définitive, nous avons commencé par définir le rôle de la lettre pour l'écrivain du XIX<sup>ème</sup> siècle, en nous axant sur les figures de Gustave Flaubert et d'Émile Zola. Les mouvements littéraires ainsi que les mutations socio-économiques de ce siècle, ont fortement influencé les écrivains dans leur rapport à l'épistolaire. La palette de l'expression artistique s'est considérablement amplifiée, durant cette période, transparaissant lors de la rédaction d'une lettre. En tant que figure sociale emblématique, l'écrivain catalyse les enjeux d'une époque non seulement dans ses œuvres, mais plus encore dans ses productions d'avant-texte, où le filtre de la réception critique a semble t-il moins de prises sur ses positionnements multiples. L'écrivain a ainsi été le premier à discerner l'importance rétrospective de ses écrits dits « secondaires », dans le jargon éditorial alors émergent, au XIX<sup>ème</sup> siècle. Les lettres vont alors devenir des objets historique, notamment à travers la généralisation du legs des correspondances. Nous nous sommes intéressés aux motivations de l'auteur, puis à celles de sa famille, jusqu'à celles des institutions patrimoniales à vouloir pérenniser l'existence de ces fonds de lettres. Cela nous semblait être là le premier maillon d'une exploitation ciblée de ce qui s'apparentait alors, déjà, comme une archive épistolaire. Au XX<sup>ème</sup> siècle, la conservation des lettres manuscrites dans les institutions patrimoniales et leur édition en volumes ont été les deux approches de pérennisation, adoptées face aux fonds physiques de correspondances d'écrivain. Toutefois, ces corpus présentaient des limites archivistiques évidentes en terme de massification documentaire et d'évolutivité permanente, ce qui en rendait souvent l'accessibilité délicate. De même, nous avons relaté l'absence d'une coordination, conventionnée, ou ne serait-ce qu'implicite, entre les institutions de conservation et les chercheurs souhaitant exploiter les lettres d'écrivain du XIX<sup>ème</sup> siècle, pour leur richesse inouïe.

Mais, avec l'arrivée du Web, ces approches ont été totalement bouleversées. Les chercheurs ont profité de cet espace nouveau d'archivage et de diffusion des lettres pour mettre en place des collaborations scientifiques et institutionnelles, parfois même au-delà des frontières. Nous nous sommes alors appuyés sur les projets d'édition numérique des correspondances de Gustave Flaubert et d'Émile Zola sur le Web en décrivant chaque étape de que ces équipes scientifiques devait réaliser en vue d'une exploitation numérique de leur corpus épistolaire. Les modalités de stockage, de traitement, de diffusion, de référencement des données numériques influaient à chaque étape sur la qualité de l'archivage des lettres, en fonction de l'approche choisie par chaque équipe... Nous avons parlé à ce titre des phénomènes d'éditorialisation et de patrimonialisation des correspondances d'écrivains, une fois leur surface manuscrite numérisée et leur transcription numérique opérée. C'est en partant de ce double paradigme, linguistique et historique, que l'appropriation d'un ensemble de lettres dématérialisées pouvait prétendre à lui conférer un nouvel éclairage, une nouvelle contextualisation. Nous avons évoqué le phénomène d'« auctorialité », dans l'approche de ces documents par des chercheurs ou des volontaires, qui consistait à valoriser ces participations tout en conservant une ligne directrice éditoriale et scientifique. Il était alors déjà question de la traçabilité de toute opération menée autour des données épistolaires sur le Web, qui se posera concrètement au moment de l'indexation numérique d'une lettre. L'encodage de la correspondance selon les spécifications Dublin Core ou les balisages en TEI devait alors assurer un affichage précis de cette structure vaste et complexe sur le Web. De surcroît, cette

description devait assurer l'interopérabilité de la moindre lettre, avant son accueil sur la plateforme des équipes scientifiques sur le Web.

Là où, au XX<sup>ème</sup> siècle, le volume physique faisait foi pour accéder à un quelconque contenu textuel éclairé, les plateformes sur le Web sont aujourd'hui la porte d'entrée privilégiée pour tendre à une étude exhaustive et réflexive de la correspondance d'un écrivain. Nous avons néanmoins montré dans quelle mesure ces nouveaux espaces étaient inféodés aux mêmes conditions juridiques et intellectuelle, pour la restitution et l'exploitation du texte épistolaire en ligne. C'est pourquoi le respect des normes de standardisation des données numériques était une garantie pour ces projets, qui en mettant à la disposition du public des documents inédits et historiques, doit pouvoir faciliter leur approche. Mais afin que ces plateformes puissent être considérées en tant que véritable archives numériques, il leur fallait être au centre de comportements de la recherche, de la part de ces équipes scientifiques. Ces travaux ont consisté notamment en une organisation des données épistolaires, à travers les cadres énonciatifs pouvant être développés au sein de ces plateformes, sur le plan sémantique ou encore thématique.

Nous avons donc finalement ébauché quelques perspectives dans la lignée des Humanités Numériques, découlant de ces deux projets étudiés, qui nous semblaient significatives dans l'optique d'une valorisation toujours renouvelée de ces fonds de lettres archivées sur le Web. De nouveaux outils numériques de recherche et de visualisation de la donnée épistolaire permettent aujourd'hui un accès facilité à ces corpus textuels uniques sur le Web. De surcroît, c'est au niveau de l'ouverture de ces corpus aux collections numériques, dédiées à des domaines similaires de la recherche, qu'il faudra envisager de nouvelles collaborations, à commencer par les institutions patrimoniales. Celles-ci se posent après tout dans la tradition archivistique de l'épistolaire et peuvent apporter en cela un regard neuf sur ces plateformes d'archivage numérique des correspondances d'écrivains. Les bibliothèques numériques et les possibilités d'échanges de l'information nous semblaient être des exemples pertinents à étudier, dans une perspective d'enrichissement des métadonnées de la lettre avant leur partage sur le Web. Ces deux formes extensions de la plateforme, en tant que contenu et contenant informationnel, marquent donc des possibilités de valoriser ces fonds de correspondances numériques sur le long terme. Mais pour que ces répercussions soient palpables, la question des publics et des usages de ces nouvelles archives épistolaires nous semblait fondamental à aborder. Les correspondances d'écrivains tels que Flaubert et Zola se posent, de par leur évolutivité intrinsèque et extrinsèque liées aux perspectives humanistiques, comme des objets de recherches tout à faits originaux et susceptibles de créer de nouveaux gisements d'information.

## Annexe N°1 :

### Entretien avec M. Jean-Sébastien MACKÉ, responsable du projet CorrELEZ, pour le Centre d'Étude sur Zola et le Naturalisme (ITEM)

**- Dans le cadre des Archives Zoliennes et de l'Institut des Textes et Manuscrits Modernes, comment vous est venu l'idée d'éditer électroniquement toutes les correspondances de l'écrivain et de proposer ainsi un libre accès à ces documents, à un public plus grand que celui de la sphère des chercheurs ?**

**Jean-Sébastien Macké :** Pour nos projets d'éditions, il faut distinguer la correspondance active et la correspondance passive de Zola. Pour la correspondance active (5000 lettres), les 11 volumes édités entre 1974 et 2010 sont difficilement trouvables, en dehors des bibliothèques universitaires. Il s'agissait donc de rendre disponibles, aux chercheurs comme au grand public, ces 11 volumes, avec les annotations scientifiques, les introductions biographiques et historiques, les bibliographies, etc. De plus, l'édition numérique permet la recherche textuelle, la recherche par mot-clé... Enfin, cela nous permet d'éditer les lettres inédites retrouvées depuis et de proposer le fac-simile de la lettre quand c'est possible.

