Numériser et exploiter un corpus d'auteur

Deux cas pratiques en littérature

Marianne FROYE, France MARCHAL-NINOSQUE

Université de Bourgogne-Franche-Comté Elliadd (EA4661)

À partir de deux expériences empiriques de numérisation des manuscrits d'écrivains (Louis-Combet et Frénaud), nous proposons une réflexion sur différents paradigmes du travail du chercheur : l'expertise (pour l'établissement d'un corpus), le conseil (pour la plateforme numérique conçue par un ingénieur informaticien, en l'occurrence la plateforme numérique multimédias FANUM qui accueille des corpus littéraires et artistiques), l'interprétation (pour l'exploitation du corpus) et la nature du corpus littéraire devenu numérique, à la fois signifiant et matière.

Une fois numérisé, un corpus implique qu'il soit pensé comme tel : les choix de l'interface numérique lient l'iconicité et la plasticité d'un document, dont il faut bien admettre qu'il n'est déjà plus le document originel. Cet article s'intéressera d'une part aux choix graphiques, et à la capacité des liens hypertextes de transformer le discours que l'on peut construire à partir de l'objet originel, c'est-à-dire le manuscrit autographe d'un homme de lettres. D'autre part, il s'interrogera sur l'interface – dans le cas présent, une plateforme numérique de corpus littéraires –, et sur son influence sur l'exploitation traditionnelle des archives littéraires (génétique, éditions de texte, établissement de variantes…).

1 Présentation des projets : de la diversité des corpus vers une plateforme commune

Les deux expériences à partir desquelles nous nous proposons de réfléchir illustrent deux processus créatifs : d'un côté, des brouillons relativement propres pour le cas de Claude Louis-Combet et de l'autre, les manuscrits très raturés d'André Frénaud. Le but du projet scientifique est l'étude et l'interprétation des états intermédiaires de la création littéraire. La plateforme de visualisation de l'archive numérisée à destination

du plus grand nombre a pour but de donner à voir ces états et de faciliter le travail en génétique littéraire et l'exploitation des corpus.

1.1 Claude Louis-Combet

Claude Louis-Combet est un romancier, philosophe et essayiste encore vivant et en activité, qui réside en région Franche-Comté. C'est le romancier qui a remplacé Julien Gracq auprès de l'éditeur José Corti. Il a publié ses premiers romans chez Flammarion ou Lettres vives à la fin de la décennie 1970 et à partir de la décennie 1990, il est passé chez Corti. Toutefois, le mariage avec Corti autorise quelques infidélités : Claude Louis-Combet peut aussi publier dans des petites maisons d'édition locales, comme L'Atelier du Grand Tétras pour les livres d'artiste. Dans le cadre de son activité d'éditeur de textes spirituels anciens, il dirige la collection « Atopia » chez Jérôme Millon à Grenoble.

L'œuvre manuscrite (romans, essais, articles de nature universitaire, livres d'artiste…) de Claude Louis-Combet constitue un volume de plus de 40.000 folios autographes. Le travail de numérisation de ce fonds d'archives a comporté cinq principales étapes. Une première a concerné la collecte des manuscrits et brouillons de l'auteur ; en lien avec lui et avec son ayant-droit, des va-et-vient mensuels (le romancier vit à quelques kilomètres du laboratoire d'accueil[1]) ont facilité le projet de numérisation et de conservation de ce fonds d'archives à partir des années 1990. C'est le professeur Jacques Houriez qui a le premier souhaité constituer un tel fonds d'archives d'un romancier du XX^{e} siècle, d'abord à l'aide de photocopies dans les années 1990, puis dans le cadre d'un programme de numérisation à partir des années 2000, dont France Marchal-Ninosque a eu la charge.

