
Au croisement des lettres, de la linguistique et de l'informatique : Les Manuscrits de Stendhal en ligne

Cécile Meynard et Thomas Lebarbé



Pour citer cet article

Cécile Meynard et Thomas Lebarbé, « Au croisement des lettres, de la linguistique et de l'informatique : Les Manuscrits de Stendhal en ligne », dans *Fabula-LhT*, n° 8, « Le Partage des disciplines », dir. Nathalie Kremer, Mai 2011, URL : <https://fabula.org/lht/8/lebarbe.html>, article mis en ligne le 04 Mai 2011, consulté le 24 Avril 2024, DOI : <http://doi.org/10.58282/lht.274>

Cécile Meynard et Thomas Lebarbé, « Au croisement des lettres, de la linguistique et de l'informatique : Les Manuscrits de Stendhal en ligne »

Résumé - Des chercheurs en lettres, informatique documentaire et linguistique, se sont rassemblés autour d'un même objet d'étude, le fonds de manuscrits de Stendhal, pour le traitement duquel ils ont conçu une base documentaire, CLELIA (en ligne sur www.manuscrits-de-stendhal.org). Ce fonds est perçu et nommé différemment selon les disciplines : patrimoine littéraire pour les premiers, ensemble documentaire pour les seconds et corpus linguistique pour les troisièmes. Mais il s'est avéré que la multiplicité des représentations rendues possibles par l'intervention et le dialogue de plusieurs disciplines a enrichi les regards portés sur cet objet

Cécile Meynard et Thomas Lebarbé, « »

Au croisement des lettres, de la linguistique et de l'informatique : Les Manuscrits de Stendhal en ligne

Cécile Meynard et Thomas Lebarbé

Préambule

Depuis quelques années, les campagnes de numérisation du patrimoine culturel, et en particulier littéraire, foisonnent, souvent avec l'appui financier du Ministère de la Culture et de la Communication, et peuvent amener à poser de façon à la fois originale et prégnante la question de l'interdisciplinarité. La volonté politique et scientifique est clairement affichée d'atteindre une masse critique afin de créer des synergies. Toutefois, le terme « numérisation » est à la fois très riche et très pauvre, recouvrant des réalités très diverses. Cette numérisation peut en effet être considérée comme l'« action de numériser¹ ; le verbe « numériser », étant lui-même défini comme « représenter par une suite de valeurs numériques », englobe dans la pratique différents aspects bien distincts et généralement successifs dans la procédure :

1. représenter un objet textuel ou iconique par une image numérique
2. représenter des propriétés de classification
3. représenter une transcription
4. enrichir d'informations contextuelles et explicatives

À partir de la deuxième étape, celle de la classification, souvent effectuée par les conservateurs des bibliothèques (nommage, indexation, catalogage...), la donnée numérique peut être valorisée à grande échelle et le patrimoine numérisé se voit diffusé plus aisément, sous forme d'édition papier, d'édition sur support de stockage (CD-ROM, DVD) ou en ligne. Pour prendre un exemple, la Bibliothèque municipale de Grenoble propose ainsi en ligne², par le biais de la société Arkhenum³, des fac-simile des pages de journaux de Stendhal (Cahiers dits « Bérès »)⁴ et des fac-simile de l'édition diplomatique de la *Vie de Henry Brulard* publiée par

1

2

3

4

Gérald et Yvonne Rannaud chez Klincksieck, associant ainsi les images de pages de manuscrits, et les fac-simile de leur transcription⁵.

Cette description, souvent effectuée par les conservateurs, permet un partage au sein d'agrégateurs plus généraux. Elle est généralement fondée sur des normes de type OAI (*Open Archive Initiative*) et permet d'accéder par un même point d'entrée à différents fonds patrimoniaux éparpillés géographiquement. L'on pourra ainsi citer le réseau thématique culturel de la Commission Européenne, Europeana⁶, qui joue un rôle de portail Internet vers de nombreuses collections culturelles européennes, manuscrites ou non. Il s'agit ici de constituer des Bibliothèques numériques accessibles au plus grand nombre de lecteurs.

Les troisièmes et quatrième étapes, quant à elles, sont plus coûteuses en moyens humains et matériels et sont par conséquent beaucoup moins fréquemment atteintes aujourd'hui, notamment parce que la priorité des autorités publiques s'est limitée pour le moment aux deux premières étapes afin de mettre en ligne suffisamment de manuscrits pour constituer d'importantes bibliothèques patrimoniales numériques, à des fins de préservation et de facilitation de la mise à disposition du public de fonds souvent précieux, fragiles, voire inaccessibles. Ces deux dernières étapes, qui correspondent à la transcription et à l'enrichissement des pages manuscrites par des spécialistes du domaine, sont pourtant fondamentales dans la mesure où elles permettent d'accéder aux contenus parfois peu lisibles des patrimoines manuscrits, à la recherche en texte plein, etc.

La préservation et la valorisation numériques du patrimoine littéraire sont ainsi des processus lourds, et qui sont souvent perçus comme destinés uniquement à des publics restreints de spécialistes de littérature, contrairement à ce qui se passe pour les fonds iconographiques (manuscrits médiévaux enluminés, collections de pastels, etc.) dont la valeur esthétique intrinsèque permet d'espérer toucher plus largement les utilisateurs. Toutefois, nous posons ici un premier postulat selon lequel un patrimoine littéraire peut être perçu comme un objet d'étude pour plusieurs disciplines, et comme un objet culturel intéressant et accessible pour divers publics. Nous posons un second postulat selon lequel l'objet d'étude que constitue le patrimoine littéraire peut être perçu comme une source d'interrogation scientifique non seulement pour la discipline littéraire mais aussi pour d'autres disciplines, notamment l'informatique et les sciences du langage. Dans cette perspective et autour de l'objet d'étude, les disciplines peuvent apporter leurs contributions respectives, mutualiser les efforts, les besoins, et les résultats, pour un enrichissement réciproque, et de l'objet lui-même.

5

6

Cet élargissement soulève toutefois la question de la notion même de discipline et des rapports pouvant exister entre disciplines. Quelle la préfixation adéquate pour l'adjectif qualifiant l'objet, l'approche, et le projet lui-même : transdisciplinaire ? pluridisciplinaire ? interdisciplinaire ?

Nous appuierons notre réponse à cette question problématique sur l'expérience des Manuscrits de Stendhal en ligne (projet conçu en partenariat avec la Bibliothèque municipale de Grenoble, qui fournit les images numériques des pages de manuscrits), où l'objet qu'est le fonds Stendhal a été pris comme objet d'étude par trois disciplines : les lettres bien entendu, mais aussi l'informatique en tant que discipline scientifique et non en tant que simple outillage, et la linguistique⁷.

Ce cas d'étude concret nous permettra de définir et analyser l'articulation, complexe et loin d'être évidente, des lettres avec d'autres disciplines des sciences humaines comme des sciences exactes. Il nous permettra par ailleurs de mettre en avant les interactions et surtout les influences de chaque discipline sur les autres en montrant que c'est l'objet d'étude qui peut générer l'interdisciplinarité et apporter, par le cas concret, des rebonds méthodologiques fondamentaux à chaque discipline.

