Les études zoliennes, de l'archive numérisée à l'encodage TEI

Bilan et perspective sur dix ans d'expérience

Olivier LUMBROSO
Sorbonne-Nouvelle/ DILTEC ITEM-CNRS

Le *Centre d'étude sur Zola et le naturalisme* de l'ITEM a été fondé par Henri Mitterand en 1975 dans le but de publier, aux Éditions du CNRS et aux Presses de l'université de Montréal, la *Correspondance générale* de l'écrivain. L'entreprise s'est étalée de 1978 à 1995. Depuis ce projet d'envergure en onze volumes, le Centre Zola s'est investi dans des entreprises éditoriales collectives dont l'objectif a été de mieux faire connaître la pensée et l'œuvre du romancier, tout en s'adaptant à l'évolution des supports numériques et de leurs enjeux. Aujourd'hui, cette équipe est constituée d'un noyau permanent de cinq membres et d'une quarantaine d'associés, parmi lesquels une vingtaine collabore régulièrement aux initiatives participatives en cours depuis 2012[1].

Cet article vise à interroger les relations qu'entretiennent, historiquement et technologiquement, les humanités numériques dans le domaine de l'édition savante, et l'histoire littéraire du mouvement naturaliste, notamment celle de son chef de file, Émile Zola. La perspective pourrait être résumée au moyen d'une interrogation qui sera le fil rouge de la réflexion : comment les humanités numériques, saisies sur une période d'une dizaine d'années (2011-2021) ont-elles contribué à faire évoluer le discours de l'histoire littéraire du mouvement naturaliste ? Réciproquement, l'œuvre de Zola ne réclame-t-elle pas une réflexion continue sur les supports et les outils les mieux ajustés à la configuration de son œuvre, longue et massive, et aux finalités de sa lecture par

[1] Les membres « noyau » du Centre Zola sont : Alain Pagès (Sorbonne-Nouvelle), avec lequel Olivier Lumbroso (Sorbonne-Nouvelle) codirige le centre ; Jean-Sébastien Macke, ingénieur à l'ITEM-CNRS qui assure des rôles d'éditions scientifiques et de recherche ; Céline Grenaud-Tostain (Université d'Évry-Val d'Essonne) et Jean-Michel Pottier (Université de Reims-Champagne Ardenne).

des publics de plus en plus diversifiés ? En somme, ce travail souhaite se situer à la charnière de la théorie littéraire et de la réflexion sur les archives empiriques, de leur numérisation à leur encodage.

1 Du projet ARCHIZ (2012-2015) au projet CORREZ (2016-2019)

Pour donner, préalablement, une idée d'ensemble des projets numériques du Centre Zola sur la dernière décade, une présentation périodisée de ses livrables constitue une bonne approche. Avant 2011, il a existé deux projets informatiques : « l'iconothèque » et la mise en ligne sur Gallica, en 2003, du dossier préparatoire du *Rêve*, sous la direction de Philippe Hamon et Jean-Pierre Leduc-Adine, alors responsables du Centre[2]. À partir d'un projet iconographique inauguré dans les années 1990 par Leduc-Adine, une base, baptisée « Iconothèque Zola », a rassemblé une importante collection d'illustrations, de caricatures ou de photographies liées à l'œuvre de l'écrivain et à sa réception culturelle. Se présentant sous la forme d'un fichier informatique construit à partir du logiciel Filemaker, elle comportait, en 2011, mille-dix entrées et chaque image y était brièvement commentée. Plus tard, le site des « Archives zoliennes », ouvert en 2013 et dont l'acronyme est « Archiz », a repris les ressources de cette base[3]. Fruit d'une ANR coordonnée par Alain Pagès, le portail rassemble les principales archives d'un écrivain polygraphe, offrant ainsi au public une documentation à portée de main : les premiers romans et le cycle des *Rougon-Macquart*, les contes et les chroniques. De plus, l'internaute a accès aux manuscrits des romans et à ceux du « J'Accuse… ! », grâce à des liens directs vers Gallica[4].



Fig. 1 : Page d'accueil du site des Archives zoliennes, ©ArchiZ

Parmi les archives iconographiques, le lecteur découvre également une partie des photographies réalisées par l'écrivain vers la fin de sa vie et l'imagerie de la presse cari-

[2] Voir https://gallica.bnf.fr/dossiers/html/dossiers/Zola/, consulté le 28/07/ 2021.
[3] Voir http://www.archives-zoliennes.fr/, consulté le 28/07/ 2021.
[4] Voir par exemple le dossier du *Ventre de Paris* accessible sur ArchiZ à partir des types de documents génétiques qui le composent : http://www.archives-zoliennes.fr/manuscrits/les-rougon-macquart/le-ventre-de-paris/, consulté le 28/07/2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

caturiste d'époque qui en a fait l'une de ses cibles privilégiées. L'intérêt principal du site réside dans l'unité de lieu de la production auctoriale, qui concentre les archives romanesques, manuscrites et iconographiques d'une figure illustre du patrimoine et de la mémoire, un « grand homme », aussi, panthéonisé en 1908.

