



4^e réunion du
Consortium pour les Corpus de Français Médiéval

ATILF – Nancy, les 6 et 7 octobre 2008



Sylvie BAZIN-TACCHELLA

Projet d'édition de la traduction française de la *Chirurgia Magna* de Guy de Chauliac (XV^e siècle)

Le texte latin de Chauliac (1363), véritable monument de la chirurgie de la fin du Moyen Âge, est accessible aujourd'hui dans l'édition de M. McVaugh (*Inventarium sive Chirurgia Magna*, 1997) ; malheureusement pour le texte français très souvent cité par les historiens, on ne possède que l'édition complète de Nicaise (1890), qui a suivi et réécrit le texte de l'édition de Joubert (1580) en citant de temps à autre soit le texte latin, soit le texte d'un des manuscrits, mais sans véritable étude critique.

L'édition projetée a pour but essentiel d'établir un texte accessible et sûr, fidèle au mode d'exposition scientifique de l'époque, grâce à l'examen des différents témoins manuscrits et imprimés anciens qui offrent un large éventail de variantes de traduction (appartenant à trois versions différentes), souvent intéressantes sur le plan de la formation du vocabulaire scientifique et la diffusion concrète des idées, la confrontation systématique avec le texte latin permettant d'élucider nombre de difficultés et de comprendre les processus à l'oeuvre dans la traduction.

La première édition imprimée de ce texte, l'édition de Nicolas Panis (1478), faisait d'ailleurs partie du corpus de référence pour l'établissement du *Lexique scientifique du Moyen Français*, publié par les soins de C. Thomasset et D. Jacquart, sous l'égide de l'INALF (aujourd'hui ATILF), alors dirigé par R. Martin. Chargée de la saisie partielle de cet incunable, j'avais été frappée par l'allure très latinisante de la traduction, qui se sépare de ce point de vue assez nettement du manuscrit de Nicolas de Blanchecourt que j'édite (Paris, BnF, fr. 24249). Si le maintien partiel du latin caractérise la traduction à destination professionnelle, la confrontation des différentes versions françaises fait apparaître dans ce domaine des différences notables.

J'ai présenté dans le cadre du dossier d'HDR (2004) l'édition critique du *Chapitre Singulier* et des trois premiers traités de la traduction française, avec introduction et notes. Il me reste à poursuivre l'établissement du texte déjà collationné (traités 4 à 7), d'harmoniser l'ensemble et d'en préparer la publication. Si le texte des traités 4 à 7 est déjà saisi, il reste un travail considérable de confrontation des manuscrits et d'élucidation des difficultés, nombreuses dans cette partie de l'ouvrage.

La refonte de mon glossaire sur l'ensemble du texte se fera en suivant les normes de lemmatisation et de saisie de l'ATILF, de manière à le rendre directement exploitable par le DMF. Il s'agira de corriger, compléter et élargir le glossaire existant, en remplaçant notamment ses apports dans la tradition lexicographique (FEW, Tobler-Lommatzsch, Godefroy, TLF) et par rapport à d'autres travaux dans le même champ (en particulier G. Sigurs et S. Tittel). Le texte français de Chauliac pourrait ainsi devenir accessible à la fois dans une édition papier, conforme aux principes traditionnels de l'édition de texte médiéval, et dans la base textuelle du DMF.

La comparaison des différents états du texte soumis à la vulgarisation offre un extraordinaire terrain d'investigation pour la formation du vocabulaire et des structures mêmes du texte d'exposition scientifique. Le fait que l'oeuvre ait été lue, éditée, commentée jusqu'au XVIII^e siècle permet en outre de cerner les permanences et les évolutions lentes du discours chirurgical dans la longue durée, avant la révolution clinique.

Olivier BERTRAND

Programme ERC :

Edition critique de la première traduction en français de la *Cité de Dieu*.

Le programme de recherche propose l'édition critique de la première traduction en français par Raoul de Presles de *La Cité de Dieu* de saint Augustin (5 volumes contenant les 22 livres de l'oeuvre) ainsi qu'une large étude scientifique sur l'aristotélisme et l'augustinisme au moyen âge. L'édition de texte permettra à la communauté scientifique l'accès à la langue du moyen français (14e s.) et à la réception de l'antiquité philosophique et théologique au moyen âge. Les autres publications liées à ce large projet d'édition scientifique offriront une meilleure compréhension de notre connaissance du vocabulaire politique actuel en français dont une partie provient de cette période (14e / 15e s.).

