

Colloque international  
Société de Langues et de Littératures Médiévales d'Oc et d'Oïl

# Éditions électroniques, études de corpus et bases textuelles dans les études médiévales



Nancy  
Campus Lettres et Sciences Humaines  
16 et 17 janvier 2015

[www.atilf.fr/dmf](http://www.atilf.fr/dmf)

## L'aide des ressources numériques à l'édition de texte : l'expérience de l'édition électronique du *Psautier d'Arundel*

Le Psautier contenu dans le ms. de la BL, Arundel 230, daté probablement de la seconde moitié du XII<sup>e</sup> siècle, est un texte bilingue : la glose française court parallèle au latin, et les blocs de mots qui la composent ne semblent pas tenir grand compte de la syntaxe française, mais sont plutôt disposés de sorte à se trouver en regard des groupes de mots latins qu'ils traduisent.

Pour l'édition du texte du Psautier d'Arundel, il est donc primordial de pouvoir respecter le parti de mise en texte du manuscrit, où manifestement le copiste recherche la correspondance visuelle entre les blocs de mots de la traduction française et ceux du psautier latin.

L'édition numérique nous a paru pouvoir satisfaire au mieux à cette exigence. Les programmes élaborés en collaboration avec les membres de l'équipe ICAR de l'ENS de Lyon rendent possible l'alignement des textes, et un système de barres verticales permet de visualiser à la fois la séparation existant entre blocs de mots sur la même ligne, et la correspondance verticale qui a été recherchée par le scribe entre les textes des deux langues.

Ce dispositif, conçu pour faciliter la lecture, s'est révélé utile d'abord pour l'éditeur : il permet par exemple de mieux percevoir les correspondances imparfaites, c'est-à-dire les rares cas où un mot n'est pas traduit, ou les écarts de traduction, comme par exemple les changements de temps verbal entre le latin et le français. Il revient évidemment ensuite à l'éditeur d'évaluer l'opportunité de corriger la leçon du manuscrit.

L'index lexical généré par le programme permet aussi de faire ressortir des formes à la graphie approchante, ce qui peut éventuellement amener l'éditeur à proposer des corrections pour des graphies isolées.

Le linguiste est aussi aidé dans son travail, puisque la présentation adoptée permet de mieux comprendre des formes et constructions françaises aberrantes, pour lesquelles le scribe du ms. Arundel s'est éloigné de son modèle (vraisemblablement le Psautier français contenu dans le ms. Oxford, Bodleian Douce 320), et qui s'éclairent grâce au texte latin souscrit, puisqu'il s'agit toujours de tentatives du copiste de calquer au plus près le cadre 'physique' fourni par l'ordre des mots du texte latin souscrit : verbes passifs français qui traduisent un déponent latin (ex. *parlé est* pour *locutus est* XVI, 11), possessifs traduits deux fois, une première fois d'après la syntaxe prépositionnelle du français, et la deuxième fois pour assurer la correspondance verticale avec le possessif postposé du latin (ex. *en ta maisun en ta* V, 8), pronom personnel traduit avant et après le verbe, toujours dans l'hésitation entre deux syntaxes différentes (ex. *tu me ampliras me* XV, 11), génitif traduit par la préposition *de* répétée devant chaque membre du syntagme latin (ex. *la voiz plurement del men* pour *vocem fletus mei* VI, 8, *en le nun del Seignur de Nostre Deu* pour *in nomine Domini nostri Dei* XIX, 8), etc.

Lors de la communication, nous nous proposons de montrer comment le programme d'édition numérique nous a permis de résoudre certains problèmes éditoriaux.

### **Bibliographie :**

- Beyer A. (1887-1888), « Die Londoner Psalterhandschrift Arundel 230 », *ZrP* 11, p. 513-534 et 12, p. 1-56.
- Pignatelli C. (à paraître), « Le traitement des possessifs dans deux Psautiers anglo-normands du 12<sup>e</sup> siècle: des indices pour l'émergence d'une syntaxe française », dans *DIACHRO VI. Le français en diachronie*. Actes du Colloque international (Leuven, 17-19 octobre 2012), éd. A. Carlier / M. Goyens / B. Lamiroy, Louvain, Peter Lang.
- Pignatelli C. & Lavrentiev A. (à paraître), « Le Psautier d'Arundel : une nouvelle édition », dans *Actes du XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013). Section 13 : Philologie textuelle et éditoriale*, éd. R. Trachsler / F. Duval / L. Leonardi [édition numérique].

**Philologie numérique et critique des textes,  
vers une application en lexicologie diachronique.  
Hypothèses et expérimentations sur la *Bible Historiale***

La *Bible Historiale* de Guiart des Moulins, composée en 1295, est une traduction en français qui rassemble la *Vulgate* de saint Jérôme et l'*Historia Scholastica* de Pierre Comestor. Visant un public laïque, lettré mais ne sachant pas le latin, la *Bible Historiale* propose une version narrative et en prose de l'Histoire Sainte avec des commentaires d'autorité éclairant le sens littéral et les fondements historiques des Écritures – l'*historia*. Cette œuvre fait partie des tentatives de compilation complète du texte biblique en langue vulgaire et fut enrichie dans ce mouvement, dans la suite du Moyen Âge et jusqu'au XVI<sup>e</sup> siècle, d'autres livres bibliques, tirés de la *Bible du XIII<sup>e</sup> siècle*. Elle fut dans ses différentes versions un grand succès de librairie.

Pour notre thèse, nous souhaitons proposer une édition de sa première version et constituer le lexique afin d'en faire une étude systématique.

Œuvre composée de plusieurs textes, la traduction est réalisée par un seul auteur et le résultat est une œuvre présentant une unité propre. Malgré cela, la complexité du texte pose plusieurs problèmes ecdotiques.

Il s'agit de rendre intelligible la structuration interne d'un texte aux sources multiples et imbriquées, et dont les gloses portent différentes informations : exégèse, notes traductives, repères intertextuels. Le texte contient lui-même des balises (« Notés chi », ou « le Maistre dist chi » par exemple) dans le corps du texte qui introduisent certains des commentaires du Comestor ou de Flavius Josèphe, et un système d'appel de notes et de lemmes pour les *marginalia*. Si dans les titres des chapitres, le traducteur annonce ses sources (selon le *Bible* et *Hystoire sour le Bible* – entendons par « Hystoire » l'*Historia Scholastica*), et si ponctuellement une note marginale donne l'origine du passage, peut-on se contenter de rendre compte de ce système ou doit-on aller plus loin dans l'analyse ? Le texte présente tout de même une suite de narration et de commentaires imbriqués implicitement qu'il s'agit de démarquer, le plus finement possible. Il faut ainsi déterminer un encodage XML-TEI précis.

La distinction des sources peut-elle mener à des résultats dans le cas de l'histoire des textes ou du fonctionnement de la langue d'un corpus complexe ?

L'élaboration de cet outil et ses pistes d'exploitation sont au centre de ce que nous souhaitons présenter. Nous voulons définir quel travail de repérage des textes, à l'intérieur du texte, doit être mis en œuvre et présenter la conception, de même que la mise en application, de cet outil de présentation et d'étude d'une œuvre et de son vocabulaire.

Le texte ainsi analysé, et accessible en plusieurs modes de lecture, rendra possible une exploration hypertextuelle, permettra d'identifier l'utilisation d'une expression dans ses différents contextes, d'identifier les interventions directes de l'auteur, et de mieux comprendre les articulations du texte. Nous pourrons ainsi mener une étude lexicale de façon différenciée, sur les plans quantitatifs et qualitatifs.

**Bibliographie :**

ANDRONACHE, Marta: « Architecture des gloses dans la première traduction en français de *La Cité de Dieu* (avant 1380). In : Neveu, Franck et al. (éd.), Congrès Mondial de Linguistique Française – CMLF 2010, Paris, ILF, publication électronique

([http://www.linguistiquefrancaise.org/index.php?option=com\\_toc&url=/articles/cmlf/abs/2010/01/contents/contents.html](http://www.linguistiquefrancaise.org/index.php?option=com_toc&url=/articles/cmlf/abs/2010/01/contents/contents.html)), 95-109.

BERGER, Samuel, *La Bible française au Moyen Âge - Étude sur les plus anciennes versions de la Bible écrites en prose de langue d'oïl*, Paris, Imprimerie nationale, 1884.

*BFM - Base de Français Médiéval* [En ligne]. Lyon : ENS de Lyon, Laboratoire ICAR, 2012, pignatelli psautier d'arundel<<http://bfm.ens-lyon.fr>>.

