**Manuscrits de Stendhal : Du patrimoine papier au document électronique**

**Auriane FAURE(1), Thomas LEBARBÉ(1), Cécile MEYNARD(2), Aïcha TOUATI(1)**

**(***1) Laboratoire LIDILEM (EA 609)*

*(2) Equipe « TRAVERSES 19-21 » (EA 3748) Université Stendhal – Grenoble 3*

# 1 Du manuscrit à la plateforme en ligne

La Ville de Grenoble possède la quasi totalité des manuscrits laissés par Stendhal à sa mort, soit environ 20 000 feuillets. Cet ensemble constitue l’un des plus importants fonds de manuscrits littéraires modernes en France, et à ce titre représente un élément précieux du patrimoine culturel et scientifique.

L’Etat français et les collectivités régionales et locales ont investi des sommes importantes depuis le début du XXème siècle pour acquérir les différents documents du fonds : Pour ne prendre qu’un exemple récent, 6 cahiers des journaux de Stendhal dit « cahiers Bérès » ont été achetés pour 900 000 euros en décembre 2006. Il a d’ailleurs aussi fallu faire appel à des mécènes privés pour réunir une telle somme : se pose alors le problème légitime de la mise à disposition du public de ces manuscrits. Il n’est évidemment pas imaginable de laisser tout un chacun consulter les manuscrits, précieux et fragiles, qui ne doivent faire l’objet que de consultations ponctuelles et justifiées afin d’être préservés pour les générations futures.

La 1ère solution consiste à se contenter d’une simple numérisation des pages, les images étant ensuite mises en ligne. La Bibliothèque

municipale de Grenoble, qui a numérisé récemment le fonds des manuscrits de Stendhal (entre 2007 et 2009) a ainsi donné au public la possibilité de feuilleter les pages des cahiers Bérès[1](#_bookmark0). Mais cette mise à disposition s’avère d’un intérêt limité, car le manuscrit se réduit à n’être qu’un bel objet pour l’utilisateur curieux qui doit se contenter de tourner

1. <http://www.bm-grenoble.fr/patrimoine/acces-aux-collections-numerisees.htm>. La bibliothèque a également mis en ligne une reproduction de la belle édition diplomatique de Gérald et Yvonne Rannaud chez Klincksieck, qui offre à l’utilisateur la possibilité de consulter les images des pages et leurs transcriptions, et de faire des recherches simples dans le texte. Mais aucun enrichissement scientifique n’a été ajouté et la liberté de parcours de l’utilisateur reste limitée.

des pages sans pouvoir toujours lire les pattes de mouche de l’écriture stendhalienne (comme le note Almuth Gresillon, l’utilisateur se trouve ici dans une situation d’esthète, à « regarder le manuscrit comme on regarde un tableau » [2](#_bookmark1) ) ni forcément comprendre la logique de l’organisation interne des documents ou de leur rattachement à des ensembles plus vastes. Aucune recherche n’est par ailleurs possible dans les textes. L’intérêt scientifique, et même culturel, de la consultation reste donc limité.

C’est dans ce souci de valorisation du fonds (le rendre lisible et non plus simplement visible) que, en partenariat avec la Bibliothèque municipale de Grenoble, des chercheurs de l’Université Stendhal – Grenoble 3 se sont lancés dans la conception d’un site commun Ville / Université, couplé avec une base documentaire, CLELIA (Corpus littéraire et linguistique assisté par des outils d’intelligence artificielle). Le projet

« Manuscrits de Stendhal » s’appuie sur une collaboration fructueuse et inédite entre des littéraires de l’équipe Traverses 19-21 et des informaticiens et linguistes du laboratoire LIDILEM de l’Université Stendhal Grenoble 3.

Le principe est de donner à voir les pages numérisées des manuscrits de Stendhal, mais aussi leur transcription et différentes informations sur leur contenu textuel par le biais d’un moteur de recherche, en fournissant des modes d’accès et de représentation variés aux utilisateurs. La plateforme CLELIA a été en effet conçue en visant le plus large public possible, du

« grand public » aux spécialistes de Stendhal ou de la littérature du XIXème siècle.

**Les premiers utilisateurs de la base sont nécessairement les transcripteurs littéraires** qui vont l’alimenter progressivement. Ils doivent pouvoir saisir toutes les informations qui leur semblent pertinentes pour l’analyse des pages, et c’est pour cette raison que l’outil doit être adaptable et évolutif. Plus généralement, tous les chercheurs stendhaliens et spécialistes du XIXème siècle sont des utilisateurs potentiels de la base. Pour ne prendre qu’ un exemple, un chercheur travaillant sur le rôle du souligné et des traits en marge au crayon chez Stendhal (c’est le cas de Christopher Thompson, qui a montré combien cette pratique est intéressante et révélatrice sur la réutilisation par Stendhal d’extraits de ses textes dans d’autres textes, transcendant ainsi les genres littéraires traditionnels) doit pouvoir trouver ces informations dans la base.