Ce premier projet numérique reflète une éthique et une poétique de l'histoire littéraire. Rappelons qu'il est né dans la période de lancement des sites patrimoniaux où les grandes monographies d'auteurs émergeaient sur le Web, en ce début des années 2010. S'observait alors un emballement pour tout numériser en vue de protéger les supports fragiles de l'usure du temps (Treleani, 2017). On envisageait aussi les retombées sociales de cette ère numérique, quand on n'y décelait pas le signe d'un élan démocratique et universel du savoir en libre accès. L'éthique numérique visait des formes nouvelles d'échange, de tolérance, d'éducation sociale et de réduction des inégalités d'accès à la formation. Sur ArchiZ, l'engagement de Zola dans l'Affaire Dreyfus en était un symbole (Lumbroso, Macke, Pagès, 2016). Le site mentionne dans l'un de ses onglets la bibliographie à destination des spécialistes et inclut le référencement des articles des *Cahiers naturalistes*, la revue littéraire du Centre, éditée chaque année par *La Société littéraire des Amis d'Émile Zola* (SLAEZ). Elle se consacre, depuis 1955, aux études sur l'œuvre de l'écrivain et plus largement sur le mouvement naturaliste en France et à l'étranger. Le site des *Cahiers* permet d'entrer dans l'histoire de cette revue et dans la vie intellectuelle de la société savante, notamment avec le Pèlerinage annuel de Médan qui reçoit de prestigieux orateurs[5]. Il est un complément au site Archiz, plus philologique et documentaire. Les numéros de la revue sont en ligne sur Gallica pour les livraisons allant de 1955 à 2010 :



Fig. 2 : Accès aux *Cahiers naturalistes* en ligne. ©BNF

Cette longue tradition littéraire et culturelle alimente un patrimoine intellectuel de premier plan, au sein de l'histoire littéraire du naturalisme, tant au niveau national (une revue structurée par des « dossiers thématiques »), qu'international, selon les méthodes de la littérature comparée.

Quelques grands principes d'histoire littéraire sont communs à ArchiZ et aux *Cahiers naturalistes.* Sur le portail, les objets culturels sont classés en rubriques (œuvres, iconothèque, manuscrits, bibliographie), en genres (les romans, les chroniques, les contes, les lettres inédites), en cycles (*Les Rougon-Macquart*, *Les Trois Villes*,…), en réceptions et jugements (le « triomphe » et le « déclin » du naturalisme, l'Affaire

[5]Voir le site à l'adresse suivante : http://www.cahiers-naturalistes.com/, consulté le 28/07/2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Dreyfus…). La revue se focalise sur des auteurs (Maupassant, Paul Alexis, Georges Darrien…) ou des thèmes (« Zola journaliste », « le roman colonial »…). Le modèle taxinomique de la bibliothèque générale, de la nomenclature, de la liste et de l'index, du dictionnaire et de l'anthologie de « morceaux choisis » actualise une perspective humaniste du mouvement littéraire mis en « tomes », comme les volumes des *Cahiers naturalistes*. Le mot « tome » vient de la racine indo-européenne « *tem », signifiant « couper », qu'on retrouve dans le « templum » fondateur du temple grec, dans le « temps » qui divise la durée, dans la « contemplation » qui admire la composition d'un paysage en le découpant mentalement. Même dans cette éditorialisation numérique totalisante, le modèle cognitif d'Archiz est fortement influencé par la logique du livre.

En 2016, la rencontre entre le *Labex TransferS* de l'ENS-Ulm et un fonds épistolaire zolien inédit, à savoir les lettres internationales de soutien adressées à Zola durant l'Affaire Dreyfus et conservées par le docteur Brigitte Émile-Zola, l'arrière-petite fille de l'écrivain, a généré une nouvelle approche. Ont été mis en valeur les circulations culturelles entre grandes aires géolinguistiques et le décentrement idéologique à l'égard de l'image patrimoniale de cet auteur, considéré par l'enseignement scolaire comme un classique de la littérature française et dont les archives en ligne ont conforté l'image. Or, les lettres autographes, envoyées de l'étranger durant l'affaire Dreyfus, ni savantes ni illustres, sauf cas exceptionnels, révélaient une myriade de visages de l'écrivain : les réceptions interculturelles de ce combat, à travers le prisme épistolaire de rédacteurs étrangers et ordinaires, à faible prestige[6]. Décentrement géographique, sociologique, discursif et générique pour capturer des représentations collectives et populaires, reflétant la diversité des accueils internationaux de l'Affaire Dreyfus et de l'œuvre de Zola. Les contrastes sont, de fait, significatifs entre l'ANR ArchiZ (2012) et le projet Correz (2016), sur les plans numériques et littéraires. Ainsi, on constate qu'autour de 2016, débute une déconcentration archivistique qui affaiblit le modèle radioconcentrique de l'Auteur et le principe d'unité de lieu de la conservation muséographique en ligne.