Au début était le texte sur papier

Les réserves des bibliothèques, espaces tempérés et sécurisés où sont précieusement conservés les archives, les fonds anciens et précieux, fourmillent de pépites : manuscrits, incunables, ouvrages anciens ou qui ne sont plus édités, presse, etc. Quelques pages de manuscrits d'un auteur peuvent sembler une goutte d'eau dans cet océan méconnu. Devant une telle masse de manuscrits et d'ouvrages et faute de moyens suffisants, si l'on souhaite valoriser ces fonds patrimoniaux, il est nécessaire d'opérer une sélection pour en mettre une partie à disposition du public. L'estimation courante retenue par la BnF est de 1€ pour la numérisation en image d'une page, à compléter d'1€ pour la conservation à long terme de l'image. Le choix de numériser tel ou tel ensemble, que nous ne remettons pas en question, outre son aspect économique non négligeable, relève autant du choix de conservation (certains documents s'abîment et il est impératif de les préserver sous forme numérique pour les générations futures) que du choix d'édition, puisque la numérisation en image ouvre la porte de la reproduction numérique ou papier⁸.

7

8

Les manuscrits de Stendhal conservés à la Bibliothèque municipale de Grenoble qui sont l'objet de notre étude et le support de notre réflexion, représentent 20.000 feuillets⁹ (40.000 pages). Dans la ville natale de ce grand auteur français, alors que certaines pages se sont détériorées, suite à de trop nombreuses consultations, ou encore à l'emploi par Stendhal d'encre acides, ou de crayon, aujourd'hui presque illisible, et de papier de mauvaise qualité, un souci de conservation évident et légitime, tout autant qu'une volonté de valorisation, a présidé la démarche de numérisation de ce fonds. Ce dernier ne représente toutefois qu'un échantillon du patrimoine de la bibliothèque et y partage cette position privilégiée avec des incunables et autres somptueux manuscrits médiévaux, ou encore avec des manuscrits de Champollion et de Berlioz, dont la numérisation est également prévue à terme ou a déjà commencé.

À ce souci de valorisation culturelle s'est ajoutée une volonté de valorisation scientifique, par des chercheurs de l'Université de lettres, langues et communication Stendhal-Grenoble 3, centre symbolique des études stendhaliennes. Cette dernière n'est pas la seule à souhaiter exploiter la richesse d'un corpus d'écrits littéraires. Rabelais à Tours, Flaubert à Rouen, Montesquieu à Bordeaux, Proust à Paris, etc., il est assez aisé de trouver dans chaque grande ville des fonds de manuscrits d'un auteur célèbre et des chercheurs soucieux de le valoriser, avec plus ou moins de moyens – les lettres restant souvent le parent pauvre de la recherche.

À noter que les objectifs, les méthodologies et outils mis en œuvre, ainsi que les publics visés, varient – et font varier les plateformes numériques proposées. Ainsi, par les choix scientifiques et techniques des chercheurs qui l'ont conçue, la plateforme NietzscheSource¹⁰ s'adresse quasi exclusivement à des philosophes, voire uniquement à des spécialistes de Nietzsche. Inversement, l'atelier Bovary¹¹ à Rouen vise aussi bien un public d'élèves du secondaire et d'étudiants de lettres, que des spécialistes de Flaubert, proposant une interface simple, et des accès originaux (index, cartographies, base de données sur les comparaisons dans le roman, etc.).

Sans doute faut-il voir un progrès de la recherche littéraire vers une réelle interdisciplinarité, avec les projets d'éditions électroniques de fonds de manuscrits qui se développent actuellement (projet Flaubert, dans le cadre d'OPTIMA à l'ITEM¹², site des manuscrits de *Madame Bovary* (Université de Rouen¹³), projet d'édition électronique des dossiers de *Bouvard et Pécuchet*¹⁴, dossier génétique du *Rêve de*

9

10

11

12

13

14

Zola mis en ligne sur le site de la BnF¹⁵, etc.) pour ne prendre que des exemples de fonds du xix^e siècle). Littérature et informatique se mettent alors à dialoguer... mais bien souvent l'informatique apparaît plus comme un outil que comme une discipline à part entière et à égalité avec la littérature : en ce sens, le principe est moins celui d'un dialogue mutuellement enrichissant que celui d'un rapport client/prestataire de service.

En ce qui concerne le projet « Manuscrits de Stendhal », nous avons constitué une base documentaire, permettant désormais la mise à disposition en ligne, sur le site www.manuscrits-de-stendhal.org¹⁶, des images numériques de pages, ainsi que de leurs transcriptions et d'enrichissements divers. Il s'agit ici d'allier compétences littéraires, linguistiques et informatiques, en visant des publics de spécialistes littéraires et linguistes, mais aussi des élèves, étudiants et enseignants, et de simples amateurs de Stendhal. La diversité des publics impose par conséquent une diversité des représentations du fonds afin de répondre aux attentes et aux exigences de chacun.

Dépasser la terminologie et les représentations disciplinaires : du manuscrit à sa transcription

La question de la représentation est centrale à tout processus de numérisation, notamment du patrimoine, particulièrement du patrimoine manuscrit. L'objectif de proposer une diversité de représentations du fonds soulève le problème de la tâche de transcription : une transcription diplomatique, consistant en une transposition typographiée de la page, n'est pas la même qu'une transcription linéarisée, résultat du processus d'écriture d'une page où disparaissent les traces d'écriture (hésitations, ajouts, ratures, etc.), ni la même que le matériau langagier observable pour le linguiste.

Exiger des transcrip-teurs qu'ils produisent simultanément plusieurs transcriptions de la même page n'est pas humainement envisageable. Il est donc nécessaire d'inventer une méthode de transcription qui permette de convertir automatiquement une représentation vers les autres, sans perte ni déformation de l'information. C'est là qu'intervient non pas le simple outil informatique mais une

15

16

conceptualisation, une modélisation du document électronique, des interfaces homme-machine, afin de penser autrement la tâche de transcription.

Cette réflexion doit toutefois être menée au regard d'un certain nombre de paramètres non négligeables : en premier lieu, la prise en compte des compétences informatiques des utilisateurs, généralement habitués à des logiciels de traitement de texte traditionnels, et *a fortiori*, la nécessité de proposer un outil de transcription assistée par ordinateur facile d'utilisation et convivial ; en opposition à ce paramètre, celui de la faible capacité d'investissement dans le développement logiciel. Allier ces deux paramètres *a priori* contradictoires ne peut se faire que par une remise en question totale de l'approche informatique documentaire.

L'encodage numérique des textes n'est pas une question nouvelle. Toutefois, depuis une décennie, sous l'influence de la volonté affichée (et vendeuse) d'interopérabilité des systèmes, une norme fait surface et devient prépondérante : la *Text Encoding Initiative*. Projet ambitieux émanant des lettres, de la linguistique et de l'informatique, la *TEI* consiste principalement en une série de consignes de codage de l'information textuelle, selon une typologie complexe des documents. La dernière version inclut d'ailleurs une partie relative aux manuscrits. Toutefois, si la *TEI* est voulue comme un format d'échange et d'interopérabilité¹⁷, et à ce titre remplit cette mission à merveille, elle n'est pas conçue pour être facilement compréhensible pour les utilisateurs lambda. À l'image de tous les formalismes, il s'agit avant tout d'un langage univoque et interprétable par la machine comme des spécialistes du document électronique. Par ailleurs, si la *TEI* est accompagnée d'outils informatiques d'indexation et d'archivage, elle n'est pas accompagnée d'outils de transcription – chaque type de document, chaque utilisateur ayant ses propres caractéristiques et exigences de transcription. Nous avons préféré reléguer la *TEI* au rang de format d'échange et concevoir un format de travail, certes inspiré des préceptes de la *TEI*, mais adaptés aux besoins immédiats et à la littératie¹⁸ numérique d'une équipe sur un projet particulier (en l'occurrence les manuscrits de Stendhal). En alliant au sein d'un logiciel d'aide à l'édition de documents structurés¹⁹ :

- le potentiel d'XML,
 - une grammaire (DTD) constituée dans la terminologie et selon le point de vue des utilisateurs et
 - une feuille de style (CSS) donnant une représentation proche du rendu voulu pour une transcription linéarisé,
- nous avons été en mesure de mettre à disposition de l'équipe de chercheurs

17

18

19

littéraires un outil qu'ils ont pu s'approprier rapidement, tout en étant accompagnés dans la démarche. Nous nous sommes toutefois assurés que les données issues des transcriptions puissent être extractibles, transposables et transformables vers des formats d'échange (tels la *TEI*) et d'indexation (tels l'*OAI*).