BRACHIMONT, Bruno et CROZAT : Stéphane « Instrumentation numérique des documents : pour une séparation fonds/forme. » sur le site : *Archive ouverte en sciences de l'information et de la communication*. 05/07/2004 [En ligne] Disponible sur : [http://archive.ccsd.cnrs.fr/docs/00/06/23/99/PDF/sic\\_00001017.pdf](http://archive.ccsd.cnrs.fr/docs/00/06/23/99/PDF/sic_00001017.pdf)

*Conseils pour l'édition des textes médiévaux*, Fascicule I : *Conseils généraux*, Paris, Comité des travaux historiques et scientifiques, École Nationale des Chartes, 2002

*Conseils pour l'édition des textes médiévaux*, Fascicule III : *Textes littéraires*, Paris, Comité des travaux historiques et scientifiques, École Nationale des Chartes, 2005

DUVAL, Frédéric : « La philologie française, pragmatique avant tout ? L'édition des textes médiévaux français en France. » in *Pratiques philologiques en Europe*, études réunies par Frédéric Duval, Paris, École des chartes (« Études et rencontres », 21), 2006

DUVAL, Frédéric : *Lectures françaises de la fin du Moyen Âge*, Textes littéraires français, DROZ, Genève, 2007

MICHEL, Bénédicte : *La « Bible historique » de Guiart des Moulins : édition critique de la Genèse d'après le manuscrit Bruxelles II 987*, thèse soutenue à l'Université de Dijon.en 2004

NOBEL, Pierre : « La glose dans le texte : l'exemple de Guiart des Moulins » in *Le texte dans le texte : qu'est-ce qu'une interpolation ?*, Annie Combes et Michelle SKLINIK éd, coll., Garnier, Garnier, Paris, 2013

PIGNATELLI, Cinzia et LAVRETIEV, Alexei, « Le Psautier d'Arundel : une nouvelle édition » in *XXVII<sup>e</sup> CILPR*, Université de Lorraine - ATILF, Nancy, 2013

PORTER, Dot : « Medievalists and the Scholarly Digital Edition. », in *Scholarly Editing : The Annual of the Association for Documentary Editing*, Volume 34, 2013 [En ligne]

Disponible sur : <http://www.scholarlyediting.org/2013/essay.porter.html>

**Stefania Cerrito**

Università degli studi internazionali di Roma - unint - Italie

## **Une édition électronique en xml-tei de l'*Ovide moralisé en prose* (version de Bruges)**

Dans le riche éventail d'outils informatiques qui sont désormais au service de l'éditeur de texte, le cadre méthodologique mis au point par la *Text Encoding Initiative* me paraît lui offrir un support solide et flexible, et ouvrir des perspectives d'un intérêt majeur. L'édition que j'achève de l'*Ovide moralisé en prose* (version de Bruges), pour laquelle je ne me servais au départ que d'instruments ordinaires de traitement de texte, s'est ainsi progressivement transformée, grâce au soutien qui m'est généreusement offert à l'Atilf, en une édition en XML-TEI.

Une tradition textuelle complexe, à cheval entre le manuscrit et l'imprimé, caractérise la mise en prose de l'*Ovide moralisé* qui vit le jour dans la Bourgogne des Ducs, texte qui n'avait jusqu'à aujourd'hui jamais fait l'objet d'une édition moderne. Si le ms. BnF fr.137 est sans aucun doute le meilleur témoin qui nous parvient, les deux autres manuscrits - Saint-Pétersbourg, Bn de Russie, F.v.XIV.1; Londres, BL Royal 17 E IV - et le magnifique incunable qui fut publié en 1484 par Colard Mansion, plus la traduction en anglais par William Caxton, constituent les étapes d'une chaîne de transmission du texte riche en variance.

Le choix d'une édition en format XML-TEI a ouvert de nouvelles perspectives de méthode et d'exploitation du texte. Un système de balisage du texte satisfaisant, mais susceptible d'être encore amélioré, a été mis au point, qui prévoit l'encodage des aspects codicologiques aussi bien que

philologiques. Les noms propres, les proverbes, les sources etc. sont balisés afin de confectionner de manière mécanique des index fiables; les lexies qui présentent un intérêt spécial pour le lexicologue (picardismes, premières attestations, néologismes sémantiques etc.) sont, toujours à l'aide d'une balise, mises au service de l'étude du vocabulaire.

Le format XML-TEI a permis également le recours à l'*Outil glossaire* du DMF conçu par Gilles Souvay. Cet instrument a tout d'abord été précieux pour la correction de l'édition, dont il a mis en évidence les fautes de transcription et les formes lexicales inconnues ou douteuses. Le lemmatiseur a offert une aide précieuse à l'étude du vocabulaire du texte édité, et à la construction d'un glossaire fiable; il a enfin permis de repérer aisément les éléments lexicaux encore absents du DMF, en mettant ainsi cette nouvelle édition au service de la lexicographie du moyen français.

Lors du CILPR2013, qui eut lieu à Nancy, je présentais un premier bilan provisoire de mon édition électronique et des résultats obtenus grâce aux outils informatiques utilisés. Ma communication se propose aujourd'hui de présenter un bilan final, aussi bien que les perspectives de recherche futures.

### **Bibliographie :**

- CERRITO Stefania, « L'*Ovide moralisé* in prosa (versione del ms. BnF, fr.137) dal manoscritto alle edizioni a stampa: l'episodio della Scilla di Megara » in *Quaderni di Lingue e Letterature Straniere* 33 (2008), pp.63-78.
- , « L'*Ovide moralisé* mis en prose à la cour de Bourgogne » in *Mettre en prose aux XIV<sup>e</sup>-XVI<sup>e</sup> siècles*. Approches linguistiques, philologiques, littéraires. Actes du III<sup>e</sup> Colloque International de l'AIEMF, Gargnano, 28-31 mai 2008, publiés par Maria COLOMBO TIMELLI, Barbara FERRARI et Anne SCHOYSMAN Turnhout : Brepols (Texte, Codex et Contexte), 2010, pp.109-117.
- , « À propos de la *Bible des poètes* » in *Le Moyen Français*, 69 (2011), pp.1-14.
- , « William Caxton traduttore di Ovidio », in *Dis/affinità elettive. Studi per Matilde De Pasquale*, a cura di Novella Novelli e Marino Freschi, ROMA: Empiria, 2013, pp. 43-56.
- , « Colard Mansion relit les *Métamorphoses*: une nouvelle version brugeoise de l'*Ovide moralisé* » in *Pour un nouveau répertoire des mises en prose. Roman, chanson de geste, autres genres*, sous la direction de Maria Colombo Timelli, Barbara Ferrari et Anne Schoysman, Paris: Garnier, 2014, pp.85-99.
- JUNG Marc-René, « *Ovide Metamorphose* en prose (Bruges, vers 1475) », in « *A l'heure encore de mon écrire. Aspects de la littérature de Bourgogne sous Philippe le Bon et Charles le Téméraire* ». Études rassemblées et présentées par Claude THIRY, *Lettres romanes*, numéro h. s., 1997, pp.99-115.
- PIERAZZO Elena et REHBEIN Malte, *TEI manuscripts SIG*  
<http://www.tei-c.org/SIG/Manuscripts/genetic.html>
- SOUVAY Gilles, « *LGeRM*: un outil de lemmatisation du moyen français », in David TROTTER (éd.), *Actes du XXIV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes*, vol. I, Tübingen: Niemeyer, pp.457-466.
- SOUVAY Gilles et PIERREL Jean-Marie (2010), « *LGeRM*: lemmatisation de mots en moyen français », in *Traitement Automatique des Langues*, t. 50 (2010) n° 2.
- SOUVAY Gilles et BAZIN Sylvie, « Construction assistée de glossaires à l'aide des outils du DMF », in *Actes du XXV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes* (Valence- Espagne).
- VAN EMDEN Wolfgang, « L'histoire de *Pyrame et Thisbé* dans la mise en prose de l'*Ovide moralisé* : texte du manuscrit Paris, B.N.F. fr. 137, avec variantes et commentaires », in *Romania*, 94 (1973), pp.29-56.
- VAN PRAET Joseph Basile Bernard, *Recherches sur Louis de Bruges, seigneur de la Gruthuyse ; suivies de la notice des manuscrits qui lui ont appartenu, et dont la plus grande partie se conserve à la Bibliothèque du roi*, Paris : De Bare frères, 1831.

## Quelques réflexions sur l'édition de *La Cité de Dieu* de saint Augustin traduite en français par Raoul de Presles (1371-1375)

Le projet Histoire du lexique politique français a entrepris en 2008 l'édition critique de la première traduction en français de *La Cité de Dieu* de saint Augustin par Raoul de Presle.

Au bout de cinq années de travail, seul le premier tome du premier volume (c'est à dire les livres I à III) est publié sur papier (pour autant, les livres IV et V sont en voie de finalisation et les livres VI à X sont bien avancés). Dans la mesure où le choix du projet a été d'encoder l'édition en TEI et dans la mesure aussi où ce choix d'encoder en TEI n'est pas le plus fréquent pour ce type d'édition, il nous semble légitime de se poser la question de la pertinence d'un tel choix.

Pour comparer ce qui est comparable, nous devons préciser que le projet produit (outre la version TEI qui sera partageable avec d'autres chercheurs) un prêt à cliquer ; autrement dit, l'éditeur commercial ne fournit aucun travail de mise en page. La comparaison entre l'encodage en TEI et d'autres solutions doit donc se placer dans la perspective où les éditeurs scientifiques réalisent, ou au moins spécifient entièrement, la mise en page complète de leur travail.

Plutôt que d'envisager les divers outils de mise en page, nous prendrons, pour illustrer notre propos, l'exemple des notes qu'on veut voir apparaître dans l'édition. Ces notes sont de trois sortes : a) les notes relatives à l'édition du texte ; b) les notes relatives aux variantes au texte ; c) les commentaires de nature historique ou philologique. Quel que soit l'outil envisagé, l'éditeur doit donc pour rendre compte d'une note :

- déterminer le segment de texte sur lequel porte la note
- définir si la note est une correction, une variante ou un commentaire
- fournir le contenu de la note.

D'une certaine façon, l'encodage en TEI n'est qu'une version plus sémantique de cette spécification. Par plus sémantique, nous entendons la différence qu'il y a entre dire qu'un segment doit être mis en italiques ou encoder la raison (l'interprétation) qui motive les italiques (mot étranger ou titre d'œuvre, ou segment en emphase, ou ...).