Au niveau microscopique, celui de la page, l’outil doit permettre de reconstituer autant que faire se peut la genèse de la page en identifiant les traces et les strates d’écriture, les ratures, variantes, soulignés, traits en

1. Almuth Grésillon, « Méthodes de lecture », Les manuscrits des écrivains, Paris, Hachette CNRS éditions, sous la direction de Louis Hay, 1993, p. 138-161 (la citation se trouve page 143).

marge, interlignes, ajouts, notes... autant d’éléments qui peuvent apporter des informations essentielles aux chercheurs sur le travail d’écriture et d’auto-relecture de Stendhal[3](#_bookmark2).

Au niveau macroscopique, celui des ensembles de pages, il s’agit d’identifier et de représenter de façon rigoureuse les documents et ensembles documentaires qui ont souvent été déplacés, par Stendhal ou par les bibliothécaires au moment de la reliure des manuscrits, voire qui ont été « désossés » lors des éditions, comme s’ils appartenaient à des corpus différents, et sans tenir compte de l’unité du support. Pour ce faire, le travail sur le fonds a amené les chercheurs à rationaliser l’analyse codicologique [4](#_bookmark3) par un inventaire systématique dont les informations (dimensions des papiers, trous de couture permettant d’identifier des cahiers, des liasses, etc.) seront introduites dans la base pour permettre des regroupements de documents présentant les mêmes caractéristiques. Toutes ces informations sont essentielles pour envisager un reclassement virtuel de ces ensembles désorganisés. Des analyses littéraires sont ainsi rendues possibles par les requêtes effectuées, qui viennent infirmer ou confirmer de façon rationnelle les intuitions des chercheurs, ou peuvent même les amener à formuler de nouvelles hypothèses.

La deuxième fonction de l’outil mis en place pour ce public que constituent les transcripteurs est de permettre de produire des éditions papier à la demande, en s’appuyant réellement sur les manuscrits.

L’équipe littéraire a des exigences en termes de contenu et de mise en forme. Il est important en effet autant que possible d’être fidèle à la mise en page stendhalienne, qui a le plus souvent une signification. Ainsi il convient de conserver le statut et la présentation des titres, qui donnent souvent une dimension solennelle à un début de cahier. Autre exemple, les notes de bas de page et les marginales n’ont en général pas la même fonction, les premières contiennent les références bibliographiques et les données chiffrées, tandis que les secondes peuvent être soit un commentaire du texte en regard duquel elles se trouvent, soit une simple notation diariste de l’état physique et mental de Stendhal et de ses activités et observations du moment.

Enfin, il s’est avéré pertinent de signaler les réclames et contre-réclames[5](#_bookmark4) pour respecter la mise en page voulue par Stendhal mais aussi pour préparer le travail sur l’édition papier. En effet, il faudra procéder à la

1. Stendhal étant coutumier de l’annotation a posteriori de ses propres écrits, nous avons ainsi créé un corpus « Notes de relecture de l’année XXXX » pour permettre des regroupements virtuels et identifier ainsi de façon rigoureuse ses centres d’intérêt selon les époques de sa vie, ses périodes de relecture active, etc.
2. L’analyse codicologique est l’analyse des caractéristiques du papier du document, voire du feuillet (identification de traces de couture, de piqûres d’épingle, etc.)
3. La réclame est un mot ou groupe de mots que le scripteur écrit sur la dernière ligne de la page en l’ (les) alignant à droite, et qu’il répète éventuellement au début de la première ligne de la page suivante (contre-réclame). Cette pratique était courante dans les manuscrits et œuvres publiées jusqu’au début du XIXème siècle.

suppression automatique de tous les mots désignés comme contre- réclames pour éviter la répétition de ces mots).

Le **deuxième type d’utilisateurs regroupe les linguistes** pour lesquels les manuscrits constituent un corpus inédit. En effet, les manuscrits de Stendhal représentent 40 années d’écriture. Sous leur forme papier, les manuscrits sont difficilement utilisables pour le chercheur en linguistique. Il en est de même pour les numérisations qui présentent le défaut de lisibilité déjà évoqué plus haut.

En revanche, transcrits et annotés rigoureusement, les manuscrits représentent un matériau langagier unique : des milliers de pages d’écriture, appartenant à différent styles (diariste, ébauches et critiques littéraires et théâtrales…), dont tous les composants sont délimités et identifiés, quelque soit le grain hiérarchique (du bloc de texte au mot biffé en passant par lignes, paragraphes, ajouts en marge…).