Pour la correspondance passive (20000 lettres), il s'agit d'un projet très ancien mais qui n'était pas réalisable au format papier. Des correspondants illustres de Zola ont vu leurs lettres éditées, parfois dans une correspondance croisée avec Zola, mais les milliers de lettres d'inconnus restaient inédites. Dans le cadre du Labex TransferS, nous avons obtenu le financement d'un projet visant à étudier les correspondants étrangers de Zola au moment de l'affaire Dreyfus, ce qui permettait de réduire le nombre de lettres (2000) et de faciliter leur édition dans un temps court (2 ans), sachant qu'il fallait d'abord les numériser (d'après des photocopies puis d'après les originaux conservés chez les descendants de Zola).

**- Dans un sens général, quelle source d'intérêt peut présenter selon vous ce matériau, ce médium spécifique de la lettre chez l'écrivain, au-delà de sa propre « littérature » ?**

**J.-S.M. :** Chez Zola, la correspondance de travail tient une très grande place. Elle nous permet donc de suivre, dans certains cas, le processus de création, le travail avec l'éditeur français, avec les traducteurs étrangers. Donc, sur le plan de la génétique textuelle, la correspondance tient une place importante. C'est aussi la possibilité de comprendre un milieu social, la sociabilité des gens de lettres, peintres, journalistes de la seconde moitié du 19<sup>ème</sup> siècle. Si on considère que les lettres ouvertes (*J'Accuse*, par exemple) font partie de la correspondance générale (ce que nous faisons), c'est la possibilité de rendre compte de la vie politique complexe de la France de cette époque. Enfin, les lettres plus intimes permettent de montrer l'homme, le marié, l'amant, le père derrière l'homme de lettres.

**- J'ai également remarqué votre collaboration avec les infrastructures en Humanités Numériques : dans quelle mesure ce domaine de recherche vous a-t-il aidé à consolider l'avancée de votre projet ?**

**J.-S.M. :** Personnellement, le numérique m'a toujours intéressé. De plus, l'équipe Zola s'est très tôt investie dans le numérique (création du site des *Cahiers naturalistes* en 2000, édition numérique du *Rêve* sur Gallica en 2001). Puis, en 2011, nous avons obtenu le financement de l'ANR pour le projet ArchiZ : édition numérique des archives d'Emile Zola. C'est ce projet

qui nous a amenés à ce qu'on appelle maintenant les humanités numériques. Avec des tâtonnements, des erreurs, des avancées tout de même. Je me suis moi-même formé à la TEI, intégré nos projets d'édition numérique au consortium CAHIER de manière à pouvoir participer aux différents groupes de travail du consortium, à rencontrer d'autres porteurs de projets qui rencontrent les mêmes difficultés que nous ou qui ont déjà trouvé des solutions. Enfin, l'ITEM possède une équipe Humanités numériques avec laquelle nous travaillons beaucoup (autour de la plateforme EMAN : édition de manuscrits et d'archives numériques). En définitive, le projet ArchiZ nous a permis de mieux structurer nos projets numériques, ce qui fait que nos projets d'édition de correspondance avancent plus vite et avec de meilleurs résultats car ils ont bénéficié d'une première expérience riche d'enseignements. Je dirais que, maintenant, nous avons les outils nécessaires, ainsi que les compétences, à l'édition numérique selon les standards internationaux en vigueur actuellement.

**- Une importante communauté de chercheurs prend part à ce projet : comment s'articulent les différentes phases de travail entre ces spécialistes ? Suivent-ils un plan défini par votre équipe à l'origine ?**

**J-S.M.** : Correspondance active : Cette édition réunit une équipe restreinte de chercheurs (5-6 personnes), constituant le moteur de l'équipe Zola de l'ITEM. Je prépare les données (OCR des lettres) qui sont ensuite relues par les différents chercheurs. Simplement, nous avons fait un premier test sur le 1<sup>er</sup> volume avec une chercheuse qui a procédé à la relecture, afin de rédiger un petit guide de recommandation pour la relecture des volumes suivants. Nous allons maintenant procéder à ces relectures durant les prochains mois puis importer les contenus dans notre édition numérique.

Correspondance passive : Ici, l'idée est un peu différente. Tout en partant de l'équipe Zola de Paris, nous avons souhaité fédérer nos collègues français et étrangers autour de ce corpus de lettres. Nous souhaitons ainsi bénéficier de leur expertise selon des aires géographiques (Europe de l'Est, Europe du Nord, Amérique du Nord, Amérique du Sud, Orient, etc...). Donc, nous avons très rapidement mis les lettres à leur disposition de manière à ce qu'il puisse y travailler individuellement. Nous avons organisé des bilans réguliers par des réunions en présentiel et avec vidéoconférence pour les collègues étrangers. Avec eux, nous avons défini le travail qu'il était possible de réaliser sur ces lettres : relire, corriger et compléter les métadonnées DublinCore, proposer une transcription et/ou une traduction. Mais nous n'avons pas demandé un travail systématique sur l'ensemble des lettres. Enfin, nous avons proposé comme étape le colloque que nous avons organisée les 23 et 24 mai 2019 : « Naturalismes du monde : les voix de l'étranger ». L'idée du colloque était d'inviter les collègues participant au projet afin de proposer une analyse de leurs corpus par aires géographiques. A cet effet, nous avons conçu avec eux une grille d'analyse possible.

**- Dans quelle mesure le format Dublin Core a-t-il été pertinent à adopter en ce qui concerne la description des métadonnées de ces documents, spécifiquement ?**

**J-S.M.** : Les champs DublinCore nous semblent suffisants pour une description minimale des contenus. Disons que c'est le minimum exigible. Dans le cadre d'un projet collaboratif comme le nôtre, les 15 champs DublinCore ont aussi l'avantage de ne pas décourager les collaborateurs, souvent peu férus d'informatique et pas habitués à créer des métadonnées.

Dans EMAN, nous avons des métadonnées descriptives supplémentaires, que nous pouvons créer selon nos besoins spécifiques. Ainsi, avons-nous créé des champs complémentaires pour la correspondance, la presse, les documents pédagogiques.

**- D'autres universités et bibliothèques ont-elles été impliquées ? Si oui, comment votre équipe a-t-elle amené ces institutions du savoir à se joindre à cette aventure ?**

**J-S.M. :** Pour le moment, aucune autre institution n'est véritablement impliquée. Probablement que nous ferons appel à la BnF, avec qui nous travaillons habituellement, lorsque nous déciderons de mettre les facsimile des lettres de Zola en ligne, si celles-ci sont à la BnF. Et nous serions amenés à faire de même avec de nombreuses institutions françaises et étrangères. Mais cela viendra dans un second temps du projet, la priorité étant la mise en ligne du texte des lettres. Nous avons surtout travaillé avec les descendants de Zola, son arrière petite-fille en particulier, que nous connaissons bien et depuis longtemps. Elle a accepté que nous fassions la numérisation et la mise en ligne sur une plateforme de travail. Nous devons maintenant discuter avec elle afin de rendre la plateforme public, en officialisant notre collaboration par une convention écrite.