Edition critique

Les traductions en moyen français des textes de science politique (grecs et latins) de l'antiquité ne sont que partiellement éditées. C'est pourtant par elles que se constitue une grande partie du lexique politique que nous utilisons en français aujourd'hui. Un des manques criants est une édition moderne de la première traduction en français (fin 14e siècle) de *La Cité de Dieu* de saint Augustin par Raoul de Presles.

Le texte de *La Cité de Dieu* est un ouvrage fondamental dans l'histoire des idées en Occident. Sa traduction en français, accompagnée de la glose du *translateur*, offre un champ d'investigation scientifique majeur en matière de lexicologie historique.

Le programme ERC entend proposer une édition de texte moderne de la première traduction de *La Cité de Dieu* en Français par Raoul de Presles (1371-1375). Cette édition aura pour manuscrit de base un exemplaire de la Bibliothèque Nationale de France (FR 22912 et FR 22913, BnF, département des manuscrits occidentaux). L'ouvrage se compose de 22 livres, eux-mêmes divisés en de nombreux chapitres. L'entreprise d'édition est donc monumentale :

- Etape 1 : premier volume (Introduction, Livres 1-3 et appareil critique) [années 1 et 2]
- Etape 2 : deuxième volume (Livres 4-9 et appareil critique) [année 2 et 3]
- Etape 3 : troisième volume (Livres 10-14 et appareil critique) [années 2 et 3]
- Etape 4 : quatrième volume (Livres 15-18 et appareil critique) [années 3 et 4]
- Etape 5 : cinquième volume (Livres 19-22 et appareil critique) [années 4 et 5]

Martin-Dietrich GLESSGEN, Harald VÖLKER

La linguistique variationnelle empirique et l'incorporation des données métatextuelles

Lors des dernières années, les méthodologies philologique et scriptologique ont fait l'objet d'un débat concernant les présupposés théoriques de l'analyse linguistique. Ce débat a mené, entre autres, à un élargissement des objectifs poursuivis, traditionnellement limités à l'étude de la diatopie, vers d'autres catégories diasystématiques.

La mise en oeuvre concrète de cette redéfinition et réorientation méthodologique nous oblige à améliorer nos données brutes à deux niveaux : au niveau des données textuelles (qui implique la problématique des critères de la transcription et de l'édition) et au niveau des données métatextuelles dont la vocation est celle de décrire les textes et de les positionner dans le diasystème. Notre contribution vise ce dernier domaine et se propose d'illustrer de quelle manière les 'métadonnées' peuvent répondre aux nouvelles exigences introduites par la linguistique variationnelle empirique.

Céline GUILLOT, Alexei LAVRENTIEV

**"Principes de codage et de description des données numériques :
état des lieux dans le projet BFM"**

L'objet de cette intervention est de présenter de manière succincte et synthétique l'état d'avancement de l'équipe de l'ENS LSH (Lyon), du point de vue technique et technologique, dans différents domaines :

- identifiant des œuvres et des textes : présentation des pratiques lyonnaises et des échanges en cours avec l'équipe du DEAF (Heidelberg) ;
- genres et domaines textuels : présentation de la liste actuelle des domaines et des genres et compte-rendu des discussions/échanges sur ces notions au sein du projet CORPTEF soutenu par l'ANR (Corpus représentatif des premiers textes français) ;
- proposition de codage des dates (conformément à ce qui avait été demandé lors de la dernière réunion du CCFM).

Richard INGHAM

**Grammaire et phonologie de l'anglo-normand tardif:
L'apport du Corpus de Correspondance Anglo-Normande**

Le corpus de correspondance anglo-normande est un petit corpus de 30.000 mots environ, comprenant une soixantaine de lettres écrites en français insulaire entre 1260-1350, pour la plupart par des ecclésiastiques. Il permet de mieux connaître les caractéristiques de l'anglo-normand tardif à une époque où il tendait à assumer, à côté du latin, certaines fonctions administratives. Ont été retenues seulement des lettres de particulier à particulier. Celles écrites soit à l'intention du roi, soit par sa chancellerie, ont été écartées, à cause de l'excès de formalité stylistique susceptible de s'y trouver.