Dans la version complète, nous défendrons l'idée que le choix de la TEI se justifie par la qualité de l'édition qui en est obtenue mais aussi en termes d'économie de temps ; le cas particulier de ce projet est que les éditeurs ont dû apprendre le format TEI et que, néanmoins, sur un texte de cette ampleur, ce temps d'apprentissage arrive à se rentabiliser. Bien entendu, nous nous plaçons dans la perspective du contrôle entier de la forme obtenue par les éditeurs scientifiques (rien dans la forme n'est sous-traité à l'éditeur commercial). Nous nous contenterons par ailleurs de mentionner rapidement d'autres avantages que présente la forme TEI, notre propos visant uniquement à démontrer l'intérêt du format TEI pour qui ne souhaiterait diffuser que la version papier d'une édition critique.

### **Bibliographie :**

- Bertrand, Olivier, 2007. « Le vocabulaire politique aux 14<sup>e</sup> et 15<sup>e</sup> siècles : constitution d'un lexique ou émergence d'une science ? », dans *Lexiques scientifiques et techniques. Constitution et approche historique*, dir. Gerner Hiltrud, Stumpf Béatrice, Bertrand Olivier, Palaiseau, École polytechnique, 2007, pp. 9-23.
- Bertrand, Olivier, 2011. « *La Cité de Dieu* de saint Augustin traduite par Raoul de Presles (1371-1375) », édition du ms BnF, fr. 22912 sous la direction d'Olivier Bertrand, volume 1, tome 1, Paris, éd. Champion, coll. « Linguistique : traduction et terminologie », N°1, 2013, 984 p.
- Burghart, Marjorie, 2011. « Digital Editions as the Myth of Sisyphus », TEI Members Meeting and Conference, Würzburg, 2011, <[www.zde.uni-wuerzburg.de/tei\\_mm\\_2011/abstracts/abstracts\\_micropapers/#c249141](http://www.zde.uni-wuerzburg.de/tei_mm_2011/abstracts/abstracts_micropapers/#c249141)>.

Duval, Frédéric, 2006. « La Philologie française, pragmatique avant tout ? L'édition des textes médiévaux français en France », dans *Pratiques philologiques en Europe*, études réunies par Frédéric Duval, Paris, École des chartes (« Études et rencontres », 21), 2006, pp. 115-150.

Gaiffe, Bertrand / Stumpf, Béatrice, 2011. « A large scale critical edition : first translation of St Augustine's City of God by Raoul de Presle », TEI Members Meeting and Conference, Würzburg, 2011, <[www.zde.uni-wuerzburg.de/en/tei\\_mm\\_2011/abstracts/abstracts\\_papers/](http://www.zde.uni-wuerzburg.de/en/tei_mm_2011/abstracts/abstracts_papers/)>.

Gaiffe, Bertrand / Stumpf, Béatrice, 2013. « Une édition au format TEI de la première traduction française de *La Cité de Dieu* par Raoul de Presles (1371-1375) », XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes, 15-20 juillet 2013, Communication retenue pour publication dans les actes électroniques.

Kuntsmann, Pierre, 2000. « Édition électronique : le cas des *Miracles de Notre Dame par personnages*, dans *Le moyen français. Le traitement du texte (édition, appareil critique, glossaire, traitement électronique)*, Actes du IX<sup>e</sup> colloque international sur le moyen français, 29-31 mai 1997, Presses universitaires de Strasbourg, 2000, pp.115-122.

Souvay, Gilles / Pierrel, Jean-Marie, 2009. « LGeRM, Lemmatisation des mots en moyen français », *TAL (Traitement automatique des langues)*, 50, 149-172. <<http://halshs.archives-ouvertes.fr/halsbbhs-00396452>>.

## Michaela Bjuggfält-Châteaux Université Lille 3

### Prémices d'une édition<sup>1</sup> scientifique numérique de *Renaut de Montauban* en prose

La geste de *Renaut de Montauban* ou *Les Quatre fils Aymon* est l'une des plus productives du Moyen Âge. Les mises en prose, particulièrement nombreuses, sont cependant mal connues car pour la plupart encore sous forme de manuscrits non publiés. Pourtant ce sont elles qui sont à l'origine du développement dans toute l'Europe des légendes des *Quatre fils Aymon* et de leur cheval-faé Bayart. Étant donné que les chercheurs et lecteurs n'ont jusqu'à présent eu accès qu'aux versions traditionnelles en vers<sup>2</sup> ainsi qu'au cinquième volume, consacré à Mabrien, de la mise en prose aristocratique<sup>3</sup> – l'objectif premier de cette édition, débutée récemment dans le cadre d'une thèse de doctorat, est de donner accès à la partie dédiée aux Quatre fils Aymon de la version aristocratique en nous appuyant sur les trois témoins de cette mise en prose :

- Ms. (*Am*) Paris, Arsenal 5.072-5.075 complété par Munich, Bayerische Staatsbibl., Gall. 7 ;
- Ms. (*Lf*) Paris, BnF fr. 19.173-19.177 ;
- Ms. (*Pm*) Pommersfelden, Graflich Schönbornische Bibliothek 311-312.

Il a été choisi d'encoder les manuscrits en XML, en suivant les recommandations de la TEI, en raison des avantages d'interopérabilité et de pérennité qu'offre la structuration de textes dans ce langage de balisage.

L'étude en profondeur d'autres travaux de traitement de textes, notamment ceux de la BFM<sup>4</sup> et de l'Atilf<sup>5</sup>, nous aura permis de définir nos propres choix de balisage en vue d'une édition parallèle de deux témoins pour commencer, dans le but non seulement de donner accès au *Renaut de*

<sup>1</sup> Il s'agit d'un travail ayant pour objectif, outre la version papier, la réalisation d'une édition scientifique numérique structurée en XML de *Renaut de Montauban* en prose.

<sup>2</sup> Le manuscrit Douce *D*, édité par Jacques Thomas (1989), et du manuscrit La Vallière *L*, édité par Ferdinand Castets (1909).

<sup>3</sup> Le manuscrit BnF 19.177 édité par Philippe Verelst (1998).

<sup>4</sup> <http://bfm.ens-lyon.fr/>

<sup>5</sup> <http://www.atilf.fr/dmf/>

*Montauban* en prose, mais également de faire une comparaison fine de plusieurs versions d'un même texte et de contribuer, à notre niveau, à l'extraction de données et l'enrichissement des connaissances des faits linguistiques pour la période qui couvre le Moyen Français. Ainsi, notre démarche diffère des travaux de traitement de textes précédemment cités, en particulier celui de la *Queste del saint Graal* de la BFM, par la mise en parallèle de plusieurs témoins très proches d'un même texte.

Après une brève description de la phase d'acquisition du corpus et du processus de prétraitement des données, nous proposons de détailler, au moyen de plusieurs objets textuels précis, nos choix de balisage en appréciant les qualités et lacunes des différents choix pour les manipulations d'analyse du corpus et en argumentant ceux-ci, en particulier lorsqu'ils diffèrent des travaux de la BFM ou de l'Atilf évoqués. Enfin, nous partagerons les résultats déjà obtenus en concluant par les perspectives offertes par les premières analyses.

### **Bibliographie :**

#### **Éditions critiques**

CASTETS (Ferdinand), « La Chanson des Quatre fils Aymon », *d'après le manuscrit La Vallière*, Montpellier, Coulet, 1909.

THOMAS (Jacques), « Renaut de Montauban », *édition critique du manuscrit Douce*, Genève, Droz, 1989.

VERELST (Philippe), *Mabrien – Roman de chevalerie en prose du XV<sup>e</sup> siècle*, Genève, Droz, Romanica Gandensia, 28, 1998.

#### **Critique littéraire**

BAUELLE-MICHELS (Sarah), *Les avatars d'une chanson de geste : de « Renaut de Montauban » aux « Quatre Fils Aymon »*, Paris, Honoré Champion, 2006.

PIRON Maurice, « L'art des réécritures : de Maugis à Mabrien », Jean Dufournet (éd), *Si a parlé par moult ruiste vertu. Mélanges de littérature médiévale offerts à Jean Subrenat*, Paris, Champion, 2000 (Colloques, congrès et conférences sur le Moyen Âge), p. 455-465.

SUARD (François), « Le développement de la Geste de Montauban en France jusqu'à la fin du Moyen Âge », H.-E. Keller (éd) *Romance epic. Essays on a medieval literary genre*, Medieval Institute Publications, Studies in Medieval Culture, XXIV, Kalamazoo, 1987, p. 141-161.

THOMAS (Jacques), « Les mises en prose de Renaut de Montauban. Classement sommaire et sources », *Fin du Moyen Âge et Renaissance. Mélanges de philologie française offerts à Robert Guiette*, Anvers, Nederlandsche Boekhandel, 1961, p. 127-137.

VERELST (Philippe), « Renaut de Montauban, textes apparentés et versions étrangères : essai de bibliographie (nouvelle édition, revue et augmentée) », *Olifant*, 18, 1993, p. 94-190.

\_\_\_\_\_, « Renaut de Montauban et textes apparentés (Maugis d'Aigremont – Vivien de Monbranc – Mabrien) Recherches récentes (1990-2006) », *Le Souffle épique : l'Esprit de la chanson de geste : études en l'honneur de Bernard Guidot*, Sylvie Bazin-Tacchella, Damien de Carné et Muriel Ott (éds), Dijon, Éditions Universitaires de Dijon, 2011, p. 133-143.