Ainsi structuré, l’ensemble des pages de manuscrits forme non seulement un ensemble de corpus dans le sens littéraire du terme, mais aussi un corpus au sens linguistique du terme. Parmi les objets d’études linguistiques, nous envisageons notamment la caractérisation du sabir[6](#_bookmark5), de la dysgraphie (ou paragraphie [7](#_bookmark6) ), la description linguistique des phénomènes de réécriture assimilables à des formes de disfluences écrites…

**Les professeurs de lycées et leurs élèves** constituent le troisième type d’utilisateurs visés, sachant que l’étude de la genèse des œuvres littéraires est au programme de français en seconde, afin de permettre aux élèves de mieux comprendre le processus de création chez les écrivains. Des parcours pédagogiques simples d’accès et d’utilisation, éventuellement téléchargeables, doivent donc être prévus, en gardant à l’esprit qu’il s’agit de distinguer les manuscrits de leur image souvent un peu poussiéreuse et ennuyeuse, en montrant aux jeunes générations à quel point l’analyse de la genèse d’une œuvre peut prendre des dimension inattendues d’enquête à partir d’indices. On mettra ainsi à disposition des enseignants des dossiers portant par exemple sur les ensembles significatifs de pages illustrant la démarche de création des personnages de romans chez Stendhal, sur les pratiques d’écriture autobiographique, ou sur le plagiat, en laissant bien sûr à l’enseignant, voire à ses élèves, la possibilité de se constituer des dossiers personnalisés.

1. Le sabir est une pratique récurrente chez Stendhal qui intègre dans ses écrits des séquences en langue étrangère (Bordas, 2007), cette pratique variant dans le temps, atteignant son paroxysme à la fin de sa vie dans « Earline ».
2. Les termes de dysgraphie et de paragraphie sont tous deux utilisés pour désigner les variantes orthographiques par rapport à la norme. Seule l’étude approfondie de ces phénomènes nous permettra de déterminer lequel des deux termes est le plus adéquat.

Enfin, dans ce souci légitime d’exhaustivité et de rigueur scientifique, il ne faut pas, oublier un public essentiel : **les amateurs éclairés et les simples curieux**. De fait, la valorisation du patrimoine culturel et scientifique doit se faire en ayant en tête un souci de vulgarisation pour que la diffusion de ces informations culturelles ne concerne pas qu’une élite. D’où la nécessité de mettre en place des parcours guidés ludiques et interactifs dans le fonds des manuscrits. Le principe sera ainsi, entre autres exemples, de donner à voir et à entendre par le biais d’hyperliens (faire voir la reproduction d’un tableau dont parle Stendhal, ou faire entendre un extrait d’un morceau de musique évoqué), de faire écouter un commentaire oral sur une page de manuscrit, de permettre un affichage dynamique de la page en cours d’écriture ou de correction. Il s’agit donc de montrer un auteur vivant à travers son œuvre.

Les recueils physiques sont reproduits numériquement, mais l’utilisateur peut aussi accéder aux manuscrits par le biais de parcours guidés, de regroupements par types d’écrits mais aussi par recherche simple ou avancée. La quantité et l’affinement des annotations sont adaptés au type d’utilisateur pour ne pas surcharger l’affichage inutilement. La plateforme permet par ailleurs aux utilisateurs de constituer leurs propres recueils de pages manuscrites. A terme, cette fonctionnalité permettra des éditions numériques et papier à la demande.

# 2 Une description « sémantique » des pages manuscrites

La page de manuscrit est un objet complexe à décrire. Y intégrer des informations paratextuelles, d’ordre scientifique ou didactique, augmente et complexifie la tâche de description.

Il existe de nombreuses normes de description et d’encodage de textes, parfois contradictoires ou antagonistes, représentant des points de vues différents sur l’objet textuel de manière général. La TEI (*Text Encoding Initiative*) joue un rôle majeur dans cet univers depuis plus de deux décennies[8](#_bookmark7), plus encore depuis la plus récente création de *guidelines* pour la transcription des manuscrits[9](#_bookmark8).

Toutefois, dans le cadre de ce projet, nous devions répondre à trois impératifs : 1) un impératif scientifique de description précise, 2) un impératif économique dû aux faibles ressources humaines pour le développement logiciel et 3) un impératif d’accessibilité à des utilisateurs peu formés aux outils et formalismes informatiques. Afin de répondre à ces trois exigences, nous avons opté pour une grammaire de description