Du côté des humanités numériques, ces années sont aussi celles de l'épaississement du tissu institutionnel, de la densification d'une expertise qui a stimulé les pratiques mutualisées, normalisées et transférables et celles de la diffusion massive de nouveaux outils de création de bibliothèques numériques en ligne (comme le CMS Omeka[7]). En raison du rattachement scientifique et institutionnel du Centre Zola, les lettres de l'étranger ont également profité de l'entreprise éditoriale Correz (« Correspondances zoliennes »). Achevé pour une part, en cours de réalisation pour l'autre, le projet vise l'édition raisonnée des correspondances de Zola dites « active » (les 5000 lettres de l'édition en volumes) et « passive » : ces 2000 lettres de soutien, envoyées à Zola de-

[6] Pour une présentation du projet « Naturalisme du monde : les voix de l'étranger » qui inclut le traitement du corpus des lettres internationales dans le contexte de l'affaire Dreyfus, voir http://www.transfers.ens.fr/naturalismes-du-monde-les-voix-de-l-etranger#outil_sommaire_0, consulté le 28/07/2021.

[7] La plateforme d'édition de manuscrits modernes « EMAN » a été lancée à l'Institut des Textes et Manuscrits Modernes en 2016, sous l'impulsion de Fatiha Idmhand, Claire Riffard et Richard Walter. Développée avec le logiciel Omeka, elle permet de diffuser des archives accompagnées de champs de métadonnées à la fois annotatifs et descriptifs. Depuis 2017, cette plateforme est maintenue par Richard Walter, Ingénieur de recherche CNRS rattaché au laboratoire THALIM (UMR7172). Voir Idmhand *et al.*, 2017.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

puis les pays étrangers[8]. Une analyse contrastive des choix effectués pour le projet ArchiZ de 2012 d'un côté, et pour le projet Correz de 2016, de l'autre, aboutit aux remarques suivantes. « ArchiZ », géré par une petite équipe, est un portail panoramique de reproduction d'archives brutes, non-encodées, à métadonnées réduites, proposant un espace de documentation classé par genres ou par types d'archives, mettant à disposition aussi un moteur de recherche lexicale dans les romans des *Rougon-Macquart*. « Correz », dans le cas des lettres internationales, propose, avec la structure de la trame EMAN, un traitement systématique de la lettre originelle, sur les plans iconique et textuel : référencement, arborescence, métadonnées épistolaires, géographiques, historiques, avec un classement topographique par pays ouvert à des ajouts de lettres, bénéficiant enfin d'une géolocalisation de l'épistolier attractive lors de la consultation. L'édition a toutefois des limites : les lettres ne sont ni traduites ni transcrites ni annotées, sans analyse codicologique, sans mots clés attribués, limitant considérablement les recherches et les exploitations. La part instrumentale de consultation du document a été volontairement réduite :