**- Comment avez-vous procédé concernant l'édition numérique de ces correspondances ? Via un encodage du contenu en Text Encoding Initiative puis en liant les documents numériques et en constituant leur corpus ? Via un logiciel documentaire, en tant que premier outil de description des correspondance, avant leur encodage numérique par la suite ? Selon une autre méthode d'édition numérique ?**

**J-S.M. :** Pour les deux projets, nous préparons les données dans EMAN (Omeka). Nous procéderons ensuite à un export de la correspondance de Zola vers un site développé en WordPress (par un prestataire extérieur), avec un thème dérivé du site ArchiZ. Cette correspondance sera donc disponible à cette adresse : <https://www.correspondance-zola.fr/>

Ce site est lui-même inspiré de l'édition numérique de la correspondance de Van Gogh. L'idée est d'avoir plusieurs fenêtres ouvertes en même temps et de pouvoir choisir celles que l'on veut ouvrir ou fermer, de pouvoir les déplacer horizontalement : facsimile, transcription, notes d'origine, notes secondes (nouvelles notes ajoutées à l'édition papier, afin d'apporter une précision nouvelle, une correction, etc...). Donc, pour le moment, il n'y a pas d'encodage TEI. Cela viendra dans un second temps. Mais j'ai déjà développé un schéma TEI, grâce à une collègue de l'ENS-Lyon spécialiste de ce domaine, qui permet un encodage simple d'une correspondance. Mais il y a 5000 lettres à encoder. Donc il va se poser la question de la méthode : Oxygen ou utilisation d'une plateforme collaborative de transcription (plateforme TACT du groupe ELAN de l'université de Grenoble).

**- Au gré de mes recherches, j'ai pu relevé divers problématiques sur le plan technique autour de votre projet : la création d'une interface de saisie du code TEI permettant de générer du XML automatiquement, un plugin d'affichage de texte à Wordpress, ainsi que la création d'outils d'exploitation en balisage des données. Qu'en est t-il actuellement de ces projets ?**

**J-S.M. :** Dans EMAN, nous avons créé un outil de transcription qui permet l'encodage TEI sans entrer dans le code xml. L'outil fonctionne pour tous les projets hébergés sur EMAN. Il ressemble à TACT mais a moins de fonctionnalités concernant le crowdsourcing. L'export des données d'EMAN vers Wordpress fonctionne également, il a été développé par notre prestataire extérieur.

**- Serait t-il possible d'avoir un extrait du schéma TEI que vous escomptez mettre en œuvre avec votre équipe prochainement ?**

**J-S.M.** : Concernant le schéma TEI, nos ambitions sont modestes et nous devons encore faire un choix en terme de granularité dans le développement du code. Voici donc la balise <correspDesc> telle que nous l'envisagons :

```
<correspDesc>
  <correspAction type="sent">
    <persName>Zola, Émile</persName>

    <!-- <dc:date>, obligatoire, unique -->
    <!--date de rédaction de la lettre-->
    <date when="1856-06-14">14-06-1858</date>

    <!--Lieu de rédaction (Ville, lieu-dit, région, lieu relatif...)-->
    <settlement>Paris</settlement>
  </correspAction>
  <correspAction type="received">
    <persName>Paul Cézanne</persName>
    <settlement>Aix-en-Provence</settlement>
  </correspAction>
</correspDesc>
```

**- J'ai cru comprendre votre volonté de proposer un appareil critique lié notamment à la correspondance active de l'écrivain. Cette facette prendra t-elle la forme de rubriques explicatives sous la forme d'onglets à part entière au sein de votre plateforme, ou bien d'annotations collaboratives ?**

**J-S.M.** : L'apparat critique est d'abord celui des 11 volumes de la correspondance générale. Il est extrêmement précis et savant, il serait dommage de s'en priver, d'autant que les spécialistes de Zola ont l'habitude de l'utiliser. Mais nous avons eu l'idée de pouvoir associer un appareil critique nouveau (que nous appelons, pour le moment, « notes secondaires) afin de compléter éventuellement les notes existantes, d'en ajouter de nouvelles. Ceci est d'ailleurs nécessaires pour les lettres inédites. Donc, ces notes secondaires se trouvent sur le même plan que les notes primaires. L'annotation n'est pas collaborative mais réservée à l'équipe Zola. Mais on pourrait réfléchir à la possibilité offerte à l'utilisateur de proposer des notes. Une des difficultés concernant ces notes est le renvoi à une lettre qui est très souvent utilisé dans la correspondance générale : voir lettre 144, note 3 par exemple. Dans le passage à l'édition numérique, « lettre 144 » n'a plus de sens. Il nous faudra donc transformer ces notes en lien vers la lettre suggérée, grâce à une table de correspondances. Mais cela représente plusieurs milliers de notes à modifier.

**- J'ai remarqué votre intérêt pour l'analyse codicologique des manuscrits d'écrivains. Aviez-vous des exigences au niveau de la transcription de texte des lettres manuscrites d'Emile Zola? Par quelle moyen avez-vous procédé à cette opération complexe (crowdsourcing, OCR...)?**

**J-S.M.** : Effectivement, la correspondance générale imprimée ne rend pas compte (ou très peu) de la matérialité du support d'écriture. Le numérique permet de rendre cette matérialité

par l'affichage du facsimile. Ensuite, dans les champs DublinCore, nous avons le champ « format » qui nous sert à décrire la matérialité du document. Mais d'abord pour les lettres de l'étranger envoyées à Zola. Ce champ est rempli au moment de la numérisation du document mais de manière succincte, vu le volume à numériser et décrire. Pour les lettres de Zola, il nous faudrait aller à la source, dans les différentes institutions, ce qui n'est pas possible pour le moment.

**- Comment avez-vous réussi à circonscrire votre corpus, avec un nombre aussi conséquent de documents à analyser et à traiter ? Y' a t-il des lettres auxquelles vous avez du renoncer dans cette entreprise archivistique ?**

**J-S.M.** : Pour la correspondance de Zola, nous n'avons pas de limite. Nous prenons les 4700 lettres déjà éditées et nous ajoutons les lettres inédites, pour arriver à un volume d'environ 5000 lettres. Et nous ajouterons les lettres retrouvées au fur et à mesure. Donc, le corpus n'est jamais limité.

Pour la correspondance passive, nous avons limité un premier corpus à la période de l'affaire Dreyfus (fin 1897-fin 1899) et aux lettres envoyées de l'étranger, soit 2000 lettres environ (donc 10% du corpus général estimé). Nous pourrions ensuite prendre en charge d'autres corpus, toujours en trouvant des limites temporelles ou géographiques (par exemple, lettres de l'affaire Dreyfus envoyées de France, lettres liées à une période d'écriture d'un roman). Ainsi, nous tendrons progressivement à l'édition de l'ensemble des 20000 lettres envoyées à Zola. Mais ça n'est pas l'objectif initial.

**- Votre corpus de lettres pourra t-il être amené à évoluer dans son contenu, au-delà de la correspondance active/passive liée à Emile Zola, ou dans sa forme (exemple : de nouvelles fonctionnalités telles que les cartes heuristiques ou d'autres outils de visualisation de corpus...) à l'avenir ?**

**J-S.M.** : L'idée est de proposer aux spécialistes l'édition de correspondances secondaires, liées à Zola. Par exemple, pour ce qui me concerne, l'édition de la correspondance entre Alfred Bruneau (compositeur) et Louis Gallet (librettiste) pour la mise en musique du *Rêve* et de *L'Attaque du moulin*. Donc, il est possible d'éditer des corpus où Zola est concerné mais où il n'apparaît pas en tant que scripteur.

Concernant la visualisation, nous travaillons sur la géolocalisation et la cartographie. Chaque lettre fait l'objet d'un point de géolocalisation sur une carte mondiale. On peut donc accéder aux ressources à partir de la carte (ce qui pose des problèmes car le monde du 19<sup>ème</sup> siècle n'est pas celui d'aujourd'hui : Autriche-Hongrie, empire Ottoman, etc...). Concernant la cartographie, je réfléchis à la possibilité de cartographier les émotions épistolaires, le degré d'émotion de chacune des lettres, afin de voir s'il y a des zones de France ou du monde où l'on est plus ou moins dans le soutien ou l'attaque à Zola au moment de l'affaire Dreyfus. C'est un chantier encore en réflexion.