Le corpus, créé en 2007, a été partiellement étiqueté. Parmi les domaines ciblés sont ceux où l'anglo-normand (l'AN) se distinguait du français du continent, par exemple:

- L'expression formelle de l'accord en genre
- L'emploi des formes fortes des pronoms personnels par rapport aux formes atones
- L'inflexion infinitivale *-er/-ir/oir* etc.

Notre exposé portera sur le thème suivant: dans quelle mesure les graphies du corpus permettent-elles de saisir l'effet qu'ont eu sur la grammaire de l'AN les modifications phonologiques survenues depuis la conquête normande?

Pierre KUNSTMANN, Gilles SOUVAY

La navigation à partir du DÉCT: principe et fonctionnalités

Le Dictionnaire électronique de Chrétien de Troyes est fondé sur le principe de la navigation entre les documents qu'il présente. C'est sans doute ce qui constitue son originalité foncière. Cette navigation s'effectue intérieurement et extérieurement.

À l'interne, elle permet, à partir d'un passage de la transcription de la copie du scribe Guiot, d'obtenir l'image (folio) qui lui correspond dans le manuscrit. En cliquant sur n'importe quel mot de cette transcription, on en trouve le lemme et le code grammatical. Un nouveau clic sur le lemme permet de rejoindre l'article qui lui est consacré dans le lexique. Dans les articles, tous les voyages sont possibles, du court déplacement (recherche plein texte dans l'article, par ex.) au long périple (tel que la recherche d'un syntagme ou d'une famille de mots dans l'ensemble du lexique). Le dictionnaire étant établi à partir d'une base textuelle entièrement codée lexicalement, l'exploration peut se faire aussi dans les textes des romans à l'aide d'une fonction portant sur la cooccurrence d'éléments (lemmes, formes graphiques, catégories grammaticales).

À l'externe, l'utilisateur plongé dans le lexique peut atteindre, dans la rubrique dictionnaire figurant sous la vedette, certains ouvrages où l'entrée retenue est également traitée ; par un clic de sa souris, il peut arriver ainsi aux pages du Godefroy, du DMF, du TLFi et bientôt de l'AND. Il peut aussi, dans le corps d'un article, voir surlignés certains noms ou titres ; en cliquant dessus, il est alors dirigé, par un lien URL sous-jacent, vers un ancrage dans les documents en question. Ce qui lui permet, par exemple, de dériver du DÉCT et d'aborder tel ou tel chapitre de *Perceval Approches*, le cours d'ancien français de May Plouzeau, placé sur le site LFA.

Dans ce travail en cours (la lettre "D" étant maintenant terminée, 32% des articles sont déjà rédigés), que nous espérons mener à bon port, nous restons bien sûr, par principe, ouverts à toute nouvelle possibilité de navigation.

James LAIDLAW

The Making of the Queen's Manuscript

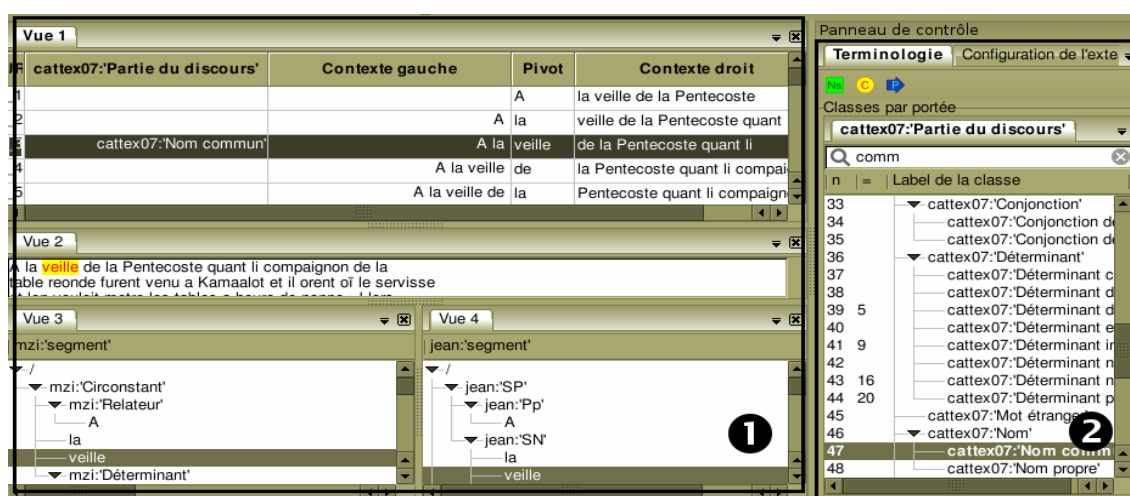
Commandé à Christine de Pizan par Isabeau de Bavière vers 1413, le ms. Harley 4431 mérite sa place parmi les 'Treasures of the British Library'. C'est un codex exceptionnel sur les plans artistique, codicologique, linguistique et littéraire. Afin de créer un recueil digne de la reine, Christine sélectionne trente ouvrages en vers et en prose qu'elle fait illustrer par des artistes de premier ordre, dont le Maître de la Cité des Dames. Elle surveille la préparation du manuscrit en neuf fascicules, et elle corrige les textes. Est-ce Christine qui copie le manuscrit en entier, comme l'affirment certains? Tous les fascicules datent-ils de 1413, ou est-ce que Christine profite de l'occasion pour y incorporer un *Epistre Othea* déjà copié et illustré pour un autre mécène?