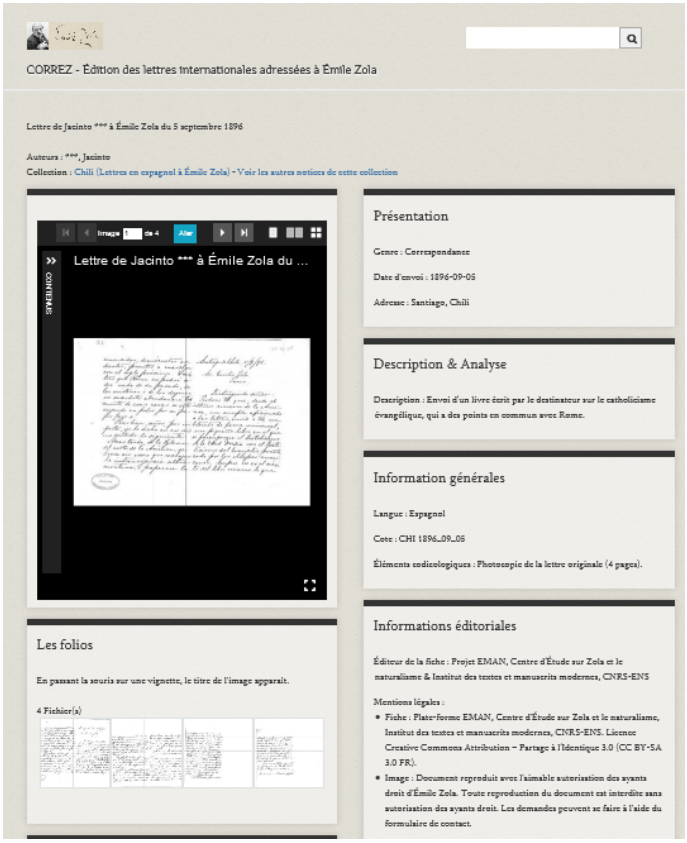


Fig. 3 : Édition numérique d'une lettre internationale ©EMAN

[8]Ce projet est coordonné par Jean-Sébastien Macke, ingénieur à l'ITEM-CNRS. Voir https://emanarchives.org/CorrespondanceZola/, consulté le 28/07/2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Sur le plan historique, Correz propose un modèle épistémologique littéraire distinct d'ArchiZ. Cultures élitiste/populaire, cultures d'érudition/de divertissement, livres à lire/à proscrire, genres majeurs/mineurs sont autant d'oppositions idéologiques manichéennes que la vitalité des pratiques d'écriture et de lecture a toujours transgressées. Mentionnons, par exemple, l'impact du feuilleton sur la création littéraire, qu'a développé Marie-Eve Thérenty avec le concept de « littérature-mosaïque », entre presse et littérature au XIX^{e} siècle (Thérenty, 2003). Un phénomène semblable se joue dans la perspective qui penserait ensemble l'histoire littéraire du naturalisme et l'édition épistolaire en ligne. Le bouleversement « révolutionnaire », au sens propre, auquel le numérique a contribué, porte sur la force inégalée d'un effritement de la normativité évaluative, souvent péjorative, qui a prévalu longtemps sur les lettres d'écrivains au regard de la textualité fictionnelle, les correspondances ayant été longtemps perçues comme document biographique secondaire voire anecdotique par rapport aux œuvres artistiques développant une esthétique du roman.

Or, la multiplication inattendue des éditions numériques épistolaires, chez Zola et chez beaucoup d'autres écrivains et artistes (par exemple la récente correspondance de Hugo et Juliette Drouet qui compte 13.232 lettres), n'est pas loin de perturber la hiérarchie texte / paratexte. Un autre élément d'innovation, à l'œuvre dans le domaine de l'histoire littéraire du naturalisme, est l'attention portée à la sociologie des pratiques scripturales, par exemple, aux écrits sociaux dans la presse d'époque, comme l'entretien ou l'interview, dont les études zoliennes s'emparent en ce moment dans une perspective d'édition numérique, en complétant l'ouvrage de Dolorès Signori et Dorothy Speirs, *Entretiens avec Zola* (1990). Ces pratiques médiatiques d'époque vont entrer dans un processus de valorisation patrimoniale et d'historicisation sociale de la vie intellectuelle et littéraire. Ainsi, l'histoire des œuvres ne se fait plus seulement par les textes et les contextes, les périodisations historiques et les jugements d'époque, les sources et les variantes, mais aussi par les pratiques ordinaires de lectures médiatisées qui passent par les discours révélateurs d'interactions culturelles comme les lettres, les interviews, les reportages réalisés chez l'écrivain. Il se révèle nécessaire d'interroger les liens de cette sociosphère avec la genèse des œuvres et leurs réceptions, notamment dans le cas de débats esthétiques en acte, entre acteurs du champ littéraire qui élaborent entre eux des positionnements (Prochasson, 1991).

Ce qui pourrait être en jeu dans ce retournement de la « pyramide » entre texte et paratexte, serait l'appréhension des faits littéraires comme processus plus horizontaux, soulignés par les hyperliens et les connexions diverses. Avec le modèle du livre, on renvoyait à une stabilité attachée à l'*opus* sans variation. Celle-ci est fragilisée aujourd'hui par la fluidité des traces numériques. Au sein de l'atlas zolien, elles nous obligent à nous demander ce que signifie éditer, écrire et lire les objets culturels naturalistes quand ils sont énoncés par un environnement technologique dynamique et structurellement processuel. Quel modèle théorique d'histoire littéraire peut prendre en charge cette représentation mobile des données empiriques qui bouleverse les cloisonnements, les genres et les ordres, sans prétendre à un nouvel ordre ou une nouvelle cohérence réticulaire *a priori*, compte tenu de la logique en archipels plus ou moins connectées et interopérables de cet espace numérique naturaliste labile, à géométrie variable ?

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

De façon complémentaire, relative à la valorisation de ces portails, quel modèle peut informer la didactique du naturalisme à destination des scolaires et des étudiants, lorsque de nouvelles formes d'apprentissage travaillent sur la simulation, le zapping, les voix collaboratives, la métamorphose de la figure du professeur en figure de passeur ? L'espace numérique est un environnement d'apprentissage souvent informel que les formations académiques, de l'école à l'université, interrogent peu. Inclure la culture exogène informelle dans la culture endogène académique est un défi institutionnel et un pari d'innovation, dont le numérique est l'une des chevilles ouvrières. De plus en plus, ces nouvelles formes d'autoapprentissage brouillent, à leur tour, les oppositions tranchées entre apprentissages académiques et informels, construisant des situations hybrides où les frontières apparaissent de plus en plus précaires.