Pour allier le besoin d'une facilité de transcription et la multiplicité des représentations, il est nécessaire d'identifier l'ensemble des informations à transcrire de manière à les retrouver dans chaque représentation où elles s'avèrent nécessaires ou utiles. La transcription ne se limite plus à une reproduction typographiée de la page mais à une description relativement exhaustive des contenus, de leur mise en forme, de leur fonctionnalité ou des rôles qu'ils jouent.

Afin de rendre la tâche abordable pour des utilisateurs néophytes, il est nécessaire de choisir les descripteurs de manière consensuelle (le terme est partagé par la communauté d'utilisateurs – transcrip-teurs), monosémique (le terme n'est pas ambigu) et définitoire (le terme est défini, contextualisé et si possible accompagné d'exemples pour éviter toute mésinterprétation). Cet exercice n'est pas aussi simple qu'il y paraît. Les arsenaux terminologiques des disciplines concernées n'ont pas une intersection nulle. Certains termes sont communs aux lettres, à la linguistique et à l'informatique, mais leurs définitions peuvent être divergentes. À titre d'exemple, le terme « corpus » en lettres correspond à un ensemble cohérent du point de vue intellectuel (unité de contenu) ou physique (unité de support, par exemple, mais aussi de scripteur, etc.), tandis que pour les linguistes, du moins de l'école sinclairienne de la linguistique de corpus, le terme correspond à un échantillon représentatif de la langue. À l'inverse, les désignés ne disposent pas toujours des mêmes désignants suivant la discipline. Ainsi, l'informatique aura tendance à utiliser le terme de « restructuration dynamique²⁰ » et les lettres celui de « reclassement virtuel » pour désigner le processus automatique qui sélectionne et réordonne des éléments de texte selon une demande de l'utilisateur. Si le terme utilisé par les lettres met en avant la réorganisation et l'automatisme, celui utilisé par les informaticiens met en avant la notion de structure documentaire et celle de dynamisme, c'est à dire de réponse aux requêtes des utilisateurs – la réorganisation n'est pas figée. Les termes sont concordants, désignent le même processus, mais ne mettent pas en avant les mêmes spécificités. La tâche de définition des descripteurs, mais de manière plus générale le dialogue entre les disciplines, est un exercice sémio-linguistique où chacun tente de s'approprier l'arsenal terminologique de l'autre tout en ayant une démarche pédagogique afin d'initier l'autre à sa propre terminologie. Il ne s'agit donc pas forcément d'aboutir à une terminologie commune pour caractériser tous les aspects de l'objet d'étude, mais de faire en sorte que les chercheurs des différentes disciplines impliquées soient

conscients quand ils emploient ou entendent tel ou tel terme, de l'ambiguïté possible dans l'interprétation de ce terme, et qu'ils en informent leur interlocuteur.

Un deuxième aspect important du travail interdisciplinaire est la définition et la dénomination des éléments pris en compte dans la description de l'objet, dans notre cas dans une page de manuscrit. Sur ce point, il a fallu que les chercheurs littéraires apprennent, par la découverte de la notion de « document structuré », à distinguer la description formelle de la description de fond. Pour prendre un exemple simple, avec un logiciel de traitement de texte, un utilisateur souhaitant créer ou identifier un « titre », peut se contenter de sélectionner une portion de texte, et de lui donner des caractéristiques visuelles (fonte, gras, centré, souligné, etc.) qu'il va associer à sa définition personnelle d'un titre. Mais dans ce cas, il s'agit d'un simple traitement formel, et en aucun cas le fragment de texte sélectionné n'est défini comme « titre » du point de vue de sa description : il reste identifié comme texte.

En l'occurrence, l'outil de description des manuscrits de Stendhal (adapté par Thomas Lebarbé à partir du logiciel Morphon) suppose que les chercheurs littéraires chargés de la transcription s'astreignent à respecter un protocole définitionnel mis en place par le dialogue littéraires/informaticiens : il ne s'agit surtout pas de mettre en forme un « titre » (c'est la machine qui s'en charge) mais d'identifier une portion de texte comme « titre », en le distinguant du « texte », mais aussi d'un « sous-titre », ou d'un « titre courant ». De la même façon, il a fallu établir les différences intellectuelles entre « foliotation » et « pagination », au lieu de se contenter de positionner des numéros à leur emplacement dans la page.

L'objectif d'intéresser plusieurs disciplines et de leur donner un objet d'étude correctement constitué impose une compréhension claire de la notion de document structuré comme arbre.

La transcription consiste ainsi à définir une série d'emboîtements, le nœud « page » contenant un nœud « descriptif » (les métadonnées contenant différents nœuds – « fichier image », « transcripateur », « cote-volume », etc. – correspondant à des informations générales sur cette page et sur le document auquel il appartient, ainsi que sur les conditions de sa transcription) et un nœud « contenu » (les données contenant elles-mêmes différents nœuds correspondant à la traduction en xml du texte rédigé sur la page par Stendhal ou éventuellement par un autre scripteur, – ce qui doit être mentionné également).

Pour prendre un exemple précis, voici la transcription XML d'une page de la copie « Arbelet »²¹ d'une partie du *Tour through Italy* de Stendhal (Arbelet-345.xml, correspondant au f° 19v du cahier n° 50) :

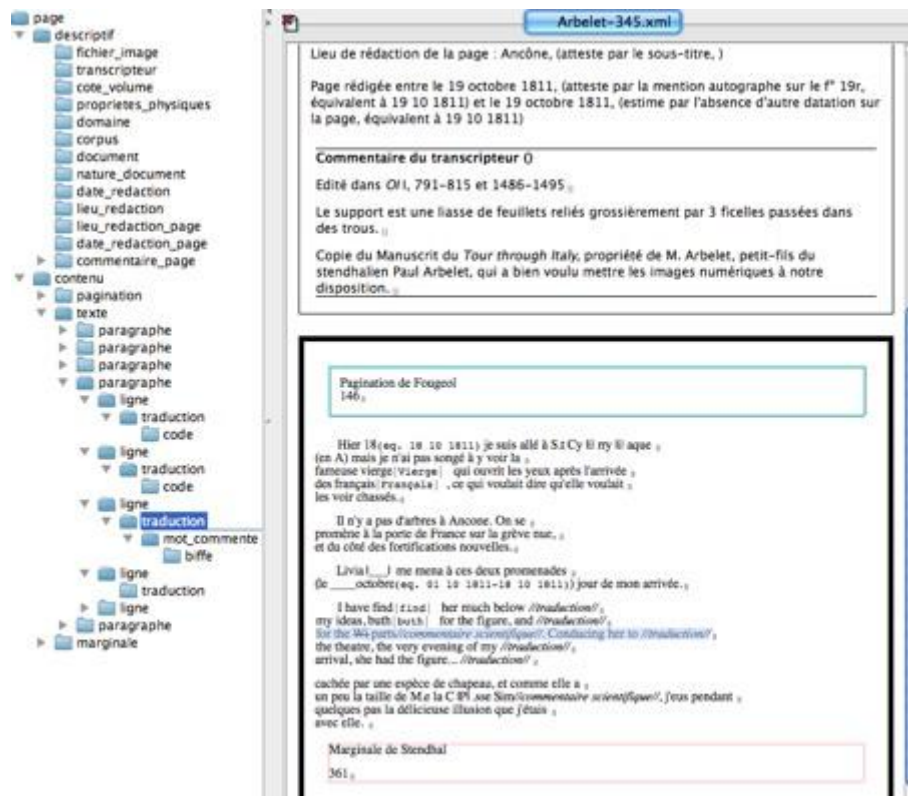


Figure 1 : Capture d'écran d'une fiche xml (détail) de transcription d'une page de manuscrit (copie Arbelet, f° 19v).