Cet exposé présentera le schéma XML que développe notre équipe internationale pour tâcher de répondre à ces questions, en éditant et glosant les trente textes 'autographes' de Christine.

Nicolas MAZZIOTTA

Logiciel NotaBene pour l'annotation linguistique Visualisations et conceptualisations multiples

Le logiciel NotaBene, dont nous avons commencé le développement de janvier à juin 2008 au laboratoire ICAR (Université Lumière Lyon 2), a été conçu pour permettre l'annotation de documents XML de la BFM. D'un point de vue technique, le logiciel a été écrit en python et utilise les standards RDF (*Resource Description Framework*) et OWL (*Web Ontology Language*) pour gérer les annotations et la formalisation de leur sémantique. On peut aborder le logiciel à l'aide de la capture d'écran suivante:



Du point de vue de l'utilisateur, NotaBene consiste en un espace de travail (zone 1) permettant de visualiser le même texte de plusieurs manières différentes (par exemple, sous forme de texte courant, de concordance ou d'arborescence syntaxique) et d'y apposer des annotations d'ordre divers. Il est ainsi possible d'afficher de la manière la plus appropriée les informations nécessaires à la tâche de l'annotateur.

Cet espace d'annotation est complété par un espace de modélisation (zone 2), qui met à la disposition des utilisateurs un éditeur d'ontologie simple. Dans cet espace, les concepts et les relations hiérarchiques qu'ils entretiennent entre eux peuvent être définis ou modifiés à la demande.

Téléchargement : <<https://sourceforge.net/projects/notabene/>>

Sophie PRÉVOST, Achim STEIN

Syntactic Reference Corpus of Medieval French (SRCMF)

The two largest machine-readable text corpora of Medieval French have been built in two independent, though cooperating, projects in France (ENS-LSH Lyon: ICAR, UMR 5191) and Germany (ILR, University of Stuttgart). The central theme of this project is the syntactic annotation of both corpora in a joint project, so that researchers can base their research on both resources, according to shared principles.

The project will use and, if necessary, develop tools for automatic and manual annotation on both morphology and syntax levels. It will keep tools and resources as system-independent as possible and thus ensure their reusability in future research.

The project will be an important contribution to closing the gap between English and French medieval corpus resources and constitute an important impact for future work in diachronic syntax. The collaborating scholars world-wide will embed the annotation project in their diachronic linguistic research on various syntactic subjects.

The added value of the German-French cooperation over two national projects is the simultaneous coordinated work on two different corpora according to shared standards. Only an intense well-structured cooperation between both groups will provide a resource on such a scale and having such an extensive chronological span (7 centuries).

The collaborating scholars will embed the annotation project in their diachronic linguistic research on various syntactic subjects, and the henceforth possibility to process such a large syntactically enriched corpus will enable to discover many new evolution processes and to verify existing theories.

Natasha ROMANOVA

Online Froissart Project: principes d'édition et d'encodage

Le texte des *Chroniques* de Froissart est un texte très long dont les quatre *Livres* nous sont parvenus en plus de 150 manuscrits médiévaux datant de la période entre la fin du 14^{ème} jusqu'à la fin du 15^{ème} siècle.