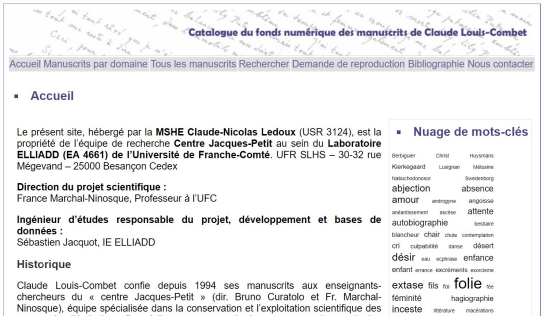


Fig. 1 : Page d'accueil du site Fanum – Claude Louis-Combet

La numérisation concerne presqu'exclusivement les photocopies des manuscrits autographes de ses œuvres publiées (romans, recueils de nouvelles, essais), agrémentés de ce que l'auteur appelle (fautivement, mais c'est une référence à Montaigne qu'il admire) ses *marginalia* : il s'agit, dans son cas, de tout ce qui n'est pas un état du texte à publier : des brouillons, notes de lecture, notes de régie, plans d'ouvrages ou de

[1] Ce laboratoire s'appelait à sa création dans les années 1960 « Le centre Jacques-Petit » (il est encore connu sous le nom de son fondateur, un des précurseurs de l'école génétique) ; puis il a été regroupé dans une plus grosse structure de 75 enseignants-chercheurs et une centaine de doctorants, ELLIADD (pour Éditions, Langages, Littératures, Informatique, Arts, Didactiques, Discours – EA 4661).

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

chapitres, parfois photocopies d'ouvrages sources. L'auteur étant encore en activité, il s'agit d'un cas assez atypique, celui d'un fonds d'archives numériques que l'on qualifiera d'*in vivo* car il continue à être alimenté régulièrement. Claude Louis-Combet et son ayant-droit Mireille Gerschwiler ont déposé les manuscrits originaux à la BNF depuis 2015 (sauf la correspondance).

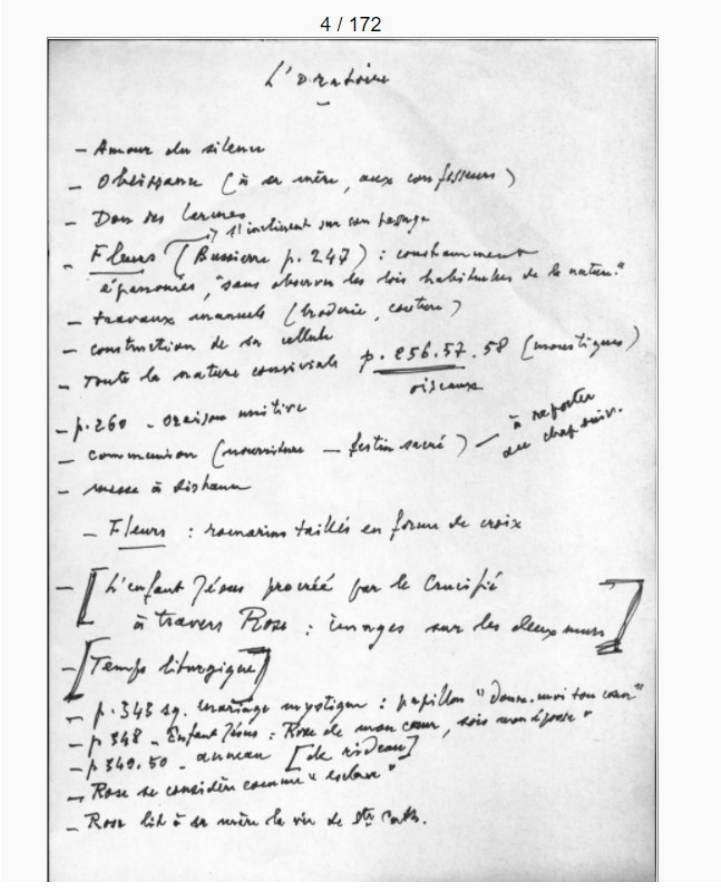


Fig. 2 : Exemple d'une page de manuscrit : Marginalia L'Âge de Rose, 4/172

La seconde grande étape de ce programme a consisté en la description des pièces de ce fonds. Des fiches signalétiques, conçues sur le modèle de la codicologie, ont été produites pour les 125 dossiers du fonds, chaque dossier ayant été conçu comme un « dossier génétique » de l'œuvre[2] puisqu'il comprend à la fois le manuscrit autographe et les *marginalia*. Chacun des 125 dossiers numériques forme désormais le catalogue en ligne de ce fonds d'archives. Il est décrit selon une fiche détaillée qui donne des informations sur le nombre de folios composant le manuscrit de l'œuvre, la nature

[2]Le dossier génétique est une production critique du chercheur qui rassemble les différentes pièces d'un même projet de création. Voir, notamment de Biasi, 2011.