Cette capture d'écran présente en bas à droite la transcription du texte, avec une ligne surlignée en bleu, correspondant à l'élément « traduction », également surligné dans l'arborescence à gauche : contrairement à ce que ferait peut-être spontanément un littéraire en travaillant avec un traitement de texte (à savoir indiquer sa traduction en français par une note à la fin du passage en langue étrangère), il faut ici qu'il désigne l'ensemble du passage en langue étrangère (qu'il le balise, pour utiliser le terme exact), ce qui permettra ensuite des recherches sur les locutions et termes étrangers utilisés par Stendhal, et sur son usage du sabir, qui pourront intéresser aussi bien de simples curieux (qui s'amuseront de la maîtrise approximative des langues étrangères par l'auteur) que des spécialistes littéraires. Dans sa jeunesse, Stendhal associe ainsi souvent la langue anglaise à la mention de l'argent qu'il doit à ses amis, ou que ces derniers lui doivent : pudeur, façon de mettre à distance l'évocation de cet objet ignoble mais vital ? quant aux linguistes, ils pourront s'intéresser au contexte immédiat du passage en langue étrangère, pour déterminer par exemple s'il existe des termes ou des contextes linguistiques introducteurs de sabir, qu'ils désignent plutôt par alternance codique (ou *code-switching*).

On le voit, aucun flou définitionnel n'est permis par ce système, qui peut sembler contraignant mais est la condition même d'une description fiable permettant

ensuite des regroupements automatisés d'informations significatives et rigoureusement établies.

La démarche de transcription numérique – outre la tâche non négligeable de déchiffrement du manuscrit – se résume (mais ne se réduit pas) à délimiter, identifier (nommer) et caractériser (donner des propriétés) les éléments de la page. Cette structuration de l'information, cette « mise en fond » peut ensuite être mise en forme selon des critères variant d'une représentation à une autre. Plus encore, les éléments peuvent être sélectionnés et réordonnés à l'envi. Cette approche du document structuré n'est pas novatrice en informatique – les traitements de texte tout comme les éditeurs de page web offrent les moyens de travailler sur un document structuré, mais force est de constater dans les documents électroniques produits par les non-spécialistes que ce potentiel est largement ignoré. La démarche, même si elle est adoptée par certains littéraires, reste encore marginale dans ce domaine, notamment parce que contrainte par des normes et des outils construits avant tout pour être en adéquation avec les besoins informatiques et non en adéquation avec la littératie numérique des utilisateurs.

Notre approche a tenté de pallier ce défaut et s'est centrée sur la structuration et la terminologie co-construites entre utilisateurs et informaticiens. Alliée à une représentation (en l'occurrence la transcription pseudo-diplomatique) au sein d'un outil d'aide à l'édition de documents structurés s'appuyant sur la DTD (la grammaire) et une CSS (la feuille de style assurant la mise en forme), elle devient un outil compréhensif pour le non informaticien, lui donnant en temps réel une approximation du résultat visuel (*WYSIWYnG*, *What you see is what you nearly get*) tout en le guidant dans sa tâche de transcription dans les termes qui lui sont familiers.

Recourir à différents champs disciplinaires pour donner les manuscrits à voir, à lire et à comprendre

L'enjeu, défini dès le début du projet en 2006²², est de donner les manuscrits de Stendhal non seulement à voir (ce que fait déjà en partie la Bibliothèque municipale de Grenoble sur son site), mais aussi à lire (rendre lisibles des pages complexes du fait de l'écriture ou des différentes strates d'écriture qui se succèdent voire s'empilent) et surtout à comprendre. Ce dernier objectif – donner à comprendre – est celui qui repose le plus évidemment sur l'interdisciplinarité.

Tout d'abord, l'utilisateur peut voir les manuscrits sous forme d'images numériques, c'est-à-dire feuilleter virtuellement des objets physiques, dans leur matérialité (couleur du papier, tracés...) et leur esthétique (voir figure 2); il peut aussi consulter des ensembles constitués automatiquement ou manuellement par les chercheurs à partir des informations qui ont été insérées dans les descriptions, tels que toutes les pages comportant des dessins de Stendhal, ou toutes les pages de titre (voir figure 3). L'intérêt est ici essentiellement visuel, mais peut prêter à des études sur les pratiques scripturales de Stendhal, par exemple le relevé de mises en scène récurrentes au début d'un nouveau cahier (mise en scène du titre, de la date, de commentaires en marge en haut, à gauche ou à droite du titre, etc.).

L'utilisateur peut ensuite choisir de lire les transcriptions des pages en regard des images correspondantes, sous deux formes : linéarisée (voir figure 4), c'est-à-dire sous une forme lisse, proche d'une édition imprimée (les ratures n'étant pas affichées, les ajouts étant intégrés dans les lignes, etc.) et pseudo-diplomatique (voir figure 5), c'est-à-dire respectant autant que possible la mise en page stendhalienne, les fautes d'orthographe, en affichant les mots biffés, les ajouts à leur place, etc.

Enfin, les manuscrits sont aussi et surtout donnés à comprendre grâce aux annotations descriptives, qualificatives et explicatives qui viennent enrichir les transcriptions, mais aussi grâce aux possibilités de recherche dans le texte. Les linguistes ont par exemple identifié dans la base de données la présence récurrente du syntagme verbal « je sors » associé à des références à des pièces de théâtre et à un contexte lexical du dénigrement (les pièces sont médiocres, les acteurs jouent mal, etc.). Le dialogue avec les littéraires a permis de contextualiser cette remarque : ce syntagme figure en effet surtout dans les pages des journaux de jeunesse (1804-1805), quand le jeune Beyle veut se donner une formation de dramaturge et fréquente activement les théâtres de la capitale, allant voir deux pièces dans la même soirée, comme c'est la tradition à l'époque. Mais il est particulièrement intéressant pour les littéraires de bénéficier d'une approche systématique qui vient enrichir leurs connaissances : la fréquentation assidue des théâtres s'accompagnant de remarques majoritairement négatives de la part du jeune homme, elle ne permet pas réellement la construction de modèles à suivre. Et la construction mentale du futur dramaturge se fait surtout sur le mode négatif : « ne pas faire comme... ». Les modèles de dramaturges anciens (Shakespeare, Molière...) ne sont plus que rarement joués, donc très peu évoqués après « je sors » ; inversement, les pièces modernes servent souvent de contre-modèles. La base de données permettra ainsi d'appuyer des intuitions littéraires sur des données scientifiques mais aussi d'émettre de nouvelles hypothèses²³.