2 Le projet ScéNa – Scénarios naturalistes (2019-2022)

Après une ouverture internationale de ses projets, le Centre Zola a souhaité revenir aux questions génétiques qui appartiennent aussi à son cœur de recherche. Le dernier projet présenté a voulu tenir compte des évolutions précédemment citées, tant socio-cognitives que technologiques, en imaginant une double réception, experte et novice, du document numérisé. Rappelons d'abord que les dossiers préparatoires du romancier sont toujours denses et qu'ils suivent la même organisation : un canevas général, deux « Plans détaillés », des « Fiches » personnages, des « Renseignements divers » issus de lectures ou d'enquêtes sur le terrain, une section de cartes et de dessins. Au sein de la phase de préparation, les « Ébauches » autographes des *Rougon-Macquart*, qui scénarisent le roman par une première « charpente », sont actuellement accessibles sur deux portails numériques. Elles ont été reproduites, dans leur matérialité originelle, sur le site Gallica au format PDF, et sont présentées sur ArchiZ selon une trame reprenant la composition du dossier de genèse[9]. Elles constituent, par ailleurs, le matériau d'un site de travail dans le cadre d'un projet, actuellement en cours, baptisé ScéNa (« Scénarios naturalistes »). Il vise l'encodage, au moyen de la TEI, des vingt Ébauches des *Rougon-Macquart*. Pour ce faire, l'équipe utilise la Plateforme de Transcription et d'Annotation de Corpus Textuels (TACT) développée par l'université Grenoble-Alpes[10]. Le projet a pour but de mettre en ligne les images des archives numérisées accompagnées de la transcription exhaustive des canevas scénaristiques. La plateforme TACT prend en charge, au moyen de balises présélectionnées, l'encodage en XML-TEI de façon à simplifier le travail du transcripteur. Dans un premier temps, la transcription enrichie a pour but de rendre ce corpus interrogeable selon des entrées canoniques, à destination des étudiants, enseignants et chercheurs qui travaillent sur les données romanesques : les « lieux », les « personnages », les « toponymes » des romans ou les « dates » historiques au fil des scénarios. Cette première couche de balisage permet d'appréhender le canevas romanesque selon différentes perspectives : l'onomastique, par exemple, permettra d'associer plusieurs désignations successives d'un même personnage, dès lors qu'elles auront été balisées, et de reconstituer, au moyen d'interrogations croisées de l'avant-texte, la genèse du système des personnages.

[9] Voir par exemple le schéma chronologique du dossier préparatoire du *Ventre de Paris* : http://www.archives-zoliennes.fr/manuscrits/les-rougon-macquart/le-ventre-de-paris/, consulté le 28/07/2021.

[10] Voir la plateforme TACT à l'adresse suivante : https://tact.demarre-shs.fr/project/20, consulté le 19 août 2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

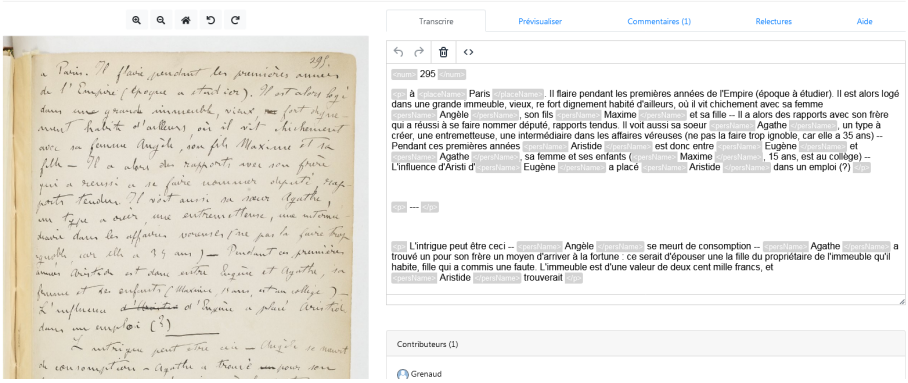


Fig. 4 : Interface TACT pour l'annotation des Ébauches de Zola. ©TACT

Qu'apporte ce système de balisage du point de vue de la recherche littéraire ? Si l'on se focalise sur la spécificité métatextuelle de ces scénarios, ce dispositif fait écho à un projet plus ancien dont il résout certaines apories. Entre 2005 et 2008, le Centre Zola, sous la houlette de Philippe Hamon, a construit le « Dictionnaire des dossiers préparatoires » de Zola, consultable sur le site des *Cahiers naturalistes*, selon un classement alphabétique de la terminologie méta-poétique et méta-énonciative de son soliloque dans les Ébauches et les plans détaillés[11] (Hamon, 2009). Si la démarche était heuristique, elle avait ses limites d'époque, de deux ordres : le principe de l'échantillonnage arbitraire (le chercheur recopie des extraits qu'il juge significatifs et décide subjectivement de la longueur des cotextes gauche et droit) et le principe lexicographique : les entrées du type « dictionnaire » privilégient le principe de la liste alphabétique d'entrées autonomisées. De fait, est peu apparent le réseau sémantique et discursif interne à l'écriture scénaristique, pouvant éclairer le « style génétique » (Herschberg-Pierrot, 2010) de Zola conçu comme tissage hétérogène. Avec le principe du dictionnaire, le document obtenu, qui constitue une bonne première photographie d'ensemble, reste une partition « trouée » qui interdit, en raison des béances entre échantillons, toute recherche textométrique, quantitative et qualitative d'ordre systémique, impliquant l'avant-texte dans sa spatialité textuelle continue et sa composition organique.