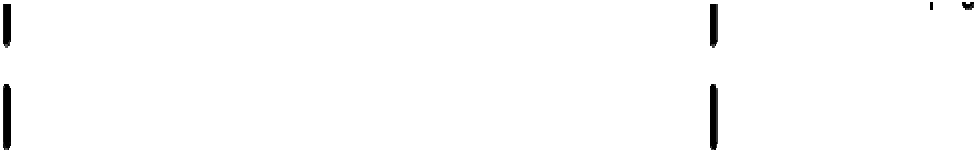
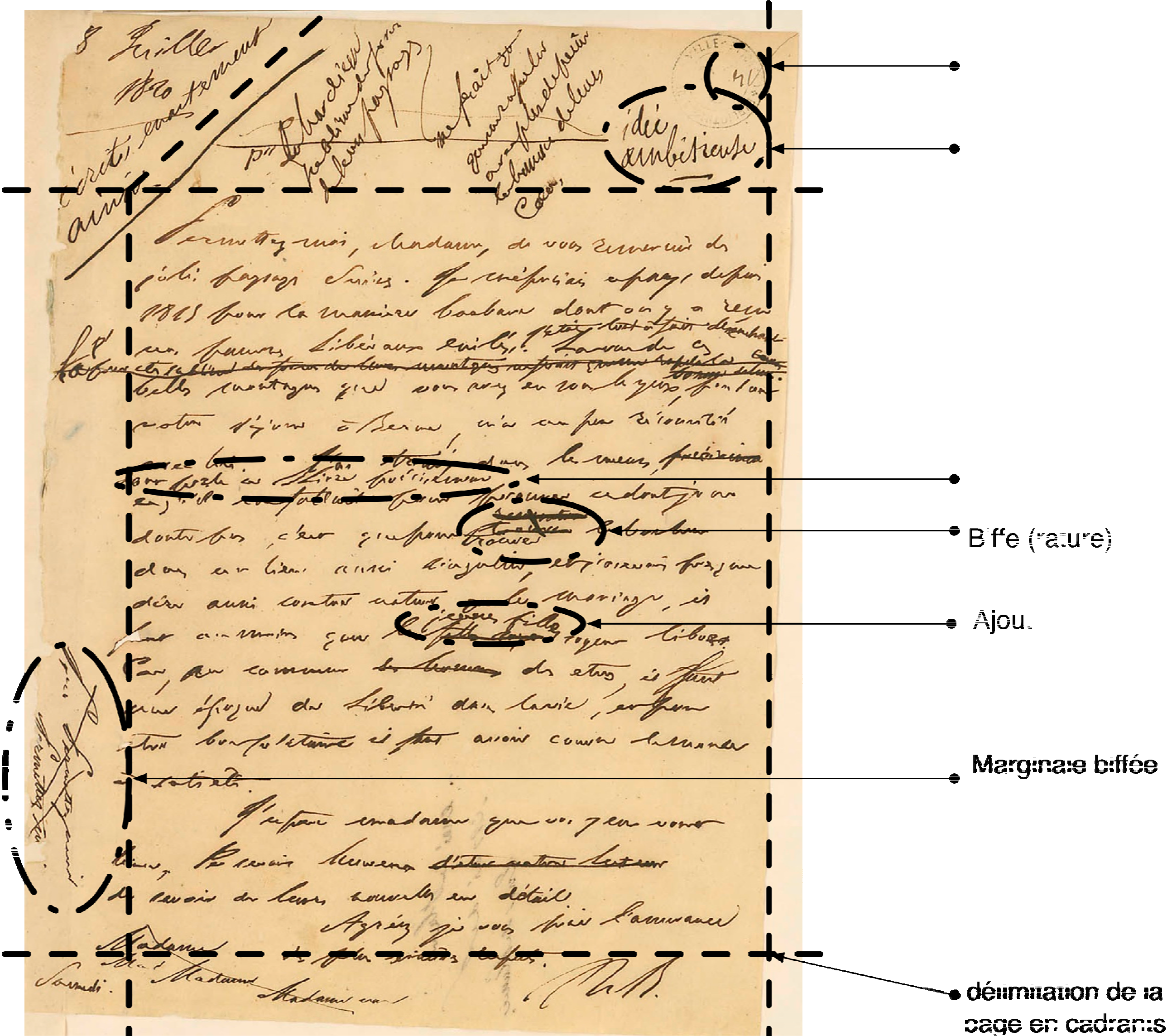
1. Text Encoding Initiative – History : <http://www.tei-c.org/About/history.xml>
2. Guideline for « Manuscript description » : <http://www.tei-c.org/release/doc/tei-p5-doc/html/MS.html>

(DTD) conçue dans un dialogue interdisciplinaire qui permette de nommer les objets et leurs propriétés dans une terminologie accessible aux transcripteurs. Cette grammaire est accompagnée d’une feuille de style permettant une visualisation approximative de la transcription dans un logiciel d’édition de fichiers XML libre de droit. Ainsi, les transcripteurs sont guidés et contraints dans leur tâche par la DTD sur un outil ne nécessitant que peu d’apprentissage, et disposent d’un rendu visuel s’approchant du rendu envisagé en ligne.