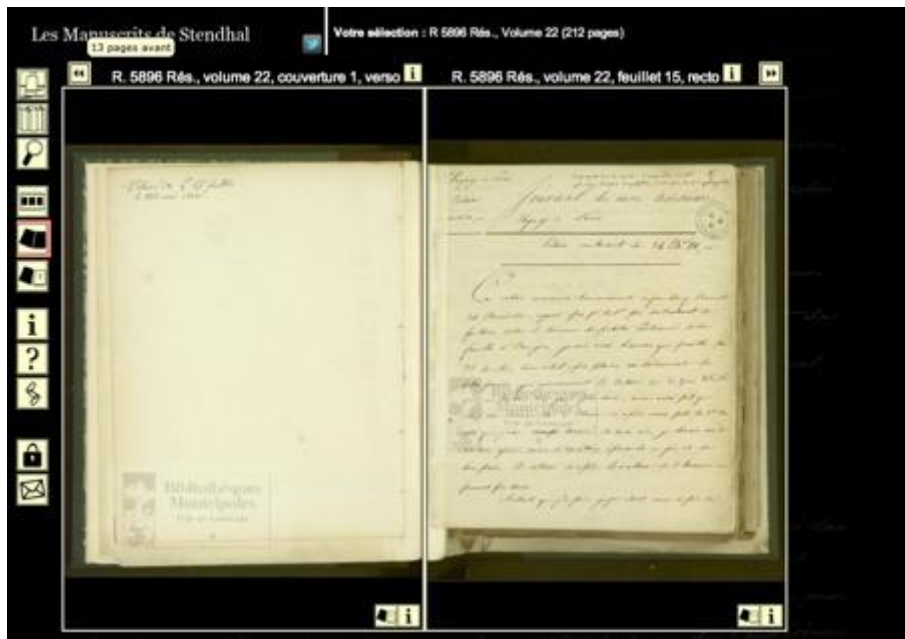


Figure 2 : Représentation en vis-à-vis d'une double page (verso-recto) du R. 5896 Rés., volume 22. Copie d'écran du site <http://manuscrits-de-stendhal.org>²⁴.

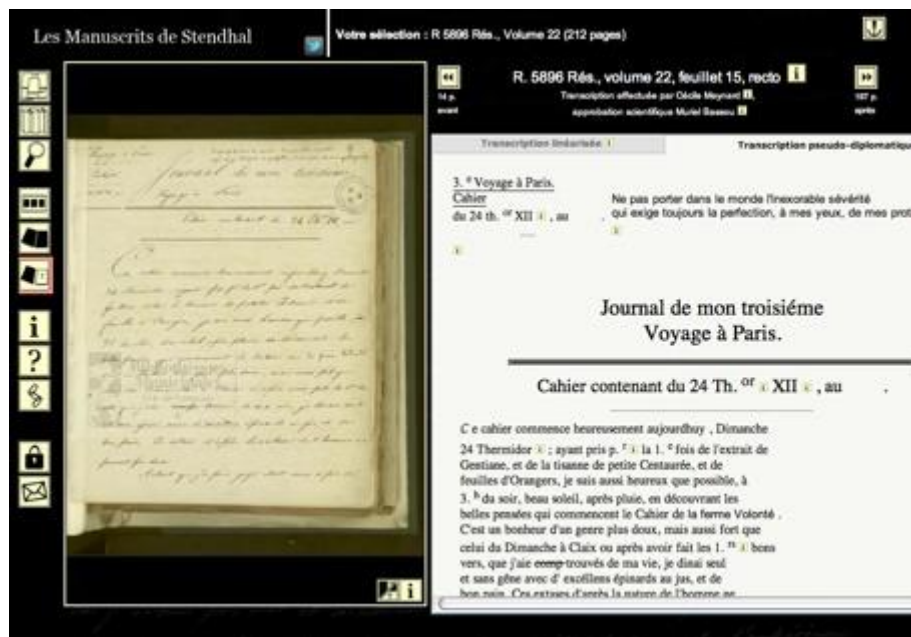
Figure 3 : Représentation en vignettes de 10 des 212 pages transcrites du R. 5986 Rés., volume 22.



Figure 4 : Représentation du feuillet R. 5896 Rés., volume 22, feuillet 15, recto, et sa transcription linéarisée.



Figure 5 : Représentation du feuillet R. 5896 Rés., volume 22, feuillet 15, recto, et sa transcription pseudo-diplomatique.



Par ailleurs d'autres types d'enrichissements, à destination de publics plus larges et moins initiés, ont été mis en place, tels que des commentaires audio et textuels, des images, ou encore une chronologie et un QCM...

Il s'agit ainsi de donner accès aux manuscrits de Stendhal de diverses manières, grâce aux apports de différents champs disciplinaires : par l'objet (on peut, comme nous l'avons expliqué plus haut, visualiser les images des pages de manuscrits, en regard ou non de leurs transcriptions), par mots-clés, mais aussi par regroupements

thématiques, ou par dossiers pédagogiques. Et d'ici deux ans, il sera possible d'interroger la base par descripteurs (dates, scripteurs intervenus sur la page, dimensions du papier, outils d'écriture utilisés), selon un principe de recherche avancée (en croisant les critères). La conception de cet outil d'interrogation fait actuellement l'objet d'une thèse en Traitement automatique des langues²⁵ qui amène la doctorante à s'appuyer sur des concepts relevant aussi bien de l'informatique que de la linguistique.

La question qui se pose dans ces conditions est de savoir si ce projet amène les chercheurs de l'équipe à travailler dans la trans- ; la pluri- ou l'interdisciplinarité.

Trans-, Pluri- ou Interdisciplinarité ?²⁶

Il n'existe pas, semble-t-il, de distinction clarifiée entre les préfixes « inter », « pluri », « multi » et « trans ». Nous souhaitons poser une distinction, du moins entre transdisciplinarité, pluridisciplinarité et interdisciplinarité. En revenant aux fondements étymologiques :

- *trans-* : à travers
- *pluri-* : plusieurs
- *inter-* : entre

Nous pouvons poser intuitivement des définitions plus concises des trois termes :

- *transdisciplinaire* : qui concerne le déplacement à travers les frontières de deux disciplines²⁷.
- *pluridisciplinaire* : qui concerne plusieurs disciplines distinctes.
- *interdisciplinaire* : qui concerne l'intersection entre des disciplines.

Plus précisément et de manière schématique, nous proposons les définitions plus élaborées et la représentation correspondante (voir figure suivante) :

- La *transdisciplinarité* consiste à transférer, transposer, transmettre (flèche) tel quel un principe, une méthode, une technologie d'une discipline A (foncé) à une discipline B (clair), à travers la frontière (diagonale) entre ces deux disciplines, pour le seul intérêt de la discipline B et de son objet d'étude (cercle).
- La *pluridisciplinarité* consiste à disposer d'un objet d'étude (cercle) commun à deux ou plus(ieurs) disciplines, chacune ayant son regard disciplinaire sur l'objet.
- L'*interdisciplinarité* consiste à disposer d'un objet d'étude commun et échanger principes, méthodes, technologies, compétences, connaissances autour de cet objet. Cet échange (que les Anglais nomment *interchange*, soulignant ainsi l'idée de

25

26

27

transmission réciproque) impose un dialogue entre les disciplines, la constitution d'un terrain commun. Par conséquent la frontière disciplinaire devient floue, il n'existe plus de « territoire » disciplinaire, il est remplacé par un continuum interdisciplinaire.

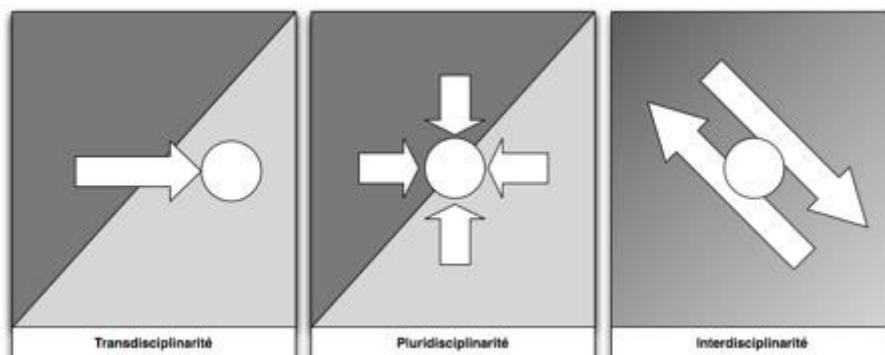


Figure 6 : Schématisation symbolique des articulations trans-, pluri- et interdisciplinaires.