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

des pochettes ou des cartons originels, la présence ou non d'éléments de datation[3], éventuellement le sommaire, des mots clés et des éléments descriptifs sur la nature du support (présence ou pas de ratures, qualité de l'encre…).

Une troisième étape a concerné le signalement, la communication de l'existence de ce fonds à la communauté internationale grâce, notamment, à la création d'une interface permettant de consulter ces archives. Le logiciel développé devait également servir aux autres fonds numériques du centre de recherche, comme celui de Claudel, ou de Lagarce, ou des fonds d'archives sonores ou visuels comme les chorégraphies de Dominique Bagouët (enregistrements faits par Jean-Paul Goux pour les « mémoires de l'enclave »). La plateforme Fanum (v. infra) a été constituée pour répondre aux besoins de ces différents fonds d'archives (sonores, textuelles, visuelles…).

Une quatrième étape a été juridique, consistant en la rédaction et la signature d'une convention avec l'auteur et son ayant-droit, permettant de mettre en ligne, et en accès libre, les archives sous leur forme numérique. Pour cette mise en ligne, nous avons privilégié les dossiers de *marginalia*, soit un dixième du fonds, pour 3363 folios[4].

La dernière étape est celle de l'exploitation de ces archives textuelles. Sont nés de ce programme de numérisation différents projets d'exploitation scientifique, tels que des colloques (voir Marchal-Ninosque et Poirier, 2012 ; Bouygues et Marchal-Ninosque, 2019 ; Bonord et Marchal-Ninosque, 2020) et des thèses doctorales (El Cheikh, 2012 ; Cabiron, 2021).

La disponibilité de ces manuscrits numérisés permet d'ouvrir d'autres possibilités, telles que l'encodage des autographes, destiné non pas à établir un texte, mais à identifier des phénomènes, à les spécifier à l'aide de balises en vue de reconstruire un ou plusieurs processus génétiques en traçant, par exemple, le destin d'un paragraphe, d'une idée, d'un personnage ou d'un lieu, depuis son apparition jusqu'à sa disparition.

1.2 Frénaud numérique [5]

Ce fonds s'est constitué en trois principales étapes. La première étape a consisté à reconstituer le fonds qui avait été désorganisé après le legs opéré par la veuve du poète en 1994.

L'intégration du projet au sein du consortium Cahier a permis d'importantes avancées dans le traitement du fonds. La numérisation exigeait des subventions peu conséquentes que les institutions ou les appels à projet peinaient à financer car la rentabilité scientifique à court terme ne leur semblait pas patente. L'aide de Cahier a permis d'avancer rapidement une fois la numérisation effectuée.

L'intérêt porté à l'œuvre de Frénaud tient en partie au fonctionnement du processus créatif. En effet, Frénaud mêle écriture de poèmes et exégèse sur les mêmes feuillets. Principalement poète, il écrit dans le même temps et sur le même feuillet son texte

[3] Ces éléments de datation sont toujours au crayon de papier, alors que les autographes sont à l'encre noire, Claude Louis-Combet voulant signifier que cet élément de datation ne fait pas partie de son texte.

[4] Par prudence, il a été décidé de ne pas mettre en ligne les manuscrits autographes des romans, beaucoup trop facilement lisibles, pour ne pas porter atteinte aux ventes des éditeurs. Les dossiers non soumis à l'accès libre sont consultables sur demande.