A l’image de nombreux formalismes de description, la grammaire développée pour les manuscrits de Stendhal se décompose en deux parties principales : une en-tête de méta-données permettant pour le référencement et les renseignements sur la page incluant un commentaire du transcripteur ; et un corps contenant la transcription elle-même. L’équipe ne souhaitait pas une transcription hyper-diplomatique, par conséquent, la page est décomposée en 9 cadrans pour le positionnement des unités textuelles et graphiques : la zone centrale principale, les quatre coins de la page, les quatre cotés (marges lattérales, supérieure et inférieure – voir figure 1 page suivante). Chacune de ces zones peut contenir des éléments textuels (blocs de texte, blocs de citation, paginations, foliotations, marginales, notes…) qui se décomposent en entités identifiables visuellement (paragraphes, lignes, interlignes, figures, tableaux…) et en entités de mises en forme (biffe, calligraphie…). Tous les éléments peuvent être enrichis d’annotations d’ordre critique (commentaires pour le grand public, pour les spécialistes, pour les membres de l’équipe, identification du scripteur, datation, localisation géographique…). Enfin, l’ensemble est complété d’un système de pointage et de références. En effet, il arrive qu’un ajout soit effectué en interligne au dessus de son point d’ancrage, puis s’enchaîne faute de place en dessous de ce point d’ancrage pour se terminer en marge. La description des éléments textuels est faite dans une représentation pseudo-diplomatique (à l’image de la page), la représentation linéarisée (à l’image de la résultante de la tâche scripturale) est calculée notamment grâce à ce système de pointeurs.

Fondée sur XML, la grammaire de description en hérite les qualités et les défauts. Ces derniers sont reconnus, notamment la contrainte d’imbrication des balises. Certes, des solutions sont proposées, telle celle proposée par (Portier, 2009) qui dissocie le contenu de son annotation, ou LMNL (Caton, 2005) qui permet l’enchevêtrement de balises ouvrantes et fermantes. Ces méthodes et techniques présentent néanmoins le défaut d’être peu outillées et peu intuitives pour l’utilisateur peu formé en informatique. Par ailleurs, les plateformes de partages de données textuelles, telles Pinakes 3 (Scotti, 2006), se fonde sur un encodage XML voire sur la TEI. C’est pourquoi nous avons planifié au sein du projet le

développement de modules de conversion des données vers la TEI. Cette prospective a certes influencé les principes d’encodages tout en laissant à l’équipe une grande liberté sur ses choix méthodologiques.



*Figure 1 : délimitation de la page de manuscrit en cadrants et dénomination des éléments (R. 5896, volume 1, feuillet 71 recto, image propriété de la BmG).*

Les transcripteurs disposent d’un outil hors-ligne pour effectuer les transcriptions. La plateforme CLELIA leur permet de déposer les transcriptions XML et de les visualiser telles qu’elles apparaîtront aux différents types d’utilisateurs et ainsi de corriger les transcriptions le cas échéant (la mise en ligne est assistée par une analyse du fichier mettant en évidence les erreurs et incohérences des données). Un processus de relecture puis de validation par les pairs, accompagné d’un code déontologique, permet de garantir la qualité scientifique des transcriptions mises à la disposition du public.

Du point de vue de l’utilisateur, l’accès aux manuscrits se fait selon trois méthodes différentes :

* par les registres physiques (à l’image des recueils conservés à la BmG) ou par regroupement cohérents d’un point de vue littéraire (corpus) ;
* par des regroupement artificiels générés automatiquement (ex : « les pages contenant des dessins de la main de Stendhal ») ou conçus par les spécialistes (ex : « les plus belles pages ») ou par des utilisateurs ;
* par recherche de mots-clés en plein texte.

La consultation des manuscrits correspondant au choix de l’utilisateur peut alors se faire selon trois modes d’affichage, l’utilisateur ayant toute liberté de basculer d’un mode à l’autre :

* par « planche contact » des pages numérisées, à l’image de leurs homonymes photographiques, permettant d’identifier rapidement la ou les pages pertinentes ;
* par « feuilletage », tel un livre dont on tourne les pages ;
* par vis-à-vis de la page et de sa transcription donnant ainsi une aide à la lecture et à l’analyse (par le biais d’infobulles).

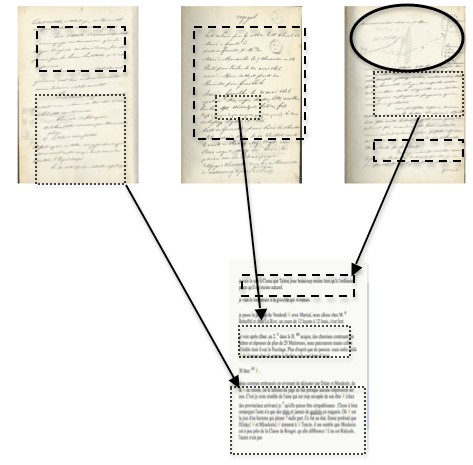
La mise à disposition des manuscrits et de leurs transcriptions sur Internet n’est toutefois qu’un des aspects (certes majeurs) du projet et de la plateforme. Les transcriptions ainsi formalisées et enrichies constituent une donnée structurée qui permet de se défaire de la page en tant qu’objet physique,