**- Quels sont les apports du numérique, en rapport à ce que l'édition traditionnelle sur support papier ne permettait pas ? Notre appréhension du document littéraire est-elle bouleversée grâce à cette interface ?**

**J-S.M.** : L'apport du numérique est déjà de pouvoir procéder à une recherche textuelle, par mots-clé, de procéder à des analyses statistiques, lexicométriques, etc... Probablement que cela reconfigure en partie notre lecture de cette correspondance. Dans l'édition papier, chaque volume peut se lire du début à la fin, dans l'ordre chronologique, de manière à lire au jour le jour le romancier à l'œuvre dans son travail, dans son intimité. Le numérique incite davantage à une lecture parcellaire, en fonction d'une recherche bien précise. Ce que le papier permet aussi. Donc, je ne pense pas que notre appréhension du document littéraire en soit

bouleversée. Par contre, cela remet au centre la correspondance passive, celle à laquelle Zola ne répond pas car elle provient d'inconnus, les lettres arrivent par dizaines chaque jour et qu'il est ensuite en exil en Angleterre. Cette correspondance qui n'avait jamais été étudiée apparaît comme un formidable document sur la réception du naturalisme et de l'affaire Dreyfus dans le monde entier.

**- D'un point de vue technique, la réalisation de cette plateforme d'archivage des correspondances de Zola sur le Web est-elle en passe de correspondre à ce que vous souhaitiez initialement à son sujet, en terme de d'accès et de navigation?**

**J-S.M.** : Oui, nous parvenons vraiment au résultat espéré en terme de navigation, grâce à nos informaticiens qui sont à l'écoute de nos demandes. EMAN est développé de manière identique pour tous les projets hébergés sur la plateforme, la personnalisation pour chaque projet est limitée. Par contre, WordPress nous permet de développer une interface qui correspond à nos attentes. Il ne semble pas y avoir de limites techniques car, finalement, nos demandes sont relativement simples.

**- Et du point de vue intellectuel ? Ce vaste corpus Web présente t-il selon vous une nouvelle image de l'artiste, de nouveaux enjeux historiographiques l'entourant ?**

**J-S.M.** : Encore une fois, je ne pense pas que cette édition bouleverse notre connaissance du romancier, la représentation que nous pouvons en avoir. Mais je me place du côté des spécialistes. Peut-être que cette correspondance, une fois rendue publique, donnera la possibilité à un public plus large d'entrer dans le quotidien épistolaire du romancier. Nous l'encourageons fortement, notamment du côté des publics scolaires (en travaillant avec le rectorat de Paris, les enseignants du secondaire).

Par contre, la correspondance passive de correspondants inconnus offre au spécialistes la possibilité de recherches nouvelles qui devrait nous donner une image nouvelle, ou plus précise, du renom de Zola et du naturalisme dans le monde entier. Dans le cadre des transferts culturels, cette correspondance offre des perspectives nouvelles.

**- Que pensez-vous à terme d'un réseau de corpus de correspondances d'écrivains du XIX<sup>me</sup> siècle, archivé en ligne, sur la base de votre travail ? Quelles perspectives pourraient découler d'une telle entreprise de mutualisation des travaux en édition numérique ?**

**J-S.M.** : C'est une idée que nous avons. Notre correspondance doit pouvoir être interopérable avec la correspondance numérique de Flaubert par exemple. Ce qui sera tout à fait possible d'autant que nous connaissons bien les flaubertiens de Rouen. Ce réseau de correspondances d'écrivains du 19<sup>ème</sup> (et au-delà) s'est déjà constitué dans le consortium CAHIER, où nous avons constitué un groupe de travail autour de la correspondance, qui a donné lieu à l'édition d'un guide méthodologique pour l'édition numérique de correspondances : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHER.pdf)

Idem dans EMAN, où plusieurs projets de correspondances sont développés (correspondance de Baudelaire par exemple). Les perspectives seraient probablement de montrer la richesse des échanges intellectuels dans la seconde moitié du 19<sup>ème</sup> siècle, entre romanciers, mais aussi avec les peintres, les musiciens, les journalistes, le monde politique, etc... Ca n'est pas forcément nouveau mais le numérique nous permet d'étudier à l'infini ces réseaux sociaux, de les visualiser par cartographies, graphes....

## Annexe N°2 :

# Présentation visuelle des plateformes d'archivage ou d'exploitation numérique des correspondances évoquées :

### - Centre Flaubert :

CENTRE FLAUBERT CÉRÉDI

Contacts | Mentions légales

## Correspondance

Édition électronique  
par Yvan Leclerc et Danielle Girard

- Copyright
- Actualités : vente de lettres
- Appel à manuscrits
- Lettres inédites
- Nombre de lettres consultables : 5599
- Lettres écrites par Flaubert : 4488
- Lettres reçues par Flaubert : 1111

**CONSULTER**  
Présentation des lettres :  
par ordre chronologique  
par correspondant  
par lieu de rédaction  
par lieu de conservation

**ATELIER**  
Thèmes des lettres  
Index des noms propres  
Cartes du voyage en Orient  
Ancienne édition Louis Conard

**ÉDITION**  
Présentation  
Principes de transcription  
L'Équipe  
Partenaires et soutiens

**GUSTAVE FLAUBERT**  
1821-1880

CENTRE FLAUBERT

Correspondance de Flaubert

<- Retour | Accueil | Recherche | Recherche multicritères | Lettres inédites | Index des noms propres | Suggestions | Codes

INDEXATION THÉMATIQUE : Les thèmes de la correspondance | Index thématique

PRÉSENTATION DES LETTRES PAR : ordre chronologique | correspondant | lieu de rédaction | lieu de conservation | Accès direct à une lettre

Flaubert à Ernest Chevalier, Rouen, 21 août 1839 | Correspondant | Notice | Télécharger en pdf

Masquer la transcription

À ERNEST CHEVALIER

[Rouen, 21 août 1839.]

Mercredi 21.

Est-ce que ce cher Ernest m'aurait gardé rancune pr la brièveté de ma dernière lettre ? Je suis sûr que c'est un trop bon bougre et qu'il n'y pense pas plus qu'il ne pense à me répondre gredin que tu es (j'abandonne ici la troisième personne) tu ne peux prtant pas alléguer tes nombreuses occupations, car je crois que tu n'as rien à faire qu'à fumer et à te chauffer les couilles au soleil. prquei c'est là ce que je fais maintenant et ce que nous ferons encore plus merveilleusement quand nous allons être ensemble ce qui sera sous peu. Me voilà sorti de la classe de rhétorique. Dans un an à nous deux oh alors quels dînés quels dînés ces messieurs se seront donnés ou plutôt se donneront se fouteront s'ingurgiteront s'introduiront - quelle soupers\* et festins salons de 420 couverts quelle pantagruélisme ! quel punch pr dire adieu au Collège - Je ne dois pas prtant me plaindre, car l'année de rhétorique passée sous ce divin père Magnus m'a paru assez courte et si l'autre ne me paraît qu'une fois plus longue je serai fort heureux.

La plateforme d'archivage numérique des correspondance de Gustave Flaubert (accessible via cette adresse Web : <https://flaubert.univ-rouen.fr/correspondance/edition/>)

## - Omeka :



Capture d'écran de la plateforme Omeka exploitée par les chercheurs du projet CorrELEZ. Cette image est issue de l'intervention de Jean-Sébastien Macke au cours de la Conférence intitulée « Usages pédagogiques et didactiques des archives numérisées : le cas d'Emile Zola », qui s'est tenue le 10 février 2016 à l'ITEM de Paris.