Persuadés de l'intérêt scientifique de véritables dialogues et interactions entre disciplines, nous avons fait le choix pour le projet « Manuscrits de Stendhal » d'une démarche « interdisciplinaire », tout en postulant qu'il ne s'agit évidemment pas de nier l'intérêt et la force des approches disciplinaires. Notons d'ailleurs que le projet des manuscrits de Stendhal en ligne aurait pu se cantonner à un projet transdisciplinaire (transmission d'un savoir-faire technologique afin de mettre les manuscrits en ligne), ou à un projet pluridisciplinaire (appropriation des manuscrits en tant qu'objet d'étude linguistique indépendamment de leur appropriation littéraire et génétique).

Thomas Samuel Kuhn définit une discipline comme « un ensemble de connaissances et de compétences construites et standardisées, par un groupe de personnes ayant des intérêts/objectifs communs, en fonction d'un paradigme, pour répondre à des questionnements »²⁸. Ces connaissances et compétences varient bien entendu d'une discipline à une autre en fonction d'intérêts et d'objectifs différents. Frank Alvarez-Pereyre, quant à lui, décrit la discipline sous un angle légèrement différent, de l'ordre de la description sociologique, voire ethnologique : « Chaque discipline correspond à un certain regard sur les objets – une langue, un ensemble de rituels, une entreprise, un savoir technique²⁹ ». En d'autres termes, quand il s'agit de faire dialoguer les disciplines entre elles, il faut tenir compte de ces paramètres :

- une langue : outil de communication entre les spécialistes, elle peut être source d'incompréhensions quand la terminologie n'est pas définie, ou source de

28

29

quiproquos quand la terminologie est commune mais les acceptions n'ont pas été désambiguïsées.

- des rituels : ils sont eux aussi inhérents à chaque discipline. Ils transparaissent dans les méthodes de travail, les modalités de communication scientifique, la reconnaissance des travaux, etc. L'absence de reconnaissance de ceux-ci et surtout de leurs différences peut s'avérer aussi contreproductive que l'absence de clarification terminologique, notamment dans le milieu universitaire.

- l'entreprise : elle est étroitement liée aux rituels, ceux-ci régissant les desseins propres à chaque discipline. Ainsi, si l'édition numérique d'un patrimoine commence tout juste à être reconnue au sein de la communauté des lettres – le papier ayant encore une valeur hypostasiée, l'édition électronique d'un corpus³⁰ est reconnue à sa juste valeur de contribution scientifique parmi les linguistes. Les deux entreprises présentent pourtant des similitudes d'objectifs et de méthodes, mais aussi des divergences qui doivent être assimilées par chacune des parties prenantes d'un tel projet interdisciplinaire.

- un savoir technique : il s'agit non seulement d'un savoir mais d'un savoir-faire qu'il est nécessaire de partager, d'explicitier. Le faire sous forme de dialogue lui confère une dimension pédagogique indéniable, nécessaire à la bonne appréhension du domaine pour tous les chercheurs concernés.

Ainsi, pour concevoir l'interdisciplinarité, le cloisonnement disciplinaire doit être transgressé, mais ne peut l'être qu'à la condition que le dialogue s'instaure et que chaque discipline soit à même d'appréhender la langue, les rituels et le savoir technique de l'autre.

Le principe du dialogue entre lettres, linguistique et informatique qui constitue la base même d'un tel projet implique, comme les exemples ci-dessus visent à le démontrer, une constante gymnastique intellectuelle dans le but de se comprendre entre spécialistes de disciplines différentes, qui peuvent par là découvrir certes des points de convergence insoupçonnés, mais également et surtout des zones, non pas forcément de divergence, mais de méconnaissance liées à des conceptions et des approches différentes de l'objet. Ce sont essentiellement ces zones situées aux marges de toutes les disciplines concernées qui à la fois posent problème, suscitent le débat et peuvent même devenir de nouveaux centres d'intérêt. Il en va ainsi par exemple de la zone correspondant à la définition problématique de la notion de « document », qui s'est avérée différer selon les littéraires, les informaticiens-linguistes et les documentalistes de l'équipe. Les disciplines concernées vont ainsi pouvoir interagir et s'enrichir mutuellement.

Interactions interdisciplinaires

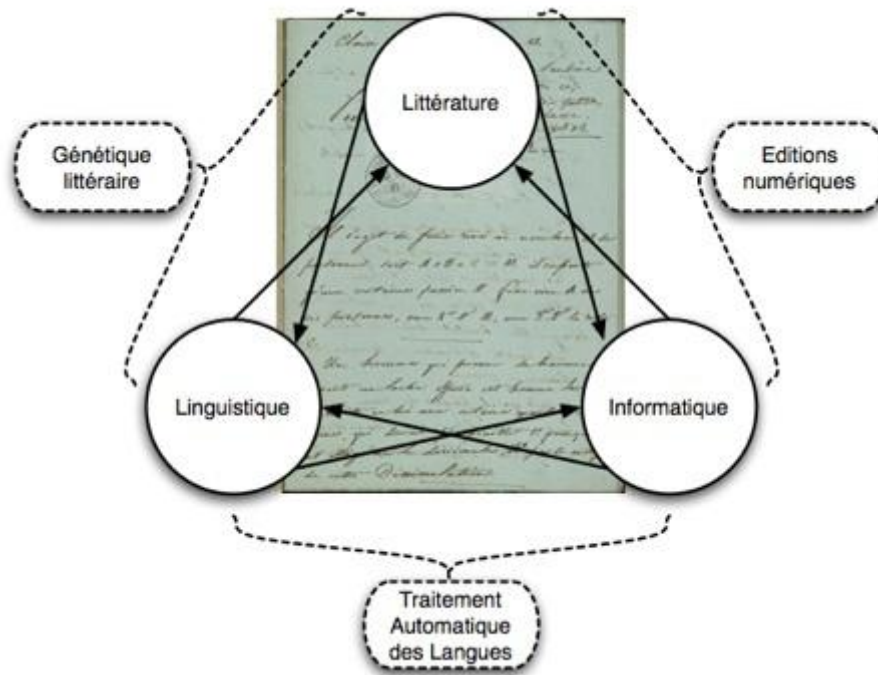


Figure 7 : Articulation interdisciplinaire littérature, linguistique et informatique autour de l'objet manuscrit.

Notre travail est fondé sur le principe d'un cycle empirique incrémental, ou pour dire les choses plus humblement et avec humour (ce dernier étant d'ailleurs sans doute l'un des ingrédients nécessaires de toute entreprise interdisciplinaire par sa capacité à dédramatiser tensions et malentendus) de « l'esclavage réciproque » : à l'issue d'une conceptualisation commune des grandes lignes de l'objet d'étude, les informaticiens conçoivent un modèle de représentation des manuscrits, et le proposent aux littéraires qui, en utilisant l'outil de transcription proposé et le site sur lequel est mise en ligne cette édition électronique inédite, vont avoir des pratiques spécifiques, et laisser des traces pour les linguistes qui vont pouvoir définir des usages, analyses qui servent aux informaticiens pour proposer de nouveaux outils informatiques plus adaptés aux besoins des transcrip-teurs et utilisateurs du site. Représenter ce cycle (c'est à dire présenter à nouveau, mais en y portant un regard analytique et critique) n'est pas simple et nous le résumons dans le schéma ci-dessous, qui permet de visualiser les différentes étapes de travail de collaboration (flèches horizontales) et d'esclavage (flèches en arcs) qui parfois résultent en des cycles empiriques incrémentaux (boucles rouges).