[5] Voir www.frenaudnumerique.fr, 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

poétique et le commentaire de ce dernier. Le corpus qui semblait alors le plus approprié pour affiner la connaissance de son geste d'écriture était composé de deux œuvres : *La Sorcière de Rome* et *Gloses à la Sorcière*. La sélection de ces deux œuvres a mêlé arguments scientifiques, pratiques, pragmatiques et génétiques. Notre volonté était d'étudier le processus de création de Frénaud. Nous souhaitions avoir en regard l'œuvre poétique et son exégèse, rédigée par le poète lui-même, pour mettre en évidence sa pratique d'écriture. L'ensemble le plus adéquat correspond à ces deux dossiers de brouillons en raison de leur complétude et de leur facilité d'accès.

L'exploitation numérique visée est de plusieurs ordres : elle a en tout premier lieu vocation à rendre plus accessibles les manuscrits de Frénaud en alimentant les recherches sur cet auteur. Frénaud est un poète majeur du XX^{e} siècle, mais sa présence est trop discrète dans les manuels d'histoire littéraire. Le travail sur le fonds de ce poète est une première approche d'une réflexion plus large sur l'ensemble de son œuvre et sur la poésie du XX^{e} siècle. Cette première campagne de numérisation amorce une réflexion sur la façon dont les référents culturels chrétiens et païens irriguent son œuvre. De façon plus large, le cas Frénaud initie une réflexion théorique sur la constitution d'un corpus à l'ère numérique ; il ouvre la voie à une plus vaste politique scientifique de numérisation et d'exploitation de fonds de poètes contemporains, dont le but est de mieux comprendre, à un niveau microtextuel, le rapport souvent tumultueux à l'alexandrin de toute une génération de poètes[6]. Les projets en Humanités Numériques portent souvent sur des textes en prose, tandis que les textes versifiés posent des questions plus complexes d'encodage ; pour réaliser notre étude, le projet débute par la conception d'un manuel d'annotation en XML/TEI, précisant l'utilisation des différentes balises à la disposition des encodeurs et par la création d'une interface de visualisation.

1.3 FANUM / FANA : Multiarch[7]

L'ingénieur du laboratoire ELLIADD, Claude-Nicolas Ledoux, a conçu la plateforme FANUM (pour Fonds d'Archives Numériques) pour les manuscrits de corpus littéraires. FANUM est destiné à fédérer les efforts d'acquisition et de mise en forme de ressources numériques entre l'Université et les institutions de conservation de patrimoine et d'archives (notamment régionales) et les éditeurs scientifiques. Il permet de mutualiser les compétences scientifiques et les moyens technologiques des laboratoires franc-comtois qui constituent, exploitent, valorisent des données textuelles et/ou audio-visuelles numérisées dont il s'agit de garantir la visibilité et l'accessibilité nationale et internationale. La plate-forme offre un outil de visualisation, mais permet aussi de télécharger les documents numérisés pour les exporter dans d'autres environnements.

FANUM rassemble sept fonds d'archives textuelles et audiovisuelles numérisées, depuis une vingtaine d'années, dans différentes disciplines (principalement en lettres, arts de la scène, histoire et sciences du langage). Son but est de favoriser leur conservation et leur exploitation dans le cadre d'activités de recherche. Il accueille ainsi,

[6] Ce projet vient d'être reconnu et soutenu financièrement par la région Franche-Comté et la MSHE Claude Nicolas Ledoux.

[7] FANUM : Fonds d'Archives Numériques, FANA : Fonds d'Archives Numériques Audiovisuelles.

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

en partie ou en totalité, les archives de l'écrivain Claude Louis-Combet, des danseurs et chorégraphes Dominique Bagouet et Ingeborg Liptay, des dramaturges Jean-Luc Lagarce et Gilles Laubert. Sont actifs dans cette plate-forme des littéraires, des textométriciens, des spécialistes de l'histoire de la danse entre autres, de l'Université de Bourgogne-Franche-Comté, mais également de Lyon et de l'Université de Lausanne. D'autres programmes portant sur des archives (textuelles, visuelles, sonores) numériques en ligne ont intégré cette plate-forme, notamment des programmes d'études de corpus (de contes cambodgiens, de correspondances des déportés de la Seconde Guerre mondiale en lien avec le musée bisontin mondialement connu de la résistance et de la déportation).