## - CorrELEZ :



Capture d'écran de la plateforme d'archivage numérique des correspondances de Zola issue du projet CorrELEZ. Cette image est issue de l'intervention de Jean-Sébastien Macke au cours de la Conférence intitulée « Usages pédagogiques et didactiques des archives numérisées : le cas d'Emile Zola », qui s'est tenue le 10 février 2016 à l'ITEM de Paris.

## - Vincent Van Gogh : « The Letters »

005 005 « 004 | 006 »

To Theo van Gogh. The Hague, Monday, 17 March 1873. SEARCH THIS LETTER PRINT

original text + line endings facsimile translation notes artworks

1v:1  
den Haag 17 Maart 1873

005  
Br. 1990. 005 | CL: 5  
From: Vincent van Gogh  
To: Theo van Gogh  
Date: The Hague, Monday, 17 March 1873  
[more...](#)

original text + line endings facsimile translation notes artworks

1v:2  
1v:3

Waarde Theo,  
Het wordt tijd dat je weer eens iets van mij hoort, & ik ben ook verlangend om te hooren hoe het je gaat & hoe of het met Oom Hein is, en hoop dus dat je mij als je tijd kunt vinden eens schrijven zult.

Je hebt zeker al gehoord dat ik naar Londen ga, & dat waarschijnlijk heel spoedig. [1](#) Ik hoop zoo wij elkaar voor dien tijd nog eens zien zullen.

Als ik maar eenigzins kan, ga ik met Paschen naar Helvoirt, [2](#) maar dat zal afhangen van de nouveautés, waar Iterson mede op reis gaat. [3](#) Voor hij [1v:2](#) terug is zal ik niet weg kunnen.—

Het zal te L. een heel ander leven voor mij zijn, daar ik waarschijnlijk alleen op een kamer zal moeten wonen & dus voor veel zaken zal moeten zorgen waar ik nu geen last van heb.

Ik ben zeer verlangend om L. te zien, zoo als je wel denken kunt, maar toch spijt het mij van hier te moeten gaan. Nu het bepaald is dat ik weg moet merk ik pas hoe ik aan den Haag gehecht ben. Maar enfin, dat is nu niet anders & ik ben van plan de dingen maar niet te zwaar op te nemen.— Ik vind het heerlijk voor het Engelsch, ik kan dat wel goed verstaan, maar spreken gaat toch lang zoo goed niet als ik zou willen. [4](#)

Aperçu de la plateforme d'archivage numérique des correspondances de Vincent Van Gogh (accessible sur le Web via cette adresse : <http://vangoghletters.org/vg/>)

## - Mapping The Republic of Letters



La plateforme « Mapping the Republic of Letters » présente, entre autres représentations visuelles de ses corpus épistolaires numériques, des cartes interactive matérialisant, les échanges épistolaires qui eurent lieu notamment durant la période des Lumières en Europe.

# Bibliographie

## Ouvrages généraux sur les correspondances d'écrivains :

- BOSSIS, Mireille. *L'épistolarité à travers les siècles*. Stuttgart : Franz Steiner, 1990.
- CHARTIER, Roger. *La correspondance. Les usages de la lettre au XIXème siècle*. Paris : Fayard, 1991.
- DIAZ, Brigitte. *L'épistolaire ou la pensée nomade*. Paris : PUF « Ecriture », 2002.
- DUCHÊNE, Roger. *Comme une lettre à la poste*. Paris : Fayard, 2006.
- GRASSI, Marie-Claire. *Lire l'épistolaire*. Paris : Armand Colin, 1998.
- KHOUZEIMI, Sami. *L'interaction épistolaire au XVIIIème siècle*. Université d'Orléans, Pôle Universités Centre Val de Loire, 2013. En ligne : [ftp://ftp.univ-orleans.fr/theses/sami.khouzeimi\\_3194\\_vm.pdf](ftp://ftp.univ-orleans.fr/theses/sami.khouzeimi_3194_vm.pdf)
- LANSON, Gustave. *Choix de lettres du XVIIIème siècle*. Paris : Hachette, 1921.
- LE BRAS, Laurence. *Les correspondances : réception ; inventaire ; communication*, Ensib : Mémoire d'étude DCB, 2002.
- MARTIN, Phillipe. *La correspondance : le mythe de l'individu dévoilé ?*. Presses universitaires de Louvain, 2015.
- TISSERON, Serge. *L'intimité surexposée*. Paris : Hachette Littératures, 2002.

## Articles sur les correspondances d'écrivains du XIXème siècle:

- CHOTARD Loïc. « Correspondances : une histoire illisible ». *Romantisme*, n°90, 1995, pp. 7-26. En ligne : [https://www.persee.fr/issue/roman\\_0048-8593\\_1995\\_num\\_25\\_90](https://www.persee.fr/issue/roman_0048-8593_1995_num_25_90)
- DAUPHIN, Cécile. « Les correspondances comme objet historique ». *Sociétés et représentations*, n°13, 2002, pp. 43-50. En ligne : <https://www.cairn.info/revue-societes-et-representations-2002-1-page-43.htm>
- DIAZ, José-Luis. « Le XIXe siècle devant les correspondances ». *Romantisme*, n°90, 1995, pp. 7-26. En ligne : [https://www.persee.fr/doc/roman\\_0048-8593\\_1995\\_num\\_25\\_90\\_3049](https://www.persee.fr/doc/roman_0048-8593_1995_num_25_90_3049)
- FERREYROLLES, Gérard. « L'épistolaire, à la lettre ». *Littératures* 2010/1, n°71, pp. 5-27.
- HOOCK-DEMARLE, Marie-Claire. « Correspondances féminines au XIXème siècle : de l'écrit ordinaire au réseau », *Clio : Femmes, Genre, Histoire*, n°35, 2012, pp.67-88. En ligne : <https://journals.openedition.org/clio/10507>
- JAUBERT, Anna. « La correspondance comme genre éthique ». *Argumentation et Analyse du Discours*, 5 | 2010, pp. 1-12. En ligne : <http://journals.openedition.org/aad/985>
- MONTENOT, Jean. « Ce que révèle la correspondance des écrivains ». *L'Express*, 01/07/2007. En ligne : [https://www.lexpress.fr/culture/livre/ce-que-revele-la-correspondance-des-ecrivains\\_812403.html](https://www.lexpress.fr/culture/livre/ce-que-revele-la-correspondance-des-ecrivains_812403.html)

- PAGES, Alain. « Stratégies textuelles : la lettre à la fin du XIX<sup>ème</sup> siècle ». *Littérature*, n°31, 1978, pp.107-116. En ligne : [https://www.persee.fr/doc/litt\\_0047-4800\\_1978\\_num\\_31\\_3\\_1168](https://www.persee.fr/doc/litt_0047-4800_1978_num_31_3_1168)
- PROCHASSON, Christophe. « Les correspondances : sources et lieux de mémoire de l'histoire intellectuelle ». *Varia*, 8 / 1991, pp. 1-5. En ligne : <https://journals.openedition.org/ccrh/2824>
- SIMONET-TENANT, Françoise. « À la recherche des prémices d'une culture de l'intime ». *Itinéraires*, 2009 / 4, pp.39-62. En ligne : <https://journals.openedition.org/itineraires/1466?lang=en>
- STANLEY, Liz. « The Epistolarium : On Theorizing Letters and Correspondences », *Auto/Biography*, n°12, 2004, pp.201-235. En ligne : [https://www.researchgate.net/publication/250263857\\_The\\_Epistolarium\\_On\\_Theorizing\\_Letters\\_and\\_Correspondences](https://www.researchgate.net/publication/250263857_The_Epistolarium_On_Theorizing_Letters_and_Correspondences)
- TOUATI, Paul. « De la médiation épistolaire dans la construction du savoir scientifique ». *Revue d'anthropologie des connaissances*, 2010/3 (Vol 4, n°3), pp.451-475. En ligne : <https://www.cairn.info/revue-anthropologie-des-connaissances-2010-3-page-451.htm>