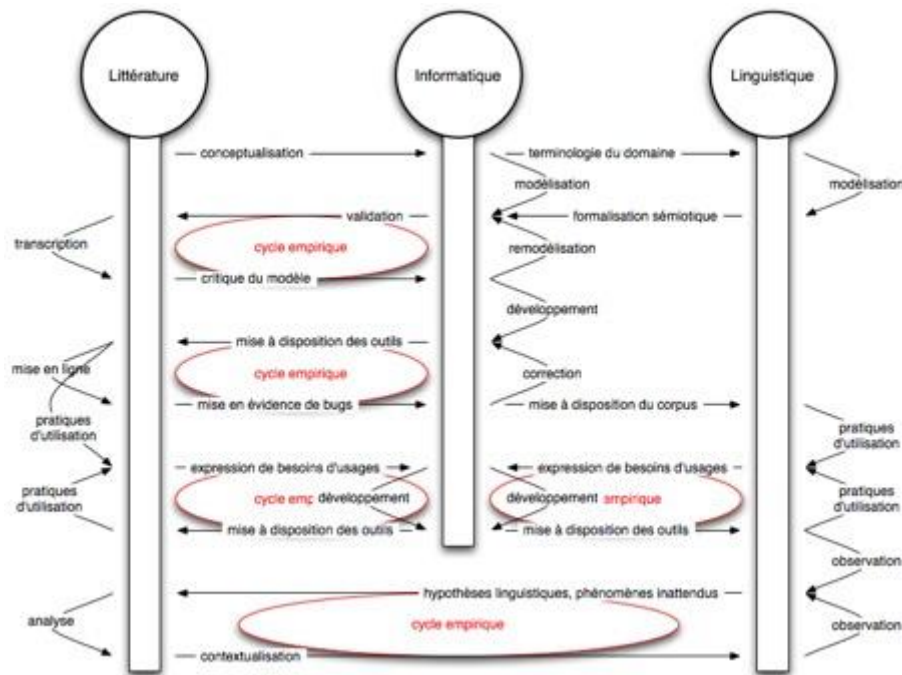


Figure 8 : Cycle(s) empirique(s) jalonnant l'élaboration du projet des Manuscrits de Stendhal.

Un autre principe fondamental du projet est celui des concessions mutuelles que doivent se faire les représentants des trois disciplines. Comme le note Franck Alvarez-Pereyre, « la relation ou l'articulation – des données et des disciplines – sont le lieu où apparaissent au plus haut point les tensions qui naissent de cette contradiction constitutive entre la diversité des lieux d'où chacun parle et la cohérence comme effort permanent³¹ ». Ces concessions peuvent être considérées comme constructivistes dans la mesure où elles permettent à chaque discipline, lorsqu'elles sont réciproques, de faire un pas vers l'autre, à défaut de quoi chacune se cantonnera dans ses propres représentations et s'isolera de toute approche constructive et bénéfique.

En construisant ensemble un nouveau modèle, tous acceptent la remise en question de leurs modes de pensée, de leurs modèles et de leurs certitudes. En effet, pour arriver à la constitution d'un « terrain commun »³², il est nécessaire de faire preuve de souplesse intellectuelle, de compétences de pédagogue, bref de s'aventurer peu ou prou en terrain inconnu : chacun se trouve à un moment ou à un autre dans la situation délicate d'employer le vocabulaire, de recourir aux notions de l'autre, ou de devoir inversement expliquer, voire justifier des notions qui lui semblent évidentes ou qu'il n'a jamais discutées. Par ailleurs, l'interdisciplinarité suppose d'accepter les limites imposées aux désirs de chacun. Pour que les littéraires aient un outil informatique simple, il faut ainsi que les informaticiens

31

32

acceptent de sacrifier les esthétismes informatiques, qui, le plus souvent amèneraient à concevoir des outils trop complexes d'utilisation pour des chercheurs peu compétents en informatique. Inversement, les littéraires doivent accepter que tout ne soit pas possible du point de vue informatique. Il n'est par exemple pas possible de mettre l'élément « biffe » que sur une seule ligne à la fois, même quand plusieurs lignes d'un paragraphe ont été rayées par Stendhal : ce serait faisable techniquement, mais l'outil deviendrait alors trop complexe à utiliser par des non spécialistes en informatique³³.

Les linguistes acceptent des littéraires l'idée que leur corpus aura une représentativité limitée (un seul auteur, une seule période...), ce qui va à l'encontre de la définition de la linguistique de corpus donnée par Sinclair. Les linguistes et les littéraires acceptent des informaticiens que le programme ne soit qu'un calcul et non une « intelligence » (même si on est dans le domaine de l'intelligence artificielle) qui présentera par conséquent des défauts et n'aura pas de compétences d'interprétation. Les informaticiens, quant à eux, acceptent l'imperfection humaine des littéraires dans la création du matériau.

De telles contraintes et concessions réciproques supposent plus de travail pour chacun puisque le projet change notre façon de travailler sur notre objet d'étude et même de le penser. Il a ainsi été nécessaire d'inventer des normes et des méthodologies inédites ; un cahier des charges a été initialement défini, mais il a été constamment aménagé depuis, ce qui suppose une grande souplesse dans la modélisation informatique du fait de la définition de nouveaux besoins par les littéraires au fur et à mesure de l'avancement de leur travail sur le fonds.

Mais tout en étant un lieu de tensions, d'efforts, cette articulation permet par une volonté de décloisonner les disciplines, de créer de nouveaux champs d'exploration et d'analyse pour chacune, en suscitant de nouvelles représentations de l'objet. Le travail des littéraires fournit un matériau concret, définit des besoins précis et amène les informaticiens et linguistes à remettre en question leurs préjugés en matière d'adaptabilité et de compétence informatique des utilisateurs d'outils de TAL. Et ces outils permettent aux littéraires d'accéder à de nouvelles représentations de leur objet (par exemple la représentation par cartographie de termes stendhaliens).



Figure 9: Exemple typique de représentation d'occurrences de deux champs lexicaux (« amour » et « prison ») dans la *Chartreuse de Parme*, où dans chaque bloc qui représente un chapitre figurent en noir le nombre d'occurrences des termes de l'« amour » et en rouge celui des termes de la « prison ».

Cette démarche amène ainsi chaque discipline à enrichir les autres disciplines de ses problématiques, de ses méthodes, de ses savoirs. En tenant compte de situations inédites, il s'agit pour l'équipe de régir précisément la « construction de représentations au carrefour de plusieurs disciplines³⁴ ».

Conclusions

Alors, « la littérature est-elle en mal de discipline ? » Notre réponse est évidemment négative et vise à montrer à quel point il est enrichissant pour elle de partager son objet d'étude avec d'autres disciplines, qui profitent elles aussi de cet échange. Le projet « Manuscrits de Stendhal » participe en ce sens à la réactualisation de l'interdisciplinarité : le fonds de manuscrits de Stendhal, contrairement à ce qu'on aurait pu penser naïvement, n'est pas un objet purement et exclusivement littéraire, et il s'est avéré bien au contraire que la multiplicité des représentations rendues possibles par l'intervention et le dialogue de plusieurs disciplines a enrichi les regards portés sur cet objet. Nous faisons ainsi nôtre l'affirmation de Frank Alvarez-Pereyre selon laquelle « aucune discipline n'épuise [...] un objet à elle seule [...]³⁵ ». Il serait présomptueux de prétendre que l'articulation interdisciplinaire est en mesure d'épuiser un objet, ce n'est certainement pas le cas. L'interdisciplinarité permet tout au plus de croiser plusieurs points de vue, plusieurs interprétations de l'objet, ce qui est déjà mieux qu'un regard disciplinaire.