2 En amont et en aval : le *corpus* comme colonne vertébrale

Le premier bilan tiré de ces deux expériences de publications numériques de corpus littéraires en humanités numériques consacre l'importance du *corpus*. Si les compétences des ingénieurs informaticiens, des juristes ou des archivistes sont nécessaires à la mise en ligne d'archives littéraires, il revient au spécialiste de l'œuvre et de l'auteur de définir un corpus raisonné et réfléchi, exploitable.

Le chercheur, en amont du projet, possède l'expertise par sa connaissance du fonds à traiter, par sa capacité à déchiffrer l'écriture parfois illisible de l'auteur étudié, sa capacité à définir ce qui relève de l'œuvre, de l'avant-texte, et de l'étape du processus de création (une note de régie n'ayant pas la même importance dans la genèse de l'œuvre qu'un faux-départ). Son expertise sur l'histoire de sa discipline, des écoles critiques, et sur l'histoire littéraire est également essentielle à la constitution optimale d'un corpus d'étude. Enquêteur à la recherche d'indices dans les manuscrits, il reconstruit le cheminement de la création et devient l'épicentre des ressources qui sont éparses dans le fonds et qu'il a la capacité de repérer, comme autant de promesses de la recherche future. Cette étape de constitution du corpus, essentielle, reste encore finalement peu visible dans l'expertise scientifique des littéraires. Le chercheur en littérature offre donc une triple expertise : historique, physique (par sa connaissance de l'écriture de l'auteur et des matériaux utilisés) et génétique.

Un autre rôle essentiel du chercheur se situe en aval du projet, au moment de l'interprétation des données recueillies. En conjuguant ses pratiques de recherche à celles d'autres disciplines, le chercheur interprète le sens des occurrences lexicales, des modifications opérées par l'auteur ou encore des correspondances textométriques. Une telle enquête nécessite un encodage de ces données. Il lui revient alors d'être celui qui choisit quelle balise le texte nécessite : balise génétique quand le texte est un manuscrit ou un brouillon d'auteur, ou stylistique et sémantique quand le texte est un tapuscrit ou un imprimé.

Le chercheur est aussi l'expert qui peut proposer une interprétation historique du corpus ainsi constitué. La numérisation facilite l'étude des influences d'un auteur, met en évidence les phénomènes d'intra- et d'intertextualité[8]. Le chercheur qui étudie

[8] Dans les corpus numérisés de Claude Louis-Combet, par exemple, les prises de notes documentaires parfois copieuses du romancier permettent aux chercheurs de dresser la liste des textes qui furent ses ressources, voire ses sources d'inspiration. Ces folios de prises de note sont parfois accompagnés de notes

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

des manuscrits littéraires (les correspondances tout particulièrement) peut constituer plus facilement le réseau intellectuel explicite ou implicite de l'auteur étudié. Le passage au numérique apporte une facilité d'accès, mais également une nouvelle iconicité du document numérique (Meunier, 2017). Ces données permettent une interprétation stylistique de plus grande ampleur, grâce au balisage numérique du fonds. L'outil informatique permet une visualisation optimisée de différents manuscrits, à condition d'être très vigilant à la qualité de la numérisation (pour pouvoir en effet différencier la teneur, la couleur, la tenue des encres), ce qui n'est pas toujours le cas malheureusement (Bernard, 1999).

Cependant, il faut se retenir de verser dans un certain angélisme au sujet des apports du numérique. Alpha et omega, le chercheur en littérature maîtrise-t-il le reste de « l'alphabet numérique » ? Il peut parfois se mettre en retrait pour être davantage un consultant qu'un acteur lors de l'étape médiane où interviennent les développements informatiques. L'ingénieur est la cheville ouvrière de la conception de la plateforme de visualisation ; il est surtout le concepteur de l'interface, voire du logiciel qui met les données à la disposition du public. Dans sa création de l'interface, il opère des choix de navigation ou des choix visuels qui déterminent l'accès aux documents numériques. Pour le dire autrement, la version numérisée et la mise en ligne n'est déjà plus la version papier qu'on feuillette.