### Ouvrages et articles sur les manuscrits d'écrivain :

- CRASSON, Aurèle. « Archives manuscrites littéraires : l'apport du numérique pour l'édition et la recherche scientifique », *Genesis*, 30 | 2010, pp.43-47. En ligne : <http://journals.openedition.org/genesis/112>
- CRASSON, Aurèle. MASPERO, Janeta. « Annotation collaborative en ligne de l'archive manuscrite ». *Genesis*, 27 | 2006, pp.165-166. En ligne : [https://www.persee.fr/doc/item\\_1167-5101\\_2006\\_num\\_27\\_1\\_1398](https://www.persee.fr/doc/item_1167-5101_2006_num_27_1_1398)
- DORD-CROUSLE, Stéphanie. « Les archives de Flaubert : conserver pour prouver ». Claude Millet. *Les archives au XIX<sup>e</sup> siècle. Nouveaux partages, nouveaux usages*, Centre de ressources Jacques-Seebacher, 2017. En ligne : <http://seebacher.lac.univ-paris-diderot.fr/bibliotheque/items/show/41>
- HENRYOT, Fabienne. *L'historien face au manuscrit*. Presses universitaires de Louvain, 2012, 366 p.
- LECLERC, Yvan. « À la recherche des manuscrits de Flaubert... ». *Le Blog Gallica*, 20 octobre 2017. En ligne : <https://gallica.bnf.fr/blog/20102017/la-recherche-des-manuscrits-de-flaubert>
- LERICHE, Françoise. MEYNARD Cécile, *De l'hypertexte au manuscrit*. Grenoble : Uga, 2008.
- LERICHE, Françoise. PAGES, Alain. *Genèse et correspondance*. Paris : Éditions des Archives contemporaines, 2012.
- PINÇON, Juliette. « Les archives des écrivains, leur place en bibliothèque ». *Enssib : Mémoire d'étude DCB*, 2017. En ligne : <https://www.enssib.fr/bibliotheque-numerique/documents/67311-les-archives-des-ecrivains-leur-place-en-bibliotheque.pdf>
- SANDRAS, Agnès. VIAL, Charles-Éloi. « Zola à la Bibliothèque nationale ». *Genesis : Manuscrits – Recherche – Invention*, 42 / 2016, pp.145-152. En ligne : <https://journals.openedition.org/genesis/1660>

- SIMONET-TENANT, Françoise. « La critique génétique : définition, intérêts, limites ». *Génétique*, © Publications numériques du CÉRÉDI, « Ressources », 2019. En ligne : <http://publis-shs.univ-rouen.fr/ceredi/index.php?id=603>

### Corpus et éditions numériques des correspondances d'artistes :

- Groupe « Correspondance » du Consortium Cahier. *L'édition numérique de correspondances : guide méthodologique*. Consortium Cahier, Huma-Num, 2017. En ligne : [https://cahier.hypotheses.org/files/2018/03/Correspondance\\_CAHIER.pdf](https://cahier.hypotheses.org/files/2018/03/Correspondance_CAHIER.pdf)
- DUCHATELET, Bernard, LE GUILLOU, Louis. *Petit guide de l'éditeur de correspondances (XIX-XXème siècles)*. Centre Brestois du Greco 53 du C.N.R.S, 1986.
- GALLERON, Ioana. DEMONET, Marie-Luce. MEYNARD Cécile. IDMHAND Fatiha. PIERAZZO, Elena. WILLIAMS, Geoffrey. BUARD, Pierre-Yves. ROGERI, Julia. « Les publications numériques des corpus d'auteur ». Consortium Cahier, Huma-Num. 2015. En ligne : [https://f.hypotheses.org/wp-content/blogs.dir/1993/files/2018/12/guide\\_edition\\_EVENT\\_2018\\_1.pdf](https://f.hypotheses.org/wp-content/blogs.dir/1993/files/2018/12/guide_edition_EVENT_2018_1.pdf)
- IRIO(D'), Paolo. « Qu'est ce qu'une édition génétique numérique ? ». *Genesis*, n°30, 2010. En ligne : <https://journals.openedition.org/genesis/116#ftn7>
- JAKOBI, Marianne. « L'édition électronique des lettres d'artistes : le cas Van Gogh ». *Perspective : actualité en histoire de l'art*, 2 / 2011, pp. 807-813. En ligne : <https://journals.openedition.org/perspective/814>

### Sur l'archive et son appropriation culturelle :

- BARTHES, Roland. *Le degré zéro de l'écriture*. Paris : Seuil, 1953.
- BARTHES, Roland. *Le bruissement de la langue : Essais critiques IV*, Paris : Editions du Seuil, 1993.
- BERT, Jean-François. RATCLIFF, J. Marc. *Frontières d'archives : recherches, mémoires, savoirs*, Paris : Archives contemporaines, 2015.
- DIAZ, Brigitte. DIAZ, José-Luis. « Le siècle de l'intime ». *Itinéraires*, 4 / 2009, pp.117-146. En ligne : <https://journals.openedition.org/itineraires/1052>
- FOUCAULT Michel, *L'archéologie du savoir*, Paris : Éditions Gallimard, 1969.
- KOPP, Robert. « Du sacre de l'écrivain à la mort de l'auteur ». *Revue des Deux Mondes*, 2017. En ligne : <https://rddm.revuedesdeuxmondes.fr/archive/article.php?code=73694&show=picture&redirect=no>
- PASTORE Graziella, *Les coopérations entre chercheurs et bibliothécaires dans le cadre des projets de numérisation de corpus documentaires*, Enssib : Mémoire d'étude DCB, Mai 2017.

### Ouvrages sur les archives numériques :

- BACHIMONT, Bruno. *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris : Hermes Science Publications, 2007.

- BALLABRIGA, Michel (dir.). RASTIER, François. *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation*, Actes du colloque international d'Albi, juillet 2006. En ligne : <http://www.revue-texto.net/1996-2007/Parutions/Livres-E/Albi-2006/Sommaire.html>
- BERMES, Emmanuelle. MARTN, Frédéric. *Le concept de collection numérique*. Bulletin des Bibliothèques de France, Mai 2010. En ligne : <http://bbf.enssib.fr/consulter/bbf-2010-03-0013-002>
- COUTURE Carol, LAJEUNESSE Marcel. *L'archivistique à l'ère du numérique : les éléments fondamentaux de la discipline*. Québec : Presses de l'université du Québec, 2014.
- DELPIERRE, Nicolas. HIRAUX, François. MIRGUET, François. *Les chantiers du numérique : Dématérialisation des archives et métiers de l'archiviste*, Paris : L'Harmattan, 2012.
- ESSEVAZ-ROULET Baptiste, *La numérisation d'archives : des fondamentaux techniques aux programmes de numérisation*, Voiron : Territorial Éditions, 2010.
- GINOUVES, Véronique. GRAS, Isabelle. *La diffusion numérique des données en SHS : guide des bonnes pratiques éthiques et juridiques*. Aix-en-Provence : Presses Universitaires de Provence. 2018.
- HENRY, Clara. *L'archivage de la littérature numérique en ligne*. Enssib : Mémoire ARN, 2018. En ligne : <https://www.enssib.fr/bibliotheque-numerique/documents/68377-l-archivage-de-la-litterature-numerique-en-ligne.pdf>
- HOLZEM, Maryvonne. LABICHE, Jacques. *Dessillement numérique : énonciation, interprétation, connaissances*. Bruxelles : P.I.E-Peter Lang S.A., 2017.
- MAHE, Annaïg. Prime-CLAVERIE, Camille. « Le défi de l'interopérabilité entre plates-formes pour la construction de savoirs augmentés en sciences humaines et sociales ». ISTE éditions, 2017.  
En ligne : [https://archivesic.ccsd.cnrs.fr/sic\\_01511618/document](https://archivesic.ccsd.cnrs.fr/sic_01511618/document)
- PEDAUQUE T. Roger, *La redocumentarisation du monde*, Toulouse : Cepaduès Éditions. 2007.
- PEDAUQUE T. Roger, *Le document à l'ère du numérique*, Caen : C&f Éditions. 2006.