\_\_\_\_\_, « Le cycle de Renaut de Montauban : aperçu général et réflexions sur sa constitution », B. Besamusca, W. P. Gerritsen et alii (éds) *Cyclification : the development of narrative cycles in the Chanson de geste and the Arthurian romances (Proceedings of the Colloquium on the development of narrative cycles, Amsterdam, 17-18 december, 1992)*, Amsterdam, 1994, p. 165-170.

#### **Édition de textes**

KUNSTMANN (Pierre), MARTINEAU (France) et FORGET (Danielle), *Ancien et Moyen Français sur le web. Enjeux méthodologiques et analyse du discours*, Ottawa, Éditions David, 2003.

KUNSTMANN (Pierre), « L'électronique à l'aide de l'éditeur : miracle ou mirage ? Bilan de quatorze années de travaux au LFA », *Perspectives médiévales* [En ligne], 34 | 2012, mis en ligne le 01 septembre 2012. Page consultée le 10 décembre 2013. URL : <http://peme.revues.org/2245>.

ROQUES (Gilles), « À propos d'éditions récentes de textes de Moyen Français. Problèmes et méthodes en lexicologie médiévale », dans Marc Wilmet (éd.), *Sémantique lexicale et sémantique grammaticale en Moyen Français, actes du Colloque organisé par le Centre d'Études Linguistiques et Littéraires de la Vrije Universiteit Brussel, Brussel, 28-29 septembre 1978*, Bruxelles, V.U.B. Centrum voor Taal- en Literatuurwetenschap, 1979, p. 3-21.

#### **Corpus**

GUILLOT (Céline), HEIDEN (Serge), LAVRENTIEV (Alexei) et MARCHELLO-NIZIA (Christiane), « Constitution et exploitation des corpus d'ancien et de moyen français », *Corpus*, 7, 2008. [En ligne] Mis en ligne le 13 novembre 2009, consulté le 19 novembre 2013. URL : <http://corpus.revues.org/1495>.

PRÉVOST (Sophie), « Corpus informatisés de français médiéval : contraintes sur leur constitution et spécificités de leurs apports », *Corpus*, 7, 2008, [En ligne] Mis en ligne le 13 novembre 2008, consulté le 17 janvier 2013. URL : <http://corpus.revues.org/1500>.

PRÉVOST (Sophie) et HEIDEN (Serge), « Etiquetage d'un corpus hétérogène de français médiéval : enjeux et modalités », dans *Romanistische Korpuslinguistik : Korpora und gesprochene Sprach / Romance Corpus Linguistics : Corpora and Spoken Language*, 1<sup>st</sup> Freiburg Workshop on Romance Corpus Linguistics, Freiburg, 2004, p. 127-136.

SERP (Claire), LAURENT (Anne), ROCHE (Mathieu) et TEISSEIRE (Maguelonne), « La quête du Graal et la réalité numérique », *Corpus*, 7, 2008. [En ligne] Mis en ligne le 12 novembre 2009, consulté le 17 janvier 2013. URL : <http://corpus.revues.org/1512>.

### Structuration et balisage informatiques

BISSON (Marie), « L'édition numérique structurée des Curieuses recherches du Mont Saint-Michel de dom Thomas Le Roy », *Schedae*, 2011, prépublication n° 12 (fascicule n° 1), p. 115-126.

HEIDEN (Serge), GUILLOT (Céline), LAVRENTIEV (Alexei) et BERTRAND (Lauranne), « Manuel d'encodage XML-TEI des textes de la Base de Français Médiéval », Projet BFM, UMR5191 ICAR, CNRS/ENS de Lyon, version 4.0, août 2010 [En ligne] Page consultée le 28 décembre 2013. URL : [http://bfm.ens-lyon.fr/IMG/pdf/Manuel\\_Encodage\\_TEI.pdf](http://bfm.ens-lyon.fr/IMG/pdf/Manuel_Encodage_TEI.pdf).

LAVRENTIEV (Alexei), « Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits », UMR5191 ICAR, CNRS/ENS-LSH, Lyon, version 2.1, juin 2008. [En ligne] Page consultée le 29 novembre 2013. URL : [http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss\\_Encodage-XML.pdf](http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf).

MARCHAL (Matthieu), *Gérard de Nevers : édition critique de la mise en prose du Roman de la Violette de Gerbert de Montreuil*, Thèse de doctorat en Langue et Littératures françaises sous la direction de Marie-Madeleine Castellani, Université Charles-de-Gaulle Lille 3, 2009.

—, « Structuration et balisage informatiques du glossaire de Gérard de Nevers, mise en prose du *Roman de la Violette* », [En ligne] <http://euro-lab.meshs.fr/documents/pdf/2010/Structuration-balisage.pdf>. Page consultée le 28 janvier 2013.

PASUPATHY (Subha-Sree), *Edition scientifique numérique (XML, TEI) de projets de l'Abbé de Saint-Pierre (ensemble de politique culturelle)*, thèse de doctorat sous la direction de Carole Dornier et Claudine Poulouin soutenue le 6 mars 2012 à l'Université de Caen, Caen, 2012.



Dominique Lapierre  
Université de Rouen

## Analyse linguistique automatique. Application au *Devisement du Monde* de Marco Polo

*Le Devisement du Monde* est une œuvre singulière où les voix de l'auteur, de l'écrivain, et des copistes successifs se sont entremêlées durant des siècles. Ce qui nous est donné à lire aujourd'hui, quelques cent quarante-trois manuscrits rédigés en treize langues différentes avec des titres qui leur sont propres, est moins le récit d'un individu - Marco Polo - que le résultat d'une accumulation de versions, copiées, traduites et remaniées depuis la rédaction du manuscrit original en 1298.

Cet article s'inscrit dans le domaine de l'application des outils de l'informatique linguistique à la recherche historique. Il concerne une mise en pratique de techniques statistiques pour l'étude d'occurrences de motifs linguistiques dans *Le Devisement du Monde* de Marco Polo. Nous présenterons notamment des exemples de mise en œuvre d'outils d'analyse automatique.

L'un des enjeux de ce travail de recherche est d'acquérir de nouveaux instruments de mesure quantitative et qualitative dans le but d'élaborer des statistiques thématiques, d'extraire des données et de déterminer l'origine commune de différentes versions du texte polien afin d'en faciliter la classification.

Développés en France depuis les années 1950 les outils d'analyse du discours ont considérablement évolué et trouvé des champs d'application variés au sein des sciences sociales. Leur usage s'avère particulièrement efficace non seulement pour évaluer la richesse lexicale ainsi que les caractéristiques d'un texte mais également pour proposer une analyse qualitative des textes grâce à de nouveaux outils statistiques. Les avantages principaux de ces derniers sont leur capacité à mémoriser de grandes quantités de données, et leur vitesse de calcul. Cependant, ces logiciels ne se substituent pas à l'analyse humaine. Il revient naturellement au chercheur de choisir le corpus sur lequel il souhaite fonder son analyse, d'encoder les données ainsi que d'interpréter les résultats.

L'interprétation du récit de Marco Polo se fait à partir d'un appareil méthodologique mettant en valeur la diversité religieuse et culturelle dans le récit. Il s'agit d'établir des connexions des lexiques *a priori* et *a posteriori* à partir d'une liste d'item afin de repérer les éléments saillants de chaque version. En fonction de la présence ou de l'absence de certains éléments, il est possible de générer une cartographie des différentes structures du texte. Le lexique religieux employé par Marco Polo relève de la liturgie, des pratiques et des croyances. Les mœurs sont étudiées selon leur importance du point de vue religieux (comme l'anthropophagie, la famille, ou la sexualité). Nous nous intéressons aux rapports entre religion et pouvoir royal (à partir des objets, reliques, monuments...) ainsi qu'aux rituels. Enfin une attention particulière est donnée à la place de l'astrologie et des rites funéraires dans le récit.

À ce jour, sept versions complètes du *Devisement du Monde* ont été numérisées et plusieurs outils d'analyse automatique utilisés en textométrie ont pu être testés. Parmi eux IRAMUTEQ, TROPES et TXM. Cet article rend compte de leurs apports et de leurs limites dans ce travail de recherche.

### Sources

- (F) : ms 1116 de la Bibliothèque Nationale de France.  
(FG) : ms fr. 5631 de la Bibliothèque Nationale de France.  
(TA<sup>1</sup>) : Magliabechiano II.IV.88 della biblioteca Nazionale di Firenze.  
(TA<sup>2</sup>) : II.IV.136 della biblioteca Nazionale di Firenze.  
(VA<sup>3</sup>) : ms CM 211 della biblioteca civica de Padova.  
(P) : B.N Vienna, B.Pal. 3273.  
(Z) : 49, 20 della Biblioteca Catedral de Toledo.

### Éditions

- (R) : Ramusio (Giovani Batista), *Navigazioni e viaggi*, vol. III, 1553.  
(G) : Groulneau (Estienne), *La description géographique des provinces & villes les plus fameuses de l'Inde orientale, meurs, loix, & coutumes des habitans d'icelles, mesment de ce qui est soubz la domination du grand Cham Empereur des Tartars, par Marc Paule gentilhomme Venetien, et nouvellement réduict en vulgaire françois*, éd. E. Groulneau, Paris, 1556.  
(M) : Müller (Eugène), *Deux voyages en Asie au XIIIe siècle par Guillaume de Rubruquis, envoyé de Saint Louis et Marco Polo, marchand Vénitien* ; éd. E. Müller, Delagrave, Paris, 1888.  
(Y) : Yule (Henry), *The Travels of Marco Polo, The Complete Yule-Cordier Edition*, vol.1, Dover Publications, New York, 1993.  
(FG) Ménard (Philippe), *Le Devisement du Monde*, Edition critique en six volumes, Droz, 2001-2009.