Ainsi, le mot « cadre », par exemple, dans des citations plus ou moins longues :

> Le Rêve (1888)
>
> Le cadre est très important. Je crois une ville de province dévote. Tout se passera autour d'une église romane (MS NAF 10323, *Le Rêve*, Ébauche, f° 233/17)
>
> […] Avec cela j'ai tout le cadre (*ibid.*, f° 234/18)
>
> Je prendrai le cadre d'une procession telle que j'en ai vu dans le midi (*ibid.*, premier plan du chapitre VIII, f° 111)

[11] Voir la présentation de ce *Dictionnaire*, dont le suivi a été réalisé par Céline Grenaud-Tostain, à l'adresse suivante : http://www.cahiers-naturalistes.com/wiki/doku.php, consulté le 28/07/2021.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

La TEI est susceptible de résoudre ces questions d'échantillonnage arbitraire, dans la mesure où le groupe de chercheurs se dote de perspectives, de choix et de protocoles communs. Sans qu'il vise, de façon illusoire, une vérité scientifique immuable, le cadre méthodologique, mieux explicité par une démarche réflexive et collective, dote les gestes des analystes d'une plus grande cohérence, d'une systématicité de la démarche comme d'une meilleure transparence à l'égard de la communauté scientifique, tout en réduisant la disparité des pratiques individuelles qui, s'additionnant les unes les autres, finissent par fausser considérablement les résultats définitifs de l'investigation.

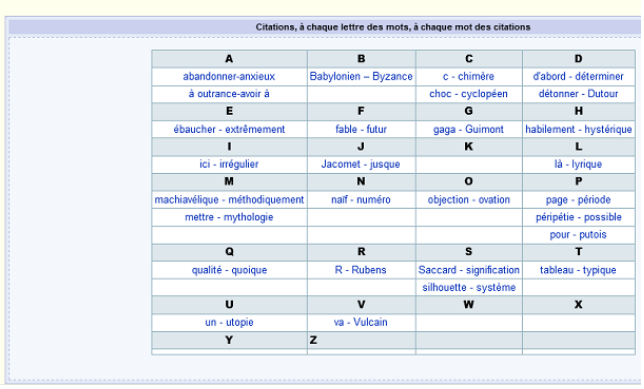
Comment cette démarche peut-elle concrètement s'actualiser dans le projet ? Première étape matérielle d'invention du scénario, l'Ébauche zolienne est singulière comme énonciation et contenus. Le projet d'encodage TEI visera à montrer la complexité du processus d'écriture scénaristique chez Zola avec une granularité plus fine. Les vingt Ébauches, de *La Fortune des Rougon* à *La Débâcle*, sont singulières : leur longueur, leur structuration, leur économie et leur stratégie diffèrent selon le projet de roman. Ainsi, vont-elles de quelques dizaines de pages à plus d'une centaine. Par ailleurs, elles s'illustrent comme des objets sémiotiques hybrides, à travers un hétéroclisme de matériaux (métaphores, injonctions, décisions, tâtonnements, bribes de dialogues). Afin de mieux cerner la nature de ces avant-textes, plusieurs propositions méthodologiques ont été formulées depuis qu'Henri Mitterand, lors du colloque *Zola. Genèse de l'œuvre*, en 1999, a ouvert le domaine de la « génétique du scénarique » (Mitterand, 1994, 2010) en soulignant ses défis et ses difficultés. En effet, les dynamiques de l'Ébauche rendent peu pertinentes les approches trop locales de l'écriture, attentives à la variante, à la rature, voire à la phrase (Gresillon, 2002), dans la mesure où la recherche scénaristique aborde les systèmes textuels globalisants qui forment des blocs et des masses aux contours flous : les variations d'une intrigue, les métamorphoses d'un personnage, les expansions et extensions du scénario, la nature des tensions narratives, la genèse des rythmes du roman, la valeur ironique du texte (Neefs, 1992 ; Hamon, 1997). Si une phrase possède un début et une fin, relevant ainsi d'un système « discret » d'unités isolables, à l'inverse, l'Ébauche est constituée de blocs de sens difficiles à isoler, dès lors qu'on les aborde comme des ensembles composés et non comme une liste d'échantillons qui, justement, propose un modèle « discret », au sens topologique, là où les phénomènes d'invention sont continus et diffus.