## Ouvrages et articles sur les patrimoines numériques :

- BACHIMONT, Bruno. *Arts et sciences du numérique : Ingénierie des connaissances et critique de la raison computationnelle*, Université de Technologie de Compiègne, 2004
- DOBEDEI, Vera. TARDY, Cécile. *Mémoire et nouveaux patrimoines*, Marseille : OpenEdition Press. 2015.
- DUPOND, Marie. FERJOUX, Céline. *Usages des patrimoines numérisés : un réseau interdisciplinaire et interinstitutionnel*, Usages des patrimoines numérisés, Paris : Université Sorbonne Paris Cité, 2017, 165 p. En ligne : [http://udpn.fr/IMG/pdf/rapport\\_de\\_synthe\\_se\\_udpn.pdf](http://udpn.fr/IMG/pdf/rapport_de_synthe_se_udpn.pdf)
- ROUSTAN Mélanie, *La recherche dans les institutions patrimoniales : sources matérielles et ressources numériques*. Presses de l'Enssib, 2016.

- SAJUS, Bertrand, Web 2.0 et mémoire : de la conversation à la conservation. *Documentaliste-Sciences de l'Information* [en ligne], 2009, vol. 46, n.1, p. 55

### Ouvrages et articles sur les Humanités Numériques :

- BOUHAI, Nasreddine. SZONIECKY, Samuel. *Intelligence collective et archives numériques*. Londres : ISTE Editions. 2017.
- CAVALIE, Etienne. CLAVERT, Frédéric. LEGENDRE, Olivier. MARTIN, Dana. *Expérimenter les humanités numériques : des outils individuels aux projets collectifs*. Montréal : Les presses de l'université de Montréal, 2017. En ligne : <https://books.openedition.org/pum/11091?lang=fr>
- FAURÉ, Christian. « Le gap sémantique et l'architecture orientée message ». Hypomnemata : supports de mémoire, mis en ligne le 5 octobre 2013. Accessible en ligne : <http://www.christian-faure.net/2013/10/05/le-gap-semantique-et-larchitecture-orientee-message/>
- KATHERINE-HAYLES Nancy. *Lire et penser en milieux numériques : attention, récits, technogénèse*. Grenoble : Éditions littéraires et linguistiques de l'université de Grenoble, 2016.
- LONGHI, Julien. « Humanités, numérique : des corpus au sens, du sens aux corpus ». *Questions de communication*, 1 / 2017, pp.7-17. En ligne : <https://www.cairn.info/revue-questions-de-communication-2017-1-page-7.htm>
- MASSOT, Marie-Laure. « Humanités numérique : Transformation des savoirs et des métiers ». *La vie de la recherche scientifique*, Sncs Fsu, 2014. En ligne : <https://hal.archives-ouvertes.fr/hal-01794171/document>
- MOUNIER, Pierre (ed). *Read/Write Book 2 : une introduction aux humanités numériques*. Marseille : OpenEdition Press, 2016.
- MOUNIER, Pierre. *Les humanités numériques*. Paris : Maison des Sciences de l'Homme, 2018. En ligne : <https://books.openedition.org/editionsmsmh/12039?lang=fr>

### Sur l'encodage numérique lié aux correspondances d'écrivains :

- DUMONT, Stephan. « Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF) », *Digiversity*, 29 octobre 2015. En ligne : <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif/>
- DUMONT, Stephan. « correspSearch – Connecting Scholarly Editions of Letters », *Journal of the Text Encoding Initiative*, Décembre 2016-Juillet 2019. En ligne : <https://journals.openedition.org/jtei/1742#tocto1n2>
- HANKINS, Gabriel, « Correspondence : Theory, Practice and Horizons », *Literary Studies in the Digital Age: An Evolving Anthology*, 2015. En ligne : <https://dlsanthology.mla.hcommons.org/correspondence-theory-practice-and-horizons/>
- Leriche, F., « Quel balisage pour les corpus épistolaires numériques ? De l'annotation traditionnelle du « document » à une analyse générique et pragmatique » dans Rastier, F. et Ballabriga, M., *Corpus en lettres et sciences sociales*, Paris, Texto, 2006; En ligne : <http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Sommaire.html>

## Études menées autour des correspondances d'écrivain (ouvrages et articles), colloques littéraires...

- C. OLDS., Marshall, « Flaubert Lacunaire », Europe : revue littéraire mensuelle, 2018, pp.149-162.
- GIRAUD, Frédérique. *Écrire pour résister au déclassé social : analyse sociologique de la carrière et des pratiques littéraires d'Émile Zola*. École Normale Supérieure de Lyon, 2014.
- GIRAUD, Frédérique. « S'imposer parmi ses pairs. Le travail réputationnel d'Émile Zola à l'assaut de la sphère restreinte du champ littéraire ». *Terrains&Travaux* N°26, 2015/1, pp.23-40. En ligne : <https://www.cairn.info/revue-terrains-et-travaux-2015-1-page-23.htm>
- MITTERAND, Henri. *Zola, tel qu'en lui-même*. Paris : Presses Universitaires de France. 2009.
- PRIVAT Jean-Marie, « La lettre et le panier. Une ethnocritique de Flaubert », *Ethnologie Française* (Vol.44), 2014, pp.651-661. En ligne : <https://www.cairn.info/revue-ethnologie-francaise-2014-4-page-651.htm#>
- SCHWEIGER, Amélie. « Flaubert en toutes lettres. L'écriture épistolaire dans la correspondance et dans l'œuvre ». Mont-Saint-Aignan : Presses universitaires de Rouen et du Havre, 2012. En ligne : <https://books.openedition.org/purh/5746>
- SZYLOWICZ Caroline, « Proust, Kolb, Kolb-Proust : emboîtement d'écritures autour de l'écriture », *Document numérique* (Vol.5), 2001, pp.135-154)
- SEGINGER, Gisèle, « Éléments pour une biocritique. ». *Flaubert : Revue critique et génétique*, Institut des textes manuscrits modernes (ITEM), 13 / 2015, pp.1-13. En ligne : <http://flaubert.revues.org/2439>.
- WALKER, John A. « Zola destinataire : 20 000 lettres à éditer ». *Cahiers naturalistes*, n°61. 1987. En ligne : <https://gallica.bnf.fr/ark:/12148/bpt6k9788419p>

## Volumes des correspondances de Gustave Flaubert et d'Émile Zola :

- ZOLA, Émile. *Correspondances*. Arvensa Editions, 2019.
- BRUNEAU, Jean (ed.), *Correspondance de Flaubert, tome I : Janvier 1830-Mai 1851*, Paris : Gallimard (Bibliothèque de la Pléiade), 1973.
- BRUNEAU, Jean (ed.), *Correspondance de Flaubert, tome II : Juillet 1851-Décembre 1858*, Paris : Gallimard (Bibliothèque de la Pléiade), 1980.
- BRUNEAU, Jean (ed.), *Correspondance de Flaubert, tome III : Janvier 1859 - Décembre 1868*, Paris : Gallimard (Bibliothèque de la Pléiade), 1991.
- BRUNEAU, Jean (ed.), *Correspondance de Flaubert, tome IV : Janvier 1869-Décembre 1875*, Paris : Gallimard (Bibliothèque de la Pléiade), 1998.
- BRUNEAU, Jean (ed.), *Correspondance de Flaubert, tome V : Janvier 1876-Mai 1880*, Paris : Gallimard (Bibliothèque de la Pléiade), 2007.