### Références

- Benzécri (Jean-Paul), *L'analyse des données*, vol.2 : Correspondances, Paris, Dunod, 1973.  
Heiden (Serge), Magué (Jean-Philippe), et Pincemin (Bénédicte), "TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement", in *JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data*, Rome, Italie.  
Marchand (Pascal), Ratinaud (Pierre), L'analyse de similitude appliquée aux corpus textuels, *lexicometrica.univ-paris3.fr*, septembre-octobre 2011.  
Mayaffre (Damon), *De la lexicométrie à la logométrie* [hal.archivesouvertes.fr/docs/00/55/.../18\\_Mayaffre\\_Astrolabe\\_2005.pdf](http://hal.archivesouvertes.fr/docs/00/55/.../18_Mayaffre_Astrolabe_2005.pdf)  
Molette (Pierre), "De l'Analyse Propositionnelle du Discours à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste", Communication au colloque Psychologie et communication de Tarbes, juin 2009.  
Ratinaud, (Pierre), Iramuteq : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. [www.iramuteq.org](http://www.iramuteq.org)  
Tehrani (Jamie), The Phylogeny of Little Red Riding Hood, *PLoS ONE* 8(11): e78871, 2013.

## Actes Médiévaux des Princes Lorrains : vers un corpus numérisé

L'objectif du projet AMPLorr, engagé depuis 2013 dans le double cadre de la MSH-Lorraine et d'un partenariat entre l'ATILF et l'Atelier diplomatique du CRULH, est de mettre à disposition de la communauté scientifique les actes numérisés des princes lorrains des XIII<sup>e</sup> et XIV<sup>e</sup> siècles (transcriptions et photographies). À terme, l'objectif est d'apporter une contribution bien spécifique aux recherches sur la pratique de l'écrit aux derniers siècles du Moyen Âge, en le situant au croisement de deux champs d'étude : d'une part le champ de **la gouvernance des États naissants à l'échelon régional** (espace lotharingien entre France et Empire) et d'autre part le domaine des **transferts culturels entre régions situées de part et d'autre de la frontière linguistique**.

Dans cette perspective, l'Atelier diplomatique propose, en partenariat avec l'ATILF, de procéder à un encodage XML réfléchi des textes, afin que l'historien puisse mener sur le corpus qu'ils composent des interrogations croisées autour des noms de personnes, de lieux ou encore à partir de mots-clefs (ex. « sel », « forêt », « abbaye »...). Il s'agit d'offrir aux chercheurs un outil privilégié au service d'enquêtes sérielles de grande ampleur dans les domaines de la prosopographie et de la définition des espaces socio-politiques lorrains au Moyen Âge : étude du gouvernement du prince dans la longue durée du règne, établissement de ses itinéraires, de ses réseaux de pouvoir, études des transferts culturels, gestion de son domaine.

Parallèlement, par leur lemmatisation, ces mêmes textes viendront enrichir la base du Dictionnaire du Moyen Français et ouvriront leur champ au questionnement des linguistes.

Le règne du duc Ferry III (1251-1303) constitue la première étape de ce projet. Le traitement des 1 500 actes produits au cours de cette période doit permettre de poser d'abord les enjeux, de mettre en place les méthodes et d'imaginer des solutions aux difficultés (scientifiques et techniques) posées par la composition multilingue du corpus global. En effet, l'une des caractéristiques de la production diplomatique des princes lorrains est son trilinguisme : moyen français, latin, moyen allemand. Cinq cents actes en moyen français sont d'ores et déjà transcrits, encodés et prêts à être mis en ligne.

Dans le cadre d'un projet ANR-FNR (TRANSSCRIPT) qui démarrera début 2015, le projet AMPLorr rejoindra une entreprise similaire menée par les médiévistes de l'Université du Luxembourg sur les actes des comtes-ducs pour la même période (mi-XIII<sup>e</sup>-fin XIV<sup>e</sup> siècle). Le corpus luxembourgeois est lui aussi trilingue.

À terme, nous souhaiterions pouvoir mener des interrogations simultanées sur les deux corpus (traités comme un seul), y retrouver la trace de personnages qui ont circulé et agi à l'intérieur d'un même espace régional, par-delà les « frontières » politiques et linguistiques. Ce projet vise également à cartographier l'emploi respectif ou cumulatif des langues dans l'espace mais aussi à pouvoir déceler les évolutions de la pratique linguistique sur la durée. Enfin, nous souhaitons contribuer à la valorisation patrimoniale des fonds d'archives en permettant au grand public d'accéder aux documents et de comprendre leur contexte de production.

## Du différentialisme au corpus : l'informatique à l'aide de Christine de Pizan

La communication proposée se situerait d'une part dans le cadre de la linguistique descriptive, d'autre part dans celui des recherches linguistiques sur le genre. Elle se donnerait pour tâche d'étudier certains aspects de la syntaxe de la prose de Christine de Pizan : le corpus serait constitué surtout par le texte du *Livre des epistres du debat sus le Rommant de la Rose* (Valentini 2014) et, comme textes de contrôle, par le *Livre de la cité des dames* (Parussa 2010), le *Livre de l'advison Cristine* (Towner 1932) et le *Livre des trois vertus* (Willard-Hicks 1989), dont il existe des versions numériques.

Certains aspects de la syntaxe analysables par l'outil informatique y seraient particulièrement étudiés : la subordination (modes verbaux, propositions infinitives, propositions hypothétiques, ruptures de construction) et l'ordre des constituants dans la proposition. Les résultats seraient confrontés surtout avec les modèles décrits dans les grandes synthèses constituées par Martin-Wilmet 1980 et Marchello-Nizia 1997<sup>2</sup>, ainsi qu'avec des études plus particulières (ex. Marchello-Nizia 1995, 9-113). (Le nombre de points abordés dans la communication orale dépendrait aussi du temps alloué à chaque communication ; certains points pourraient être mieux développés dans la version écrite.)

Cette étude permettrait sans doute d'apporter quelques pièces d'appui ainsi qu'en moindre mesure, quelques rectifications à des descriptions déjà existantes de certains points de la syntaxe du moyen français, dont la langue de Christine de Pizan constitue un exemple paradigmatique, l'écrivaine ayant vécu à Paris au début du XVe siècle. Parallèlement, elle voudrait montrer comment une étude systématique menée sur un corpus précis exploité informatiquement permet d'infirmer certaines affirmations qui semblent être dictées par des préjugés selon lesquels la langue de Pizan serait différente de celles des autres auteurs de son temps en tant que langue d'une femme : ex. Zink 1995, Kennedy 1998, xviii-xl. Ce type de thèses peuvent être rattachées à une vision différentialiste des genres sexués, selon laquelle la langue des femmes et celle des hommes auraient chacune des caractères propres « par nature ». L'approche retenue, en revanche, considère que la langue est avant tout un des aspects de la construction sociale : Christine de Pizan faisant partie d'une certaine société à une époque donnée, sa langue répondait plutôt à des impératifs d'ordre systématique d'une part (le moyen français pouvant être décrit comme un état de langue régi par sa structure propre), et d'ordre littéraire d'autre part (les contraintes de la prose argumentative).

### **Bibliographie :**

- Kennedy 1998 : Angus J. KENNEDY (éd.), Christine de Pizan, *Le Livre du corps de policie*, Champion, 1998.
- Marchello-Nizia 1995 : Christiane MARCHELLO-NIZIA, *L'évolution du français. Ordre des mots, démonstratifs, accent tonique*, Paris, Armand Colin, 1995.
- Marchello-Nizia 1997<sup>2</sup> : Christiane MARCHELLO-NIZIA, *La Langue française aux XIV<sup>e</sup> et XV<sup>e</sup> siècles*, 2<sup>e</sup> éd., Paris, Nathan, 1997 (1<sup>re</sup> éd. 1979).
- Martin-Wilmet 1980 : Robert MARTIN et Marc WILMET, *Syntaxe du moyen français*, Bordeaux, SOBODI, 1980.
- Parussa 2010 : transcription par Gabriella PARUSSA du *Livre de la cité des dames* à la page internet <http://www.pizan.lib.ed.ac.uk/xmlfinal/cdam.xml>, à vérifier dans Patrizia CARAFFI (trad. en italien) et Earl Jeffrey RICHARDS (éd.), Christine de Pizan, *La Città delle dame*, Milan, Luni, 1997, réimpr. Rome, Carocci, 2003.
- Towner 1932: Sister Mary Louis TOWNER, *L'avison-Christine*, Washington, D.C., The Catholic University of America Press, 1932, réimpr. New York, AMS Press, 1969, version électronique dans les *Classiques Garnier Numériques en Ligne*, à vérifier dans Christine RENO et Liliane DULAC (éd.), Christine de Pizan, *Le Livre de l'advison Cristine*, Paris, Champion, 2001.
- Valentini 2014 : Andrea VALENTINI (éd.), Christine de Pizan, *Le Livre des epistres du debat sus le Romman de la Rose*, Paris, Classiques Garnier, 2014, l'étude linguistique aux p. 35-106.

Willard-Hicks 1989 : Charity Cannon WILLARD (éd.) et Eric HICKS (avec la collaboration), *Le Livre des trois vertus*, Paris, Champion, 1989, version électronique dans les *Classiques Garnier Numériques en Ligne*.  
Zink 1995 : Gaston ZINK, « La phrase de Christine de Pizan dans le *Livre du corps de policie* », dans *Une femme de lettres au Moyen Âge. Études autour de Christine de Pizan*, dir. Liliane Dulac et Bernard Ribémont, Orléans, Paradigme, 1995, p. 383-395.