L'expertise de l'ingénieur pour créer un codage en vue d'une mise en ligne est déjà un acte qui intervient sur le fonds d'archives, qui le transforme dans son accès. L'archive numérique (qu'elle soit sonore, visuelle, textuelle…) devient à la fois un signifiant et une matière, et l'iconicité du signifiant qui change dans sa version numérique, risque d'être orientée par l'interface créée. Un logiciel, quel qu'il soit, transforme le rapport à cette nouvelle textualité. Car le document numérique n'est pas immatériel : il est une matière, qui vit sa vie propre, indépendamment du document origine. La moindre transformation de son iconicité (inévitable au moment de la mise en ligne) le transforme.

Les besoins des projets numériques dépassent d'ailleurs ces deux disciplines et appellent d'autres ressources, comme les spécialistes du droit. Consulter un manuscrit en bibliothèque nécessitait jusqu'à présent une autorisation des ayants droit, celle-ci est à repenser lorsqu'une vue numérisée du même document est publiée sur un site Internet. Aspect juridique qui continue à exister et même pourrait se durcir à l'avenir en raison des progrès faits en matière de propriété intellectuelle depuis une vingtaine d'années.

3 Une terminologie pour une méthodologie

L'apport du numérique au traitement des fonds d'archives d'auteurs implique aussi des changements de paradigmes. Les conséquences épistémologiques demandent à être éclaircies. Le numérique induit une redéfinition de la critique génétique, et probable-

de régie qui fixent les grandes lignes d'un plan, d'un chapitre, d'un motif ou d'une image qui va infuser le roman à venir.

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

ment de réfléchir à une nouvelle terminologie qui permettrait d'ordonner et d'organiser une méthodologie jusque-là abordée de manière inductive et empirique.

La critique génétique est impactée par le passage au numérique ; elle impose de réfléchir aux statuts des différents documents : le papier (manuscrit papier), la version numérisée (format jpeg ou IIIF), le fichier sur la plateforme, la transcription diplomatique, la transcription normalisée, transcription avec du contenu enrichi, les choix de balisage d'un manuscrit. Le numérique permet d'adjoindre à la recherche génétique *stricto sensu* d'autres métadonnées qui rendent l'œuvre interactive et interrogeable. Le format permet d'ouvrir d'autres pistes de recherche et d'allier plusieurs perspectives de manière simultanée et ne fige plus dans le temps l'interprétation d'une œuvre. Pour le projet Frénaud, dans l'immédiat, une première étape du travail concerne l'encodage des métadonnées des manuscrits. L'encodage de la transcription des textes viendra dans un second temps, et notamment l'encodage des métadonnées culturelles. Tous ces exemples confirment que l'archive numérique n'étant en aucun cas un *reprint* de l'archive papier ; elle est un document re-pensé, ré-ordonné, et donc d'une certaine manière ré-« orienté ». Ce sont deux documents qui restent fondamentalement différents et donnent des informations différentes. À ceci près que le format numérique est plus souple et permet une recherche *in progress*.

La constitution d'un corpus, comme nous l'avons évoqué à partir de deux exemples précis, impose des choix. Le chercheur qui constitue un corpus en vue de le numériser donne d'emblée une orientation scientifique à son projet. Le fruit de la numérisation est une lecture du chercheur de l'ensemble du fonds. En agissant ainsi, le chercheur offre à l'ensemble de la communauté scientifique des vues numérisées qui ne correspondent plus finalement à un fonds vierge de toute entreprise scientifique, comme celui auquel les chercheurs pourraient avoir accès lorsqu'ils se déplacent en bibliothèque pour découvrir des manuscrits.