## Interviews

- LECLERC, Yvan. JUNGERMANN, Nathalie. « Gustave Flaubert. Correspondance », FloriLettres n°91, 2008, pp.1-5. En ligne : [fondationlaposte.com/article.php3?id\\_article=980](http://fondationlaposte.com/article.php3?id_article=980)

## Vidéos, conférences

- LECLERC, Yvan. « Éditer en ligne les manuscrits de Flaubert ». Université de Rouen. 29 septembre 2015. En ligne : <https://webtv.univ-rouen.fr/videos/04-editer-en-ligne-les-manuscrits-de-flaubert-yvan-leclerc-ceredi-2/>
- DRUCKER, Johanna. « Humanistic approaches to the graphical expression of interpretation », MIT Communications Forum, 20 mai 2010. En ligne : <https://techtv.mit.edu/videos/16677-humanistic-approaches-to-the-graphical-expression-of-interpretation-mit-communications-forum>
- DUPOND, Marie. « De l'édition à l'éditorialisation. Pour une approche historique des éditions de correspondances ». Université Lille 3. 22 novembre 2016.  
En ligne : <https://live3.univ-lille3.fr/video-etudes/dhnord-2016-de-ledition-a-leditorialisation-pour-une-approche-historique-des-editions-de-correspondances.html>
- MACKÉ, Jean-Sébastien. « Éditer un corpus numérique. Baudelaire, Nietzsche, Zola », Sorbonne Université, 10 mai 2019. En ligne : <http://www.item.ens.fr/Baudelaire-Nietzsche-Zola>

## Normes, standards, rapports sur les archives numériques...

- Correspondence Metadata Interchange Format (CMIF). Documentation  
En ligne : [https://correspsearch.net/index.xql?id=participate\\_cmi-format](https://correspsearch.net/index.xql?id=participate_cmi-format)
- ITPC. En ligne : <https://imatag.com/blog/2018/05/11/metadonnees-dimages-etat-lieux-inedit-inquietant/>
- OAI-PMH : [https://www.bnf.fr/sites/default/files/2019-02/Guide\\_oaipmh.pdf](https://www.bnf.fr/sites/default/files/2019-02/Guide_oaipmh.pdf)
- CNNum, Rapport “Ambition numérique : pour une politique française et européenne de la transition numérique”, 2015. En ligne : [https://www.renaissancenumerique.org/system/attach\\_files/files/000/000/136/original/Note\\_decryptage\\_Plateforme\\_RN\\_%281%29.pdf?1508510854](https://www.renaissancenumerique.org/system/attach_files/files/000/000/136/original/Note_decryptage_Plateforme_RN_%281%29.pdf?1508510854)

# Table des matières

<b>SIGLES ET ABRÉVIATIONS.....</b>	<b>9</b>
<b>INTRODUCTION.....</b>	<b>10</b>
<b>I. LES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE : QUELLES SPÉCIFICITÉS ET QUELS USAGES ?.....</b>	<b>13</b>
<b>I.4. L'épistolaire à l'âge de l'industrie et du romantisme.....</b>	<b>13</b>
I.1.1. La lettre au sein des transformations structurelles de la société.....	13
I.1.2. Introspection lettrée au XIX <sup>ème</sup> siècle.....	15
I.1.3. Une conservation recrudescence mais contrainte des manuscrits de lettres.....	17
<b>I.5. Les correspondances de Flaubert et de Zola.....</b>	<b>19</b>
I.2.1. Ce que nous disent ces écrivains et leurs destinataires, dans les correspondances.....	20
I.2.2. Vers une approche génétique des correspondances.....	24
I.2.3. La conservation des fonds de correspondances de Gustave Flaubert et d'Émile Zola.....	27
<b>I.6. L'appropriation des fonds de lettres de ces deux écrivains..</b>	<b>29</b>
I.3.1. Les éditions papiers des correspondances : leurs limites.....	29
I.3.2. La nécessité des collaborations archivistiques pour la contextualisation des correspondances : le cas Proust.....	33
I.3.3. La construction d'une archive épistolaire du XIX <sup>ème</sup> siècle.....	35
<b>II. LES ARCHIVES NUMÉRIQUES FACE AUX CORPUS DES CORRESPONDANCES DE GUSTAVE FLAUBERT ET D'ÉMILE ZOLA : ÉTUDE DE L'EXISTANT SUR LE WEB.....</b>	<b>39</b>
<b>II.1. Les premières approches des équipes scientifiques face aux     corpus de lettres numérisées.....</b>	<b>39</b>
II.1.1. Contexte de ces projets.....	39
II.1.2. La numérisation des manuscrits de lettres.....	42

II.1.3. Enjeux de l'« auctorialité » avec la transcription et l'annotation numérique des correspondances.....	44
<b>II.2. Indexation numérique des correspondances.....</b>	<b>49</b>
II.2.1. Description minimale du contenu des lettres numériques à l'aide des spécifications de Dublin Core.....	49
II.2.2. Perspectives d'un processus d'encodage numérique plus détaillé des correspondances d'écrivain.....	54
II.2.3. Le protocole TEI-XML à la lumière de la lettre d'écrivain à indexer : détails et développements.....	56
<b>II.3. Exploitation des corpus épistolaires de correspondances d'écrivain sur les plateformes du Web.....</b>	<b>60</b>
II.3.1. Écueils et perspectives d'une base de données épistolaire.....	60
II.3.2. Délimitation du corpus de lettres numériques et mise en accessibilité sur la plateforme du Web .....	63
II.3.3. Maintien de l'archivage numérique d'une correspondance d'écrivain : à quelles conditions ?.....	65
<b>III. PERSPECTIVES DE L'ARCHIVAGE NUMÉRIQUE DES CORRESPONDANCES D'ÉCRIVAIN DU XIX<sup>ÈME</sup> SIÈCLE SUR LES PLATEFORMES DU WEB.....</b>	<b>68</b>
<b>III.1. Extension des corpus épistolaires présents sur le Web aux Humanités Numériques.....</b>	<b>68</b>
III.1.1. Le projet Mapping the Republic of Letters et les fonctionnalités de représentations visuelle de corpus.....	68
III.1.2. Gestion de l'incertitude des données épistolaires numériques et perspectives humanistiques.....	70
III.1.3. Perspectives des Humanités Numériques pour les correspondances Gustave Flaubert et d'Émile Zola sur le Web.....	73
<b>III.2. Enjeux de l'ouverture des corpus épistolaires aux collections numériques des institutions patrimoniales sur le Web.....</b>	<b>75</b>
<b>III.3. Quels publics et quels usages pour l'archive épistolaire numérique sur le Web ?.....</b>	<b>79</b>
III.3.1. Médiation scientifique par le dépouillement des données numériques des lettres au sein de nouvelles plateformes du Web.....	80

III.3.2. Médiation pour les recherches en SHS à travers l'ouverture de la plateforme d'archivage épistolaire numérique des correspondances.....	82
<b>CONCLUSION.....</b>	<b>85</b>
<b>ANNEXE N°1.....</b>	<b>87</b>
<b>ANNEXE N°2.....</b>	<b>93</b>
<b>BIBLIOGRAPHIE.....</b>	<b>96</b>
<b>TABLE DES MATIÈRES.....</b>	<b>103</b>