Geoffrey Roger  
University of London Institute in Paris

## Les *scriptae* régionales du moyen français : état des lieux

Sous sa forme écrite, le moyen français a souvent été décrit comme largement standardisé d'après l'usage « francien » ou parisien (Marchello-Nizia 1997 ; Greimas et Keane 1992 ; Price 1998) en dépit des régionalismes ou dialectalismes qui ont été signalés dans de nombreux documents de cette période (Pope 1934 ; Dees 1985 ; Buridant 2000 ; Lusignan 2012). En l'absence d'aucune étude détaillée des variations phono- et morpho-graphématiques en vigueur, les éditions modernes de textes du moyen français ne se prononcent que rarement sur leur localisation à partir de critères linguistiques (Greub 2003). Celles qui s'y aventurent feignent souvent d'ignorer la pénurie d'informations sur le sujet, et se réfèrent par défaut aux études de l'ancien français, ce qui n'est pas sans poser de risque.

Comme l'ont répété Dees (1985), Völker (2007) et Lusignan (2012), il est urgent et indispensable de procéder à une étude quantitative des variations graphématiques au sein des textes des XIV<sup>e</sup> et XV<sup>e</sup> siècles, et à terme, de produire un atlas linguistique cartographiant leur distribution à travers la francophonie médiévale. Des projets ponctuels allant dans ce sens ont déjà été entrepris (Grübl 2013), mais étant donné l'ampleur du corpus il faudra soit faire preuve d'une extrême patience, soit se limiter à des sondages localisés à la manière de Gilliéron et Edmont (1902-1910).

Dans cette attente, la question se pose de savoir dans quelle mesure nous pourrions mettre à profit le vaste corpus numérique de moyen français d'ores et déjà explorable en ligne (*DMF*, *BFM*, *AND*, *Online Froissart*, *Christine de Pizan Project*). Si le *DMF* permet d'explorer facilement les variations graphématiques sur un lemme donné, il ne fournit encore que trop rarement des renseignements précis sur la localisation de ses sources ; des informations sont parfois disponibles ailleurs (*FEW*, *DÉAF*, éditions modernes), mais souvent trop laconiques ou opaques pour être pleinement utiles. La communication proposera une méthodologie pour sonder le corpus du *DMF*, en s'appuyant sur des exemples de graphèmes régionaux détectés lors de recherches sur la localisation du ms. des *Cent nouvelles nouvelles* (Roger 2011). Ces indices peuvent contribuer à orienter la localisation d'autres sources de provenance non-enseignée, aussi bien qu'à pondérer des localisations pré-établies sur critères lexicaux et extra-linguistiques.

### **Bibliographie :**

- K. Baldinger, *Dictionnaire étymologique de l'ancien français*, Québec, 1971-.
- C. Buridant, *Grammaire nouvelle de l'ancien français*. Paris, 2000.
- A. Dees, *Eléments constitutifs du Moyen Français*. V<sup>e</sup> Colloque International sur le Moyen Français, Università Cattolica del Sacro Cuore, 1985, p. 7-19.
- J. Gilliéron et E. Edmont, *Atlas linguistique de la France*, Paris, 1902-1910.
- A. J. Greimas et T. M. Keane, *Dictionnaire du moyen français: la Renaissance*, Paris, 1992.
- Y. Greub, *Les Mots régionaux dans les farces françaises*, Strasbourg, 2003.
- K. Grübl, *Varietätenkontakt und Standardisierung im mittelalterlichen Französisch. Theorie, Forschungsgeschichte und Untersuchung eines Urkundenkorpus aus Beauvais (1241–1455)*, Narr, Romanica Monacensia, 83, Tübingue, 2013.
- S. Lusignan, F. Martineau, Y.-Ch. Morin, P. Cohen, *L'introuvable unité du français. Contacts et variations linguistiques en Europe et en Amérique (XII<sup>e</sup>-XVIII<sup>e</sup> siècle)*, Québec, 2012.
- C. Marchello-Nizia, *La Langue française aux XIV<sup>e</sup> et XV<sup>e</sup> siècles*, Paris, 1997 (1979).
- M. K. Pope, *From Latin to Modern French with especial Consideration of Anglo-Norman*, Manchester, 1934.
- G. Price, "French", *Encyclopedia of the Languages of Europe*. Oxford, 1998.
- G. Roger, « *Les Cent Nouvelles Nouvelles : A Linguistic Study of MS Glasgow Hunter 252* », University of Glasgow, 2011 (<http://theses.gla.ac.uk/2872/>).
- H. Völker, « A "Practice of the Variant" and the Origins of the Standard. Presentation of a Variationist Linguistics Method for a Corpus of Old French Charters », *French Language Studies*, t.17 (2007), p. 207-223.
- W. von Wartburg, *Französisches Etymologisches Wörterbuch*, Bonn, 1932-.
- Liens :
- Anglo-Norman Dictionary* - <http://www.anglo-norman.net/>
- Base de français médiéval*- <http://bfm.ens-lyon.fr/>
- Christine de Pizan Project* - <http://www.pizan.lib.ed.ac.uk/context.html>
- Dictionnaire du moyen français (1330-1500)* - <http://www.atilf.fr/dmf/>
- Online Froissart* - <http://www.hrionline.ac.uk/onlinefroissart/>

Pierre Kunstmann, Université d'Ottawa.  
Gilles Souvay, ATILF-CNRS, Université de Lorraine

## **D'une phase à l'autre, le Dictionnaire Électronique de Chrétien de Troyes (DÉCT) : achèvement et élargissement**

Notre communication se veut une mise à jour du point sur nos travaux que nous avons fait, il y a un an et demi, lors du CILPR XXVII organisé à Nancy par l'ATILF. Dans le premier quart d'heure, Pierre Kunstmann présentera le dictionnaire : passé, présent et avenir ; dans le deuxième quart d'heure, Gilles Souvay fera une démonstration des fonctionnalités du DÉCT dans sa version de janvier 2015, qui constitue l'état achevé de la phase 1. Les quinze dernières minutes seront réservées aux questions de l'auditoire : c'est, en effet, la première fois que nous nous adresserons directement (mais voir Kunstmann 2012) à un public aussi spécialisé (les littéraires, philologues et linguistes de la SLLMOO) qui devrait logiquement représenter les utilisateurs les plus assidus et les plus critiques de l'instrument que nous offrons sur la Toile. Sans constituer véritablement un wiki-dictionnaire, le DÉCT est un travail d'équipe : il résulte de la collaboration d'un groupe de

chercheurs, qui, au moment de commencer la seconde phase (*DÉCT2*), aimeraient bénéficier de l'aide et des observations que peuvent leur apporter les spécialistes de notre discipline.

L'équipe qui a œuvré à la première phase (*DÉCT1*) était constituée de cinq personnes : Pierre Kunstmann (Laboratoire de Français Ancien, Université d'Ottawa), qui en a assumé la direction et a rédigé les articles ; Hiltrud Gerner (ATILF, CNRS Université de Lorraine) et May Plouzeau (Université de Provence), qui ont révisé les articles ; Ineke Hardy (LFA, Université d'Ottawa) pour la correction de la version anglaise. Gilles Souvay (ATILF, CNRS Université de Lorraine) s'est chargé des développements informatiques. Le *DÉCT1* se limitait à la définition des mots lexicaux (adjectifs qualificatifs, adverbes se terminant en *-ment*, substantifs, verbes) dans la copie du scribe Guiot (ms. BnF fr. 794), un des meilleurs manuscrits et surtout le seul qui présente le texte de tous les romans.

La seconde phase est particulièrement ambitieuse : elle consiste à introduire non seulement les mots grammaticaux (quelque 553 lemmes, si l'on inclut les locutions, d'après les listes établies par Ollier 1986), mais aussi les axes de synonymie et d'antonymie avec des renvois réciproques d'un lexème, ou d'une lexie, à l'autre (sur la façon de détecter similitudes et oppositions voir Kunstmann 2011) et, *last but not least*, les variantes des autres manuscrits (variations lexicales ; variations également dans la collocation des termes – ces variations sont nombreuses, mais impossibles à chiffrer pour l'instant), passant, pour ainsi dire, du travail de Guiot à l'œuvre de Chrétien.

Nous travaillons résolument dans les trois axes retenus pour ce colloque : le *DÉCT* est une édition électronique, qui constitue un corpus ; c'est une base textuelle, qui interroge ce corpus ; c'est une étude de ce corpus par l'établissement d'un lexique complet.