Par ailleurs, l'interdisciplinarité, telle que nous la percevons, la pratiquons et la définissons, ne peut se mettre en place qu'en recherche-action autour d'objets d'étude concrets, et elle permet de réduire l'émiettement consécutif à la « multiplicité des regards spécifiques, circonscrits³⁶ » propres à chaque discipline.

34

35

En effet, contrairement aux approches pluri- ou multidisciplinaires, il ne s'agit pas simplement de juxtaposer des points de vue mais d'aboutir à une « pensée décloisonnée³⁷ ».

Qui plus est, la volonté même de partager avec le plus grand nombre le patrimoine culturel, et pas uniquement littéraire, que constituent les manuscrits de Stendhal, est partie prenante d'un dessein plus ambitieux encore, celui d'ouvrir notre projet interdisciplinaire à d'autres disciplines non encore représentées dans notre équipe. Des intérêts ont d'ores et déjà été exprimés : dans le domaine de la didactique du français en lycée – la genèse de l'œuvre est au programme des classes de seconde ; dans le domaine de l'histoire – le regard que porte le diplomate et écrivain sur la société qui l'entoure, de la révolution remémorée de son enfance aux campagnes napoléoniennes.

En revendiquant ostensiblement une démarche interdisciplinaire fondée sur le dialogue et le constructivisme, nous sommes pleinement conscients que « l'interdisciplinarité dérange³⁸ ». Elle dérange les acteurs de la démarche interdisciplinaire dans la mesure où elle déstabilise un équilibre parfois précaire du système que constitue une discipline. Elle dérange pour les mêmes raisons les acteurs disciplinaires, créant parfois des réactions exacerbées de rejet. Les connaissances nouvelles déstabilisent cet équilibre que le système doit tenter de retrouver en fonction de ce nouveau contrepoint cognitif, ce que Piaget³⁹ exprime comme un principe d'acquisition :

L'acquisition d'une information se traduit par une « perturbation » qui va entraîner chez l'individu un « déséquilibre » du champ cognitif et exiger un travail de synthèse pour assimiler, intégrer, critiquer, admettre, ajouter cette nouvelle dans un champ cognitif alors enrichi.

Le choix de l'interdisciplinarité – car il s'agit d'un choix – est parfois considéré comme une folie de jeune chercheur ou une lubie de chercheur en fin de carrière. Folie car elle constitue un risque pour la carrière du jeune chercheur qui verse dans l'interdisciplinarité puisqu'il est à cheval sur plusieurs communautés qui pourraient ne pas le reconnaître comme l'un des leurs – c'est le problème, évoqué en introduction, des catégorisations.

Les interdisciplinaires doivent enfin se revendiquer comme tels sans toutefois perdre leurs identités respectives. Auteurs de cet article et porteurs du projet manuscrits de Stendhal, nous restons pour l'une littéraire, pour l'autre linguiste-

36

37

38

39

informaticien. Si le frottement disciplinaire a permis à chacun de nous de s'approprier une partie des connaissances de l'autre, aucun n'aura la prétention d'être détenteur des connaissances et des compétences de l'autre.

BIBLIOGRAPHIE

Sitographie :

Bibliothèque numérique européenne Europeana : <http://europeana.eu>

Manuscrit du *Côté de Guermantes* de Proust : BnF, cote NAF 18317, <http://gallica.bnf.fr/ark:/12148/btv1b6000143s.image.f1.langFR>

Manuscrits de Nietzsche : <http://www.nietzschesource.org/>

Manuscrits de *Madame Bovary* de Flaubert : <http://bovary.univ-rouen.fr/>

Atelier des Manuscrits de *madame Bovary* : <http://flaubert.univ-rouen.fr/bovary/atelier/>

Manuscrits de Bouvard et Pécuchet de Flaubert : <http://dossiers-flaubert.ish-lyon.cnrs.fr/>

Manuscrits du *Rêve* de Zola : <http://gallica.bnf.fr/dossiers/html/dossiers/Zola/>

Manuscrits du *Journal* de Stendhal : <http://stendhal.bm-grenoble.fr/journal/>.

Manuscrits de Stendhal : <http://manuscrits-de-stendhal.org/>

Site de l'Institut des Textes Et Manuscrits Modernes : <http://www.item.ens.fr/index.php?id=14059>

Alvarez-Pereyre (Frank), *L'Exigence interdisciplinaire. Une pédagogie de l'interdisciplinarité en linguistique, ethnologie et ethnomusicologie*, Paris, Éditions de la Maison des Sciences de l'Homme, 2003.

Burnard (Lou), « XML + TEI, un mariage fait aux cieux », présentation à la MSH Lyon, novembre 2001, URL de référence : <http://www.tei-c.org/Talks/lyon-3.ppt>.

Clark (Herbert H.), Marshall (Catherine R.), « Definite reference and mutual knowledge » dans *Elements of discourse understanding*, sous la dir. de Joshi (A.K.), Webber (B.L.), SAG (I.A.), Cambridge, Cambridge University Press, 1981, p. 10-63.

Giffard (Alain), *Pour en finir avec la Mécroissance*, Paris, Flammarion, 2009.

Gilster (Paul), *Digital literacy*, New York, Wiley and Computer Publishing, 1997.

Glykos (Alain), « Approche communicationnelle du dialogue Artiste/Scientifique », note de synthèse pour l'Habilitation à Diriger des Recherches, Paris, Université Diderot – Paris 7, 1999.

Kuhn (Thomas Samuel), *La Structure des révolutions scientifiques*, Paris, Flammarion, coll. « Champs », n°115, 1993.

Lebarbé (Thomas), « Fonctions interdisciplinaires intrinsèques et extrinsèques du traitement automatique des langues », note de synthèse pour l'Habilitation à Diriger des Recherches, Grenoble, Université de Grenoble, 2010, URL de référence : <http://w3.u-grenoble3.fr/lebarbe/uploads/Main/hdr.pdf>

Maingain (Alain), Dufour (Barbara), sous la direction de Fourez (Gérard), *Approches didactiques de l'interdisciplinarité*, Bruxelles, De Boeck Université, coll. « Perspectives en éducation et formation », 2002.

OCDE, *La littératie à l'ère de l'information, Rapport final de l'enquête internationale sur la littératie des adultes*, 2000, URL de référence : <http://www.oecd.org/dataoecd/24/62/39438013.pdf>

Piaget Jean *et alii*, *Construction et validation des théories scientifiques*, Fondation Jean Piaget, Archives Jean Piaget, 1980.

Vanier Martin, « Territoires, territorialité, territorialisation : et après ? – Introduction », dans *Territoires, territorialité, territorialisation : controverses et perspectives*, Rennes, Presses Universitaires de Rennes, 2009.

PLAN

- [Préambule](#)
- [Au début était le texte sur papier](#)
- [Dépasser la terminologie et les représentations disciplinaires : du manuscrit à sa transcription](#)
- [Recourir à différents champs disciplinaires pour donner les manuscrits à voir, à lire et à comprendre](#)
- [Trans-, Pluri- ou Interdisciplinarité ?26](#)
- [Interactions interdisciplinaires](#)
- [Conclusions](#)

AUTEURS

Cécile Meynard

[Voir ses autres contributions](#)

Courriel : cecile.meynard@gmail.com

Thomas Lebarbé

[Voir ses autres contributions](#)

Courriel : thomaslebarbe@gmail.com