**2.1 De la transcription à l’ontologie des documents et des usages**

L’objet transcrit et affiché est la page. Il est décrit non seulement par des propriétés physiques telles que le format, le type de page, le scripteur, la date de rédaction, etc., mais aussi par son contenu. Le contenu de la page est l’ensemble des éléments textuels la constituant (pagination, foliotation, marginales, note de bas de page, ajout en interligne, paragraphe, titre, etc.). Chaque bloc de texte dispose d’un ensemble de propriétés (corpus, scripteur, emplacement dans la page, type d’écriture, etc.) où chaque élément textuel peut contenir à son tour d’autres types de blocs de texte qui héritent de ses propriétés de façon implicite ; ces blocs peuvent également disposer de propriétés différentes, spécifiées explicitement.

Une page peut être considérée comme un objet dont les éléments sont organisés hiérarchiquement. L’observation de ces éléments peut se faire aux différents niveaux de la hiérarchie, correspondant à autant de niveaux de granularité. L’objectif de CLELIA et de permettre aux utilisateurs d’interroger l’ensemble documentaire des manuscrits afin de reconstruire virtuellement des objets textuels à tous les niveaux de granularité. Au niveau macroscopique, l’outil permet de remettre de l’ordre des pages, par exemple de reconstruire le « *Journal de mon 3ème voyage à Paris,*

*1804-1805* », ensemble de cahiers aujourd’hui physiquement en désordre. Au niveau microscopique, d’une part les littéraires pourraient, par exemple consulter ***«*** *les pages de manuscrits de type diariste, contenant des figures, de la main de Stendhal, figure représentant des plans, classées selon les lieux décrits, puis par date de rédaction****».*** D’autre part, les linguistes chercheront un corpus correspondant à des critères tels que *« paragraphes comportant des biffes et des ajouts, classés par type d’écrits, puis par ordre chronologique »* pour étudier le phénomène de réécriture.



(a)

(b)

(d)

(c)

*Figure 2. Exemple construit de restructuration dynamique*

Comme nous l’avons cité plus haut**,** le formalise XML conçu pour les transcriptions permet de décrire fonctionnellement et sémantiquement les éléments textuels de la page manuscrite et d’en donner une représentation

hiérarchique. Pour répondre aux diverses requêtes, l’ensemble des éléments textuels de l’ensemble des pages est indexé, leur contenu et leurs propriétés sont ainsi enregistrées en conservant l’organisation hiérarchique. Cette indexation nous permet par la suite comme l’illustre la figure ci-dessous (Figure 1), de décomposer les éléments textuels constituants la page, pour nous donner une plus grande flexibilité de manipulation de ces objets (éléments textuels), les objets, quels qu’ils soient – de la page au mot -, peuvent être sélectionnés, en fonction de leurs propriétés, puis classé selon d’autres. Dans la figure 2, le document résultant (c) de la composition de fragments textuels présents dans trois pages manuscrites (a, b et c), est fait dynamiquement grâce aux propriétés qui caractérisent ces éléments et qui correspondent aux critères de sélection données par l’utilisateur.

La restructuration dynamique de corpus nous conduit à nous interroger sur l’accès au contenu des Manuscrits, et sur la navigation au sein des bases documentaires. Les publics de CLELIA sont nombreux, et ont des pratiques documentaires variées, que ce soit entre individus ou entre groupes. Pour permettre à chaque utilisateur de parcourir l’ensemble documentaire selon ses besoins propres, nous avons intégré à CLELIA un système de recherche d’information (SRI), couplé à une aide contextualisée. Comme l’expriment (Labiche & Holzeim, 2009) : « Nous avons estimé nécessaire de réfléchir à la conception de systèmes pour lesquels les interactions avec ces utilisateurs-concepteurs sont essentiels ».

Devant la masse de données et les différences entre utilisateurs, il nous semble important d’aider l’utilisateur de façon personnalisée. Les conditions d’interprétation d’une information sont différentes pour chacun ; La construction du sens est un acte dynamique et individuel. Afin de rendre la navigation pertinente pour chaque utilisateur, il nous apparaît nécessaire de nous inscrire dans une démarche centrée utilisateurs.