#### **Bibliographie :**

- Kunstmann Pierre (directeur du projet), Gerner Hiltrud, Hardy Ineke, Plouzeau May, Souvay, Gilles, *Dictionnaire Électronique de Chrétien de Troyes (DÉCT 1)*, nouvelle édition, Nancy, ATILF, janvier 2015. [www.atilf.fr/dect/](http://www.atilf.fr/dect/)
- Kunstmann Pierre, Souvay Gilles, « Dictionnaire Électronique de Chrétien de Troyes : fin et suite... » XXVII<sup>e</sup> Congrès International de Linguistique et Philologie Romanes, Actes électroniques, Section 16, n° 15.
- Kunstmann Pierre, « L'électronique à l'aide de l'éditeur : miracle ou mirage ? Bilan de quatorze années de travaux au LFA », *Perspectives médiévales* [En ligne], 34 | 2012. URL : <http://odel.revues.org/10/peme/2245>
- Kunstmann Pierre, « Le dictionnaire électronique au centre du travail d'édition des textes : quelques réflexions sur les rôles et rapports de l'éditeur et du lexicographe », dans D. Trotter (éd.) *Present and future research in Anglo-Norman* : Aberystwyth Colloquium, July 2011, Aberystwyth, The Anglo-Norman Hub, 2012, p. 159-162.
- Kunstmann Pierre, « Synonymie et antonymie : des propositions de Robert Martin à la préparation de la seconde phase du Dictionnaire Électronique de Chrétien de Troyes », dans F. Duval (éd.) *La « logique » du sens, Autour des propositions de Robert Martin*, Metz, Université Paul Verlaine (Recherches Linguistiques n° 32), 2011, p. 301-317.
- Kunstmann Pierre, Gerner Hiltrud, Souvay Gilles, « Dictionnaire électronique de Chrétien de Troyes : état actuel du projet », dans M. Iliescu, H. Siller-Runggaldier, P. Danler (éds) : *Actes du XXV<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes*, Berlin / New York, Walter de Gruyter, 2010, vol. 2, p. 185-192.
- Kunstmann Pierre, « Vers un dictionnaire électronique de Chrétien de Troyes », *Qui tant savoit d'engin et d'art*, *Mélanges de philologie médiévale offerts à Gabriel Bianciotto*, Poitiers, CÉSCM, 2006, p. 311-319.
- Kunstmann Pierre, Souvay Gilles, « Le DÉCT (Dictionnaire Électronique de Chrétien de Troyes) : un modèle pour la lexicographie d'aujourd'hui ? », 13<sup>e</sup> Congrès Euralex, Barcelone, juillet 2008, publication électronique sur le site [www.euralex.org/euralex/conferences/](http://www.euralex.org/euralex/conferences/)
- Kunstmann Pierre, Souvay Gilles, « Dictionnaire Électronique de Chrétien de Troyes (DÉCT) », dans P. Kunstmann et A. Stein (éds) *Le Nouveau Corpus d'Amsterdam. Actes de l'atelier de Lauterbad, 23--26 février 2006*, Stuttgart, Steiner, 2007, p. 111-120.
- Ollier Marie-Louise, *Lexique et concordance de Chrétien de Troyes*, Montréal, Université de Montréal / Paris, Vrin, 1986.

**Réunir une base de données textuelles et la lexicographie historique :  
les *Plus anciens documents linguistiques de la France*  
et le *Dictionnaire étymologique de l'ancien français***

Le projet présenté par la suite réunit deux entreprises lexicologiques et philologiques de longue haleine: le *Dictionnaire Etymologique de l'Ancien français* (DEAF), rédigé depuis quatre décennies à Heidelberg, et les *Plus anciens documents linguistiques de la France* (DocLing), poursuivies également depuis les années 1970.

Le projet « DocLing-DEAF » poursuit le but d'intégrer les données lexicales des *DocLing* dans les 'articles courts' du DEAF, c'est-à-dire dans le DEAFpré. Cela permet de renforcer la part des sources documentaires dans le DEAF, mais aussi de créer une interface entre ce dictionnaire et une base textuelle de référence. En même temps, le vocabulaire des *DocLing* serait placé dans le cadre naturel de l'ancienne langue française par une harmonisation des lemmes.

L'idée est donc d'une grande simplicité, mais sa réalisation suppose une harmonisation très précise entre la structure lexicologique des deux entreprises et la création d'interfaces informatiques performantes.

Nous présenterons les structures lexicologiques et informatiques des deux bases de données en question. Nous préciserons les interfaces concrètes entre les deux bases de données (lemme, sigle-DEAF, forme lexématique, datation et localisation, source, contexte, lien avec l'édition électronique des actes) et les interfaces pour les utilisateurs. En conclusion, nous réfléchirons sur la plus-value de cette alliance pour les deux entreprises en question et pour les études sur l'ancienne langue française ainsi que sur son impact sur la méthodologie en linguistique historique et sur la linguistique de corpus.

Très concrètement, le projet commun permettra entre autres de pallier à la faible présence des textes documentaires dans les études sur l'ancienne langue; d'un point de vue méthodologique, il pourra rompre le cloisonnement qui régit encore aujourd'hui les grandes entreprises de la recherche.

Finalement, nous demandons si un tel projet pourrait ensuite ouvrir la voie vers d'autres dictionnaires et corpus textuels, comme l'AND ou le *Nouveau Corpus d'Amsterdam* (NCA).

**Bibliographie :**

AND = Rothwell, William / Stone, L. W. / Reid, T. B. W., with the assistance of D. Evans, S. Gregory, David A. Trotter, P. Staniforth, *Anglo-Norman dictionary*, London: The Mod. Humanities Research Assoc., 1977-1992. Version AND2, Rothwell, William / Gregory, S. / Trotter, David A. et al., fasc. A-C et D-E, London (Maney/MHRA) 2005; version électronique ANDEL: <<http://www.anglo-norman.net>>.

DEAF = Städtler, Thomas / Dörr, Stephen / Tittel, Sabine / Möhren, Frankwalt (fondé par Baldinger, Kurt), *Dictionnaire étymologique de l'ancien français*, Québec, Les Presses de l'Univ. Laval, Tübingen, Niemeyer, 1971-; version électronique DEAFél: <<http://www.deaf-page.de>>.

DocLing = *Les plus anciens documents linguistiques de la France. Édition électronique*, Collection fondée par Jacques Monfrin, poursuivie par M.-D. Glessgen, en collaboration aujourd'hui avec Frédéric Duval et Paul Videsott, <[www.mediaevistik.uzh.ch/docling](http://www.mediaevistik.uzh.ch/docling)>, 2009 [nouvelle édition, 2012].

Glessgen, Martin-D., « Le long chemin d'une charte vers l'ère multimédiale », in: Schweickard, Wolfgang / Overbeck, Anja / Völker, Harald (éds.): *Lexikon, Varietät, Philologie*, Romanistische Studien. Günter Holtus zum 65. Geburtstag. Berlin, de Gruyter, 2011, 649-664.

- / Alletsgruber, Julia, « Structure de la base de données lexicologique des *Plus anciens documents* », in: BEC 168 (2010) [2011], 83-94

- / Kihäi, Dumitru / Videsott, Paul, « L'élaboration philologique et linguistique des 'Plus anciens documents linguistiques de la France, Édition électronique' », in: *Bibliothèque de l'École des Chartes* 168 (2010) [2011], 5-94.

NCA = Stein, Achim / Kunstmann, Pierre / Gleßgen, Martin-D. (edd.): *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français* (ca 1150-1350), établi par Anthonij Dees (Amsterdam 1987), remanié par A. S., P. K. et M.-D. G., Stuttgart: Institut für Linguistik/Romanistik, version 2-2, 2010 (<<http://www.uni-stuttgart.de/lingrom/stein/corpus/#nca>>).

Tittel, Sabine, « Le « DEAF électronique » – un avenir pour la lexicographie », in: *Revue de Linguistique Romane* 74 (2010) 301-311.

## Textes médicaux du moyen âge : corpus numérique et base de données morphologique plurilingue

Cette communication a pour objectif de présenter un projet de recherche qui s'appuie sur des corpus numérisés en collaboration avec l'équipe du DMF, et qui étudie pourquoi certains néologismes créés dans le domaine médical au cours du Moyen Âge parviennent à se maintenir jusqu'en français moderne, tandis que d'autres disparaissent après un certain temps. Notre hypothèse est que la morphologie, et plus en particulier la transparence constructionnelle, a joué un rôle crucial dans la préservation de ce lexique: concrètement, des termes qui présentent une relation formelle proche de l'élément latin dont ils sont issus se maintiennent mieux que des créations françaises originales, c'est-à-dire des dérivations ou des compositions réalisées à partir de bases morphologiques françaises. En tant que tels, ces premiers jetteraient les bases de la technique de composition néoclassique toujours utilisée aujourd'hui.

Le projet veut examiner les facteurs qui sont décisifs pour la survie des néologismes dans le domaine des sciences. Nous étudions d'un côté des facteurs externes, tels que le succès du texte, le prestige de l'auteur, la présence ou non de gloses explicatives, la constance dans l'utilisation du terme par l'auteur, la fréquence du terme dans les textes, etc. (cf. Goyens – Van Tricht à par.). Par ailleurs, nous analysons aussi l'impact des caractéristiques morphologiques de ces néologismes: même si certaines études mentionnent l'importance des latinismes ou des termes latins francisés (e.g. Ducos 1998), nous voudrions soumettre le vocabulaire scientifique médiéval à une étude morphologique approfondie.

Notre étude sera basée sur un corpus de textes médicaux du Moyen Âge, des traductions du latin ou des textes immédiatement composés en français, pour lesquels nous développons une version électronique et lemmatisée, en collaboration avec le DMF2012 (ATILF), ainsi qu'une base de données de termes médicaux (latins et français) et des morphèmes qui les composent (bases, affixes, latins et français). Ces données seront complétées grâce aux lemmes du dictionnaire du français scientifique médiéval de Créalscience, notamment du point de vue des informations sémantiques. Ensuite, ces éléments seront analysés du point de vue des critères externes mentionnés, et du point de vue de critères internes, comme la fréquence ("type" et "token") du terme et des critères morphologiques. Des analyses statistiques devraient enfin nous permettre d'isoler les critères qui jouent le rôle le plus significatif dans la survie des termes, et de vérifier si notre hypothèse qu'une forme proche du latin a la meilleure chance de se maintenir, est correcte.