Sur la plateforme TACT, la transcription préalable du texte complet des Ébauches par une vingtaine de membres[12] garantit la continuité linéaire du discours programmatique, au-delà de l'extrait arbitrairement représentatif qui crée une disjonction entre la surface rédactionnelle du métatexte et les structures narratologiques profondes et dynamiques qui s'élaborent en continu et pour lesquelles nous avons besoin du matériau sémio-génétique complet. Comment l'encodage TEI pourrait-il alors compléter la logique du dictionnaire de 2005 ?

Au lieu de partir du modèle alphabétique à entrées uniques, tel que présenté sur le site du *Dictionnaire* :

[12]Le projet de transcription sur la plateforme TACT est coordonné par Jean-Sébastien Macke.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352



Citations, à chaque lettre des mots, à chaque mot des citations

A	B	C	D
abandonner-anxieux	Babylonien – Byzance	c - chimère	d'abord - déterminer
à outrance-avoir à		choc - cyclopéen	détonner - Dutour
E	**F**	**G**	**H**
ébaucher - extrêmement	fable - futur	gaga - Guimont	habilement - hystérique
I	**J**	**K**	**L**
ici - irrégulier	Jacomet - jusque		là - lyrique
M	**N**	**O**	**P**
machiavélique - méthodiquement	naïf - numéro	objection - ovation	page - période
mettre - mythologie			péripétie - possible
			pour - putois
Q	**R**	**S**	**T**
qualité - quoique	R - Rubens	Saccard - signification	tableau - typique
		silhouette - système	
U	**V**	**W**	**X**
un - utopie	va - Vulcain		
Y	**Z**		

Fig. 5 : Entrées alphabétiques du *Dictionnaire des autoconsignes des dossiers préparatoires*, ©ArchiZ

il serait possible de choisir comme balises des notions poétiques complexes, auxquelles se rattachent des réseaux de termes composant une terminologie méta-poétique désignant l'intrigue, le rythme, l'ouverture et la clôture de l'œuvre, à savoir des notions aidant à comprendre quels termes et familles de termes Zola convoque, croise, disjoint, oppose pour élaborer telle caractéristique de l'œuvre. C'est le sens du tableau n° 6 qui regroupe, selon leurs fonctions narratologiques, les termes désignant les opérations métatextuelles :

COMPOSITION DES INTRIGUES	DISTRIBUTION GENERALE DES MATERIAUX	DENSITE DES MATERIAUX	CONTINUITE, RYTHME ET GRADATION	MATIERES DE COMMENCEMENT ET DE TERMINAISON
Aboutir	Abandonner	Aborder	(s') Accroître	Commencement
Accommoder	Accompagnement	Abuser	Achever	Commencer
Accrocher	A cheval	Accumuler	Alternative	Conclure
Acheminement	Afficher	Abréger	Annoncer	Conclusion
Actif	Ailleurs	Ajouter	Au fur et à	Début
Action	A l'autre	Alinéa	mesure	Débuter
Agencement	Apparaître	Allonger	(Se) balancer	Fin
Agir	Apparition	Amonceler	Batterie	Final
Agissant	Arrière (en)	Bref	Coïncider	Initial
Amener	Auparavant	Brièvement	Conducteur	Introduction
Arriver à	Ça et là	Brièveté	Conduire	Introduire
Avancer	Caser	Casser	Conserver	Se mordre la queue
Bâtir	Cinquième	Cesser	Continu	Ouverture
Carcasse	Classer	Compact	Continuation	Ouvrir
Catastrophe	Découper	Complet	Continuel	Terminer
Charpente	Dernier	Compléter	Continuellement	
Composer	Deuxième	Complètement	Continuer	
Composition	Disposer	Contenir	Contrecoup	
Concourir	Disposition	Couper	Contrepartie	
Connexité	Distance	Court	Courir	
Conséquence	Distribution	Se débarrasser	Cours	
Crise	Distribuer	Déblayer	Cran	
Découler	Endroit	Déborder	Crescendo	
Déduire	Entre	Dédoubler	Croissant	
Dénouement	Enumération	Disjoindre	Croître	
Dénouer	Enumérer	Disparaître	Déchéance	
Dépendre	Espacer	Division	Définitif	
Détente	Fourrer	Egrener	Définitivement	
Détourner	Ici	Emplir	Dégradation	
Diversion	Inclusivement	Enlever	Degré	
Echappée	Intercaler	Entamer	Dégringolade	

Fig. 6 : Termes poétiques relatifs aux grandes entrées narratologiques du roman, ©Lumbroso

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

Selon quelles combinaisons lexicales et quelles entrées narratologiques (l'espace, le personnage, l'action…) Zola programme-t-il le scénario du roman ? Le TEI permet alors la construction d'un regard panoramique sur l'ensemble des scénarios afin de les appréhender aux moyens d'instruments compatibles avec une génétique longitudinale et perspectiviste qui étudierait les phénomènes de création sur l'étendue du cycle des *Rougon-Macquart*. Les termes du métalangage pré-rédactionnel des Ébauches constituent le socle des opérations poïétiques de la composition d'un roman naturaliste : le lexique renvoyant à l'intrigue, à la distribution des matériaux narratifs et descriptifs, les termes, si importants dans les récits de Zola nommant les principes de progression et de gradation, le lexique désignant enfin l'agencement du début de l'œuvre et de sa conclusion. Ce sont ces termes qu'il s'agirait d'encoder afin que ce marquage de la *dispositio* romanesque permette de les suivre tout au long des vingt Ébauches (Hamon, 2009).