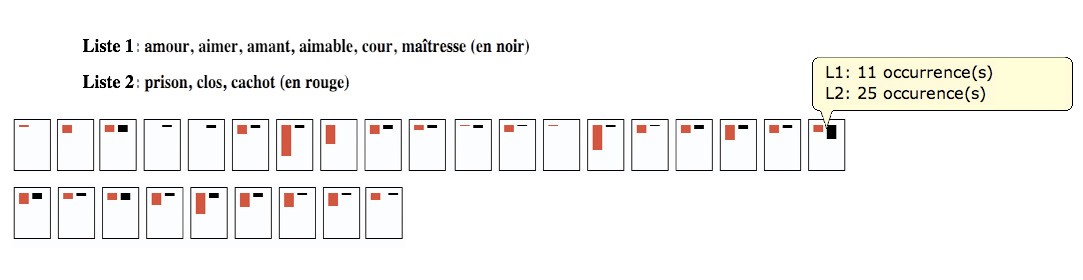
Le système de recherche d’information greffé à CLELIA fonctionne de manière simple. Il comprend une interface d’interrogation, par laquelle l’utilisateur exprime son besoin, ainsi qu’une interface de réponse, regroupant les résultats que l’utilisateur doit interpréter. L’articulation de ces deux étapes consiste en la mise en relation de l’expression du besoin de l’utilisateur et des données contenues dans la base documentaire, modélisées au préalable, pour sélection et affichage.

Dans l’interface d’interrogation, l’utilisateur exprime son besoin par une requête composée de listes de termes (une ou plusieurs, composées d’au moins un terme chacune). Par observation empirique [10](#_bookmark9) , nous pouvons

1. Expérience menée sur quatre utilisatrices dans le cadre de travaux de Master 2 : (Faure, 2008)

affirmer que les utilisateurs construisent une liste comme un ensemble lexical définissant un « thème » de recherche (soit un angle de lecture). La construction de plusieurs listes vise ainsi à caractériser ou opposer plusieurs thèmes. Nous laissons ainsi l’utilisateur couvrir des champs lexicaux les plus larges possibles.

La sélection des documents du fonds correspondant à la recherche est effectuée par l’intermédiaire d’un index, qui contient l’ensemble du vocabulaire des Manuscrits. Lorsqu’un terme contenu dans une requête d’utilisateur est indexé, la mise en relation avec le(s) document(s) dont le terme est issu est opéré. Les documents modélisés le sont donc par l’intégralité de leur contenu, et non par une liste de mots-clés issus d’une sélection (humaine ou automatique). Il n’y a donc pas d’étape intermédiaire entre le document et la formulation du besoin de l’utilisateur ; l’absence d’interprétation donnée par une tierce personne va dans le sens d’un SRI centré utilisateur.

Le résultat est ensuite affiché sous la forme d’une cartographie, représentation graphique de l’ensemble documentaire sur laquelle est projetée la requête de l’utilisateur. Cette cartographie est un support de navigation plus éloquent qu’une liste de liens vers les documents sélectionnés. Elle constitue un point d’entrée dans l’ensemble documentaire ainsi que dans le document lui-même. Elle constitue également un premier support d’interprétation pour l’utilisateur, qui peut observer les différences ou ressemblances entre documents au regard de sa recherche. Il peut décider de consulter un document, ou de réajuster sa recherche. La cartographie est donc le point de départ de la navigation dans l’espace documentaire.

*Figure 3 : Cartographie selon deux listes de termes (quantifiées)*

Afin d’apporter une aide contextualisée à l’utilisateur construisant son parcours documentaire, c’est-à-dire centrée sur ses besoins et ses pratiques, nous proposons la construction d’une ressource terminologique qui lui soit propre. Cette ressource lui est proposée quand il construit sa requête de recherche. Lorsqu’un terme ajouté à la liste en construction existe dans la ressource terminologique, ses termes frères[11](#_bookmark10) sont proposés à l’utilisateur. Dans la perspective du sens construit dans l’interprétation,

1. Par termes frères au sein de la RTO, nous entendons les termes coprésents avec le terme de la requête courante

il semblerait incohérent de proposer une seule et même ressource lexicale à tous les utilisateurs, tel un dictionnaire des synonymes, ou un panel « à l’aveugle » des dix termes les plus employés dans les Manuscrits ou dans les requêtes d’utilisateurs. Nous souhaitons proposer des ressources légères individuelles, qui soient pertinentes en contexte et de taille raisonnable[12](#_bookmark11).

Dans (Roy & Beust, 2006), l’utilisateur exprime ses connaissances sur un domaine en organisant des lexies par regroupement ou par opposition. Nous choisissons pour notre part de construire la ressource de manière automatique, en l’implémentant à chaque nouvelle requête d’utilisateur. Les traces d’interaction entre l’utilisateur et le SRI viennent ainsi compléter la ressource terminologique de l’usager, en tenant compte des termes existant au préalable et des liens qu’ils entretiennent[13](#_bookmark12).