Cette communication veut davantage explorer les possibilités qu'offre la collaboration avec l'Atilf et l'équipe du DMF et l'emploi de corpus numérisés en fonction des objectifs du projet, notamment l'analyse morphologique des formes des textes ainsi que le développement d'une base de données morphologique; enfin, dans le cas de traductions en moyen français de textes source latins, nous étudions les possibilités d'une analyse conjuguée de corpus parallèles en moyen français et en latin.

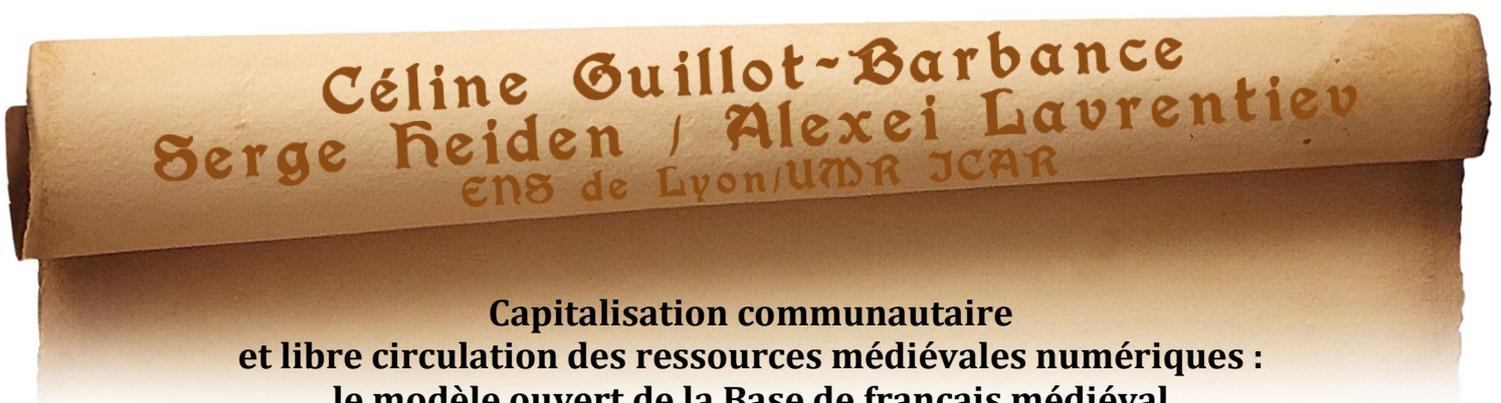
### **Bibliographie :**

- Amiot, D. & Van Goethem, K. 2012. 'A constructional account of French *-clé* 'key' and Dutch *sleutel-* 'key' as in *mot-clé / sleutelwoord* 'keyword'. *Morphology* 22. 347-364.
- Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Cottez, H. 1980. *Dictionnaire des structures du vocabulaire savant: éléments et modèles de formation*. Paris.
- Ducos, J. 1998. *La météorologie en français au Moyen Âge (XIIIe - XIVe siècle)*. Paris.
- Fradin, B. 2003. *Nouvelles approches en morphologie*. Paris : Presses Universitaires de France.
- Goyens, M. 2013. 'Le sort des néologismes dans la langue des sciences au Moyen Âge : une question de morphologie?'. *Neologica*, 7, 41-56.

Goyens, M. & I. Van Tricht. sous presse. 'Albathe face à pustule : disparition versus lexicalisation des néologismes en français médiéval'. C. Badiou-Monferran & T. Verjans eds. *Disparitions. Contributions à l'étude du changement linguistique*. («Linguistique historique»). Paris : Champion.

Lusignan, S. 1989. 'La topique de la *translatio studii* et les traductions françaises de textes savants au XIV<sup>e</sup> siècle'. G. Contamine ed. *Traduction et traducteurs au Moyen Âge*. Paris. 303-315.

Villoing, F. 2012. 'French compounds'. *Probus* 24. 29-60.

A scroll with names and affiliation. The text on the scroll is: Céline Guillot-Barbance, Serge Heiden / Alexei Laurentiev, ENS de Lyon/UMR JCAR.

Céline Guillot-Barbance  
Serge Heiden / Alexei Laurentiev  
ENS de Lyon/UMR JCAR

### **Capitalisation communautaire et libre circulation des ressources médiévales numériques : le modèle ouvert de la Base de français médiéval**

La *Base de français médiéval* (BFM<sup>1</sup>) est l'un des corpus numériques de français médiéval les plus importants (9<sup>e</sup>-15<sup>e</sup> siècles, 126 textes, 3 550 000 mots). Accessible en ligne *via* un portail TXM (<http://txm.bfm-corpus.org>), ce corpus se compose d'éditions de référence (éditions originales et éditions imprimées numérisées), encodées au format XML-TEI et enrichies à de multiples niveaux : métadonnées décrivant les textes, codage interne aux textes et segmentation graphique (*tokenisation*), étiquetage morphosyntaxique, annotation syntaxique, encodage du discours direct. Ce corpus déjà ancien (25 ans) s'est développé ces dernières années grâce à différentes initiatives collectives de recherche (programmes ANR notamment).

L'ensemble des ressources qui composent la BFM sont librement accessibles, réutilisables et rediffusables sous licence *Creative Commons* (BY-NC-SA). Ces ressources comprennent naturellement le texte de base des éditions, mais aussi les fichiers sources enrichis grâce aux métadonnées, aux balises de codage et de segmentation graphique, aux étiquettes morphosyntaxiques, etc. Le système de métadonnées créé pour la BFM, les normes de codage XML-TEI et le schéma de spécification ODD, le jeu d'étiquettes morphosyntaxiques spécialisé (Cattex2009) et ses principes d'application, le modèle morphosyntaxique (pour le logiciel TreeTagger) et le lexique de l'ensemble de la base sont également accessibles dans les mêmes conditions<sup>2</sup>. Enfin, la plateforme TXM (Heiden *et al.* 2010), qui permet la lecture, l'interrogation et l'analyse des textes de la base, est téléchargeable gratuitement<sup>3</sup> et adaptable librement (*open-source*). Le mode ouvert du développement et de la diffusion des textes de la BFM et de son outil d'analyse TXM offrent la possibilité de vérifier et de reproduire de façon aussi précise que nécessaire toutes les études réalisées à partir d'eux, ce qui garantit la scientificité de leur démarche empirique.

Les enjeux scientifiques liés au développement de la BFM, son environnement technologique et juridique ont permis la mise en place d'un espace de travail collaboratif et cumulatif, les ressources et leur appareillage étant accessibles aux collègues et utilisateurs pour l'ajout de nouvelles couches d'information. Les choix en matière de formats et de standards de codage et d'annotation rendent

possible la circulation, la bonne interprétation et la réutilisation de ces ressources. Les modes de citation permis par le numérique garantissent l'identification des auteurs et des contributeurs successifs.

Notre présentation donnera un état synthétique des ressources actuellement disponibles dans la BFM et détaillera la façon dont elles ont été constituées (échanges de textes, de jeux d'étiquettes, procédures d'étiquetage et d'annotation manuelle, procédures d'étiquetage et d'annotation par apprentissage, etc.). Nous aborderons en particulier l'évolution des conditions d'accès à la BFM et les problèmes de droits d'auteur concernant les éditions de sources médiévales. Seront également présentés les projets de développement de la base dans la perspective de libre circulation des ressources numériques.

### **Bibliographie :**

- Guillot, C., Lavrentiev, A. & Marchello-Nizia, C. (2007) « Les corpus de français médiéval : état des lieux et perspectives », *Revue française de linguistique appliquée* 121, 125-128.
- Guillot, C., Lavrentiev, A., Rainsford, T., Marchello-Nizia, C., Heiden, S. (à par. 2014) « La "philologie numérique" : tentative de définition d'un nouvel objet éditorial ». In É. Buchi, J.-P. Chauveau, J.-M. Pierrel, (éd.). *Actes du XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013)*, 3 volumes. Strasbourg : Société de linguistique romane/ÉLiPhi.
- Heiden, S., Magué, J.-P. and Pincemin, B. (2010) « TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement ». In S. Bolasco, I. Chiari and L. Giuliano (éd.). *Statistical Analysis of Textual Data - Proceedings of 10th International Conference JADT 2010, Rome, 9-11 juin 2010*.
- Pincemin, B., Guillot, C., Heiden, S. & Lavrentiev, A. (2008) « Usages linguistiques de la textométrie. Analyse qualitative de la consultation de la Base de Français Médiéval via le logiciel Weblex », *Syntaxe & sémantique* 9, 87-110.
- Rainsford, T. Guillot, C., Lavrentiev, A. & Prévost, S. (2012) « La zone préverbale en ancien français : apport des corpus annotés ». *Actes du 3<sup>e</sup> Congrès Mondial de Linguistique Française*, EDP Sciences ([www.linguistiquefrancaise.org](http://www.linguistiquefrancaise.org)), 159-176, [<http://dx.doi.org/10.1051/shsconf/20120100246>].

1 Le site du projet BFM (<http://bfm.ens-lyon.fr>) présente la base dans son état actuel et ses objectifs de recherche.

2 Ces ressources sont publiquement accessibles sur le site de la BFM ([http://bfm.ens-lyon.fr/rubrique.php3?id\\_rubrique=112](http://bfm.ens-lyon.fr/rubrique.php3?id_rubrique=112)).

3 <http://textometrie.ens-lyon.fr>.