Ce lexique métatextuel pourrait ainsi être appréhendé comme le tissu conjonctif de l'écriture scénaristique. L'encodage en XML TEI aura pour but de mettre en exergue des notions fédératrices organisées en faisceaux qui caractériseraient ainsi le style génétique zolien (« intrigue », « matériau », « scansion », « seuils »…) et permettraient d'appréhender les relations entre les balises qui pointent vers des opérations poïétiques générales. La recherche se donnera ainsi les moyens de comprendre les processus d'invention du scénario à l'état natif en termes de *concentration/éclatement* du drame (acmé, graduation, densité et gradation), *agrandissement/miniaturisation* des scènes, *marqueterie/assemblage* du chapitrage, *morcellement/continuité* des liaisons, *focalisation/diversion* des cadrages, des reliefs et des points de vue ; *tension/pause* dans la distribution des nœuds et des dénouements. Ainsi, ce que Flaubert appelait la « poétique insciente » de l'écrivain devient, pour le chercheur, davantage lisible et visible. L'exemple suivant, annoté à l'aide de marqueurs que nous appelons « poïétiques » et extrait de l'ébauche de *Son Excellence Eugène Rougon*, illustre l'intention du projet :

> La scène de la serre que je mettrai au commencement [*seuil*] du chapitre [*division*] suivant ne sera qu'une pointe aiguë [*focalisation*], amenant le dénouement[*dramatisation*] des amours et je pourrais donner ainsi plus de développement [*matériau*] au mariage de Clorinde et de Rougon que je coucherai ensemble[13].

Pour aller au-delà du classement alphabétique, cette édition numérique longitudinale devrait permettre l'enregistrement d'un métalangage en mouvement dans chaque scénario et à travers les vingt Ébauches appréhendées selon une perspective systémique neuve : apparition et disparition de termes, générativité de familles sémantiques d'Ébauche en Ébauche, augmentation et récessivité lexicales, environnement métadiscursif. L'image de ce métatexte scénaristique en ressort considérablement modifiée. Il ne puise plus dans le magasin d'un vocabulaire fini énumérable dans un répertoire, il envisage un métalangage ouvert, évolutif, selon les besoins de chaque projet et l'évolution du cycle (Lumbroso, 2013, 2017). L'encodage TEI montrera sans

[13] B.N., Ms, NAF 10292, f° 23.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352

doute plus finement les phénomènes corrélés de différentes natures, appréhendés par une génétique multidimensionnelle et réflexive au sens où « 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 » (Idmhand *et al.*, 2017).

3 Conclusion

Entre l'histoire littéraire du naturalisme et les humanités numériques, la rencontre est féconde, sans aucun doute. Elle a obligé les chercheurs de l'Équipe Zola de l'Institut des Textes Et Manuscrits modernes à s'interroger sur la plus-value sociale et scientifique des sites consultables en ligne et à tenir compte des nouvelles pluri-littératies médiatiques qui bouleversent les modes de lecture et les représentations des écrivains, lesquelles se diversifient dans l'espace social, se métissent, se teintent des détournements ou réactualisations modernistes dont se saisissent les internautes. Face aux quelques grandes tendances qui se dessinent sur la décade (le travail collaboratif, l'institutionnalisation de l'édition numérique et ses harmonisations, le passage d'une centralité d'auteur à une insularité des sites multilatéraux, liés par des ponts et des synergies qu'il reste encore à structurer), aller jusqu'au bout du numérique processuel supposerait de permettre la navigation entre un site patrimonial de correspondances, un portail des manuscrits encodés, de grandes ensembles d'archives qui se tissent entre eux, selon des liens de cohérences réticulaires et scientifiquement étayées, didactiquement exploitables. Il s'agit d'une dimension qualitative et rhizomique à très grande échelle, qui n'est pas si aisée de mener à son terme sans l'articuler à une vision renouvelée de l'histoire littéraire et, sûrement aussi, à des partenariats institutionnels avec, notamment, les ministères de l'Éducation Nationale et de la Recherche.

Published in : IDMHAND, Fatiha ; GALLERON, Ioana (dir.), *Dix ans de corpus d'auteurs*, Editions des archives contemporaines, ISBN : 9782813004352, doi : 10.17184/eac.9782813004352