La ressource construite traduit les centres d’intérêt et les pratiques de l’utilisateur. En effet, les termes régulièrement employés seront liés à un vocabulaire large, tandis que les termes n’ayant pas satisfait la recherche de l’utilisateur tomberont en désuétude. Par ailleurs, la navigation inter et intra-textuelle que mène l’utilisateur lui permet de découvrir de nouveaux termes appartenant au vocabulaire de Stendhal mais pas forcément au sien. L’intégration de ces termes dans les requêtes de recherche suivantes conduit du même coup à leur intégration dans la ressource terminologique.

De nombreux travaux se sont penchés sur l’extraction automatique de termes pour construire des ressources terminologiques, mettant au jour un vocabulaire propre aux auteurs des documents de référence. Cette perspective ne correspond pas à notre démarche centrée utilisateur. Nous privilégions une ressource issue des seules pratiques de l’utilisateur, contenant exclusivement des termes qu’il choisit d’employer en conscience. Une telle ressource modélise donc à la fois ses centres d’intérêts et ses usages, face à l’outil et à la navigation dans l’ensemble documentaire : il s’agit de modéliser son parcours interprétatif. Il devient ainsi possible d’apporter une aide plus concrète à l’utilisateur dans sa recherche, par l’accès à cette ressource durant le parcours. Ces ressources sont également mutualisées, pour générer une ressource globale afin d’émettre de nouvelles pistes interprétatives. La comparaison de traces d’utilisateurs différents devrait permettre d’esquisser des petits groupes d’utilisateurs dont les processus cognitifs et les usages sont proches, dans un objectif de modélisation.

1. Plusieurs travaux sur la création de ressources terminologiques personnelles ont inspiré cette proposition : citons (Perlerin, 2004) ou encore (Roy, 2007).
2. Pour une étude des liens entre termes d’une ressource terminologique personnelle, consulter (Faure & Lebarbé, 2009)

# Conclusion

Les méthodes et outils que nous présentons ont été développés autour du projet des manuscrits de Stendhal mais dans une perspective plus généraliste afin d’être adaptés en tant que plateforme ou en tant qu’outils indépendants à d’autres formes de ressources textuelles. Il est d’ores et déjà prévu d’étendre le champ d’application méthodologique et informatique aux Carnets des Canuts, propriété de la Ville de Lyon, autre patrimoine de l’histoire de la France. Conçus dans un dialogue interdisciplinaire permanent, ils offrent aux spécialistes comme au grand public de nouvelles approches documentaires du patrimoine culturel écrit ainsi enrichi, permettant de concevoir simultanément et complémentairement des éditions numériques en ligne et des éditions papier.

Valoriser le patrimoine apparaît ainsi comme le moyen de construire également des supports de réflexion méthodologiques pluridisciplinaires.

# Références bibliographiques

J.-L. Bouraoui, Ph. Boissière, M. Mojahid, N. Vigouroux, A. Lagarrigue, F. Vella, J.-L. Nespoulous. Problématique d'analyse et de modélisation des erreurs en production écrite. Approche interdisciplinaire. Actes de TALN 2009.

B. Boie. L’écrivain et ses manuscrits, Les manuscrits des écrivains, Paris, Hachette CNRS éditions, sous la direction de Louis Hay, 1993, p. 34- 53.

E. Bordas. Le Babeylisme scriptural de Stendhal, ou le style comme langue étrangère, Stendhal à Cosmopolis, ELLUG, sous la direction de Marie- Rose Corredor, 2007, p. 219-231.

P. Caton. « LMNL Matters », Extreme Markup Languages 2005, Montréal, Quebec

A. Faure. Cartographies et OntologieS des Manuscrits et Œuvres de Stendhal (COSMOS). Mémoire de Master 2, Université Stendhal Grenoble 3, 2008.

A. Grésillon. Méthodes de lecture , Les manuscrits des écrivains, Paris, Hachette CNRS éditions, sous la direction de Louis Hay, 1993, p. 138- 161.

J. Labiche. M. Holzem. Couplage et perturbation versus boîte noire et entrée-sortie. Journées de Rochebrune 2009, 2009.

V. Perlerin. Sémantique légère pour le document. Thèse de doctorat en Informatique, Université de Caen / Basse-Normandie, 2004.

P.-E. Portier, S. Calabretto. Modélisation des connaissances dans le cadre de bibliothèques numériques spécialisées. Extraction et Gestion des Connaissances (EGC) 2009, Strasbourg.

F. Rastier. Sémantique interprétative, Presses Universitaires de Frances, 1987.

T. Roy et P. Beust. Ressources termino-ontologiques différentielles personnelles : construction et projection sur corpus. Revue I3E, Hors- série, 2006.

T. Roy. Visualisations interactives pour l’aide personnalisée à l’interprétation d’ensembles documentaires. Thèse de Doctorat, Université de Caen/Basse-Normandie, 2007.

A. Scotti. Postgres & Java in the cultural heritage research: the Pinakes 3.0 Project. In WorldWide PG Day, Prato, Juin 2006.