Il convient donc de définir le statut de ces différentes archives physiques et / ou dématérialisées pour comprendre la perspective scientifique de la démarche du chercheur. En ce sens, la mise à disposition de données FAIRisées modifie en profondeur les rapports aux archives. Leur accessibilité et leur réutilisabilité sont deux apports indéniables qui invitent à ce que nous considérions le travail scientifique de la constitution d'un corpus. Par exemple, pour le cas des manuscrits Frénaud, sa veuve accorde ou pas l'accès et la reproduction des manuscrits papiers. Pour les documents numérisés, cette responsabilité nous échoit. C'est le laboratoire et les chercheurs qui définiront qui peut accéder aux vues numérisées. Plus le travail sur l'archive numérique est conséquent, plus émerge une nouvelle propriété intellectuelle : celle du chercheur, dont les droits seront gardés par une équipe ou un laboratoire. La propriété intellectuelle sur un fonds d'archives originelles est UNE (l'auteur ou son ayant droit après sa mort) ; elle est plurielle quand le fonds est numérisé (l'ingénieur et l'équipe avec qui il a travaillé).

Travailler sur des fonds littéraires de genres différents et sur des archives audiovisuelles montre toute la difficulté à trouver le vocabulaire adéquat pour décrire la réalité du document. Et ce n'est pas uniquement une question lexicale. Le vocabulaire employé est consubstantiel à la réalité décrite. Espérer concevoir une plateforme générique impose une grammaire commune. Or, les métalangages disciplinaires ne renvoient pas

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

aux mêmes réalités. Le Web sémantique est une piste que nous explorons pour parvenir à dépasser les idiolectes scientifiques. Plusieurs raisons en sont la cause, la première est sans aucun doute l'étape de l'encodage. Une même balise TEI peut être utilisée pour signifier des éléments différents selon que le document est sonore ou écrit. La seconde, celle de la visualisation : comment harmoniser l'affichage de différentes versions d'un document écrit et les captations d'un spectacle ? Les choix des balises d'encodage dépendent de cette grammaire ; l'arborescence du site et sa granularité également. Que définir comme « texte » ? Nous proposons en nous référant à son étymologie « ce qui en fait ordinairement le sujet, l'écriture sainte ». Que signifie « document » ? Selon la même approche, ce serait « ce qui sert à nous instruire ». La carte mentale suivante permet d'en cerner les enjeux.

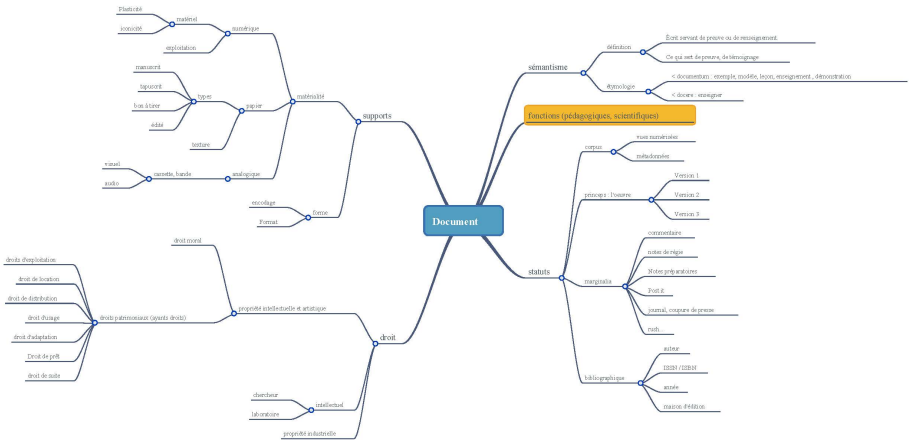


Fig. 3 : Carte mentale – Définition de « document »

Finalement, le changement de paradigme apporté par le numérique entraine la conception de nouveaux outils d'analyse, notamment ceux développés avec les chercheurs en informatique ou les ingénieurs de recherche. Le numérique automatise et fiabilise certains résultats ; le chercheur littéraire doit s'approprier de nouveaux paradigmes de recherche. Le décompte d'occurrences est certainement plus fiable lorsqu'une machine l'effectue que lorsque c'est un humain. La nécessaire sémantique de l'exploitation de textes numériques suppose que nous soyons auparavant d'accord sur le statut des différents états d'un document, sur sa terminologie. La terminologie pour l'exploitation du texte se raisonne encore bien souvent en spécialités scientifiques. L'émergence d'une école critique dépasserait les clivages disciplinaires.

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