Le projet *Online Froissart*, lancé en 2007 par une équipe de chercheurs à Sheffield et à Liverpool, est en train de créer une ressource en ligne qui mettra à la disposition de la communauté savante et du grand public des images digitales de cinq des plus anciens manuscrits des *Livres I-III*, des transcriptions complètes de ces témoins, collationnées et accompagnées par une traduction partielle, des commentaires et un engin de recherche. Par ailleurs, des transcriptions d'une sélection de passages de chaque manuscrit existant seront aussi disponibles.

Dans ce contexte, compte tenu du volume et du nombre de témoins, cette édition ne sera ni une édition diplomatique ni une édition critique. Le projet se propose de présenter le texte comme il est conservé par chaque témoin avec de minimales interventions éditoriales (les graphies manifestement erronées ou des mots omis sont corrigés et les corrections signalées) mais sans reproduire la réalité du manuscrit (les abréviations sont résolues et ne sont pas signalées).

Cette intervention sera centrée sur les principes d'encodage adoptés par le projet pour traiter ce vaste corpus et sur les possibilités que le *Online Froissart* ouvrira pour la recherche dans de différentes disciplines.

Lene SCHOESLER

Le besoin de descripteurs illustré par des cas d'étude

Depuis le mois d'août, un petit groupe informel de recherche sur la négation composé de France Martineau, Harald Völker et moi-même, essaye de mettre au pied une grille d'analyse qui combine des informations syntaxiques et des informations diasystématiques. Cette grille sera utilisée pour l'analyse de la négation et des particules de renforcement de la négation (ne + âme, goutte, mie, mot, pas, point ...), mais la négation n'est qu'un exemple; nous croyons que la grille sera utile pour l'étude d'autres phénomènes. Le but de ma contribution est de présenter et d'illustrer cette grille, tout en espérant que cette présentation sera utile pour la discussion ultérieure des besoins de descripteurs.

Actuellement, nos réflexions se sont concrétisées de façon à distinguer trois niveaux :

I. Distinction selon le type de négation / le contexte syntaxique / sémantique autour de *ne*

II. Distinction selon le contexte syntaxique et sémantique plus large (type de verbe, type de phrase, mode, diathèse, etc.)

III. Variables diasystématiques (1) Genre textuel /tradition de discours, 2) diatopie, 3) diachronie : datation, 4) diastratie : niveau social, etc.)

Pour ma présentation, je me servirai d'exemples de négation provenant des manuscrits du Charroi de Nîmes pour illustrer la grille.

Thomas STÄDTLER

L'apport du DEAF aux descripteurs

Faisant suite aux idées et réflexions présentées par Frankwalt Möhren lors des dernières rencontres CCFM et CORPTEF, nous avons continué surtout deux travaux pour contribuer au développement des descripteurs:

- la mise à point des sigles du Complément bibliographique munis du signe ÷, c'est-à-dire de ceux qui n'ont pas encore trouvé leur forme définitive
- le balisage des informations données dans le Complément bibliographique concernant les domaines et les genres.

Dans cette intervention, on fera le point sur l'état de ces travaux en cours, pour en pouvoir discuter les perspectives.

Willy STUMPF

La datation des textes du corpus DMF

Lors de l'établissement de la bibliographie des textes du corpus, il a fallu attribuer une date à chacun d'entre eux, ce renseignement faisant partie intégrante de l'abréviation des références. D'autre part, les exemples sont classés dans l'ordre chronologique à l'intérieur de chaque paragraphe des lexiques.

Nous présenterons brièvement les critères qui ont été appliqués pour dater les textes, ainsi que les difficultés que nous avons rencontrées dans ce travail.

David TROTTER

La typologie géolinguistique/scriptologique et le problème des textes à multiple origine

Il existe plusieurs cas de figure en ce qui concerne la localisation et (pour le travail du CCFM) les « descripteurs liés à la dimension spatiale/dialectale » [B].

Nous distinguons implicitement entre documents « administratifs » et textes « littéraires ». Les premiers existent d'habitude en une seule version, souvent datée et localisée selon des critères extérieurs ; il y a parfois divergence entre date de l'événement et document, parfois aussi (et plus souvent) entre endroit concerné et lieu de production du document, peut-être aussi entre ce dernier et le dialecte du copiste. Les textes littéraires par contre offrent d'habitude des divergences plus grandes et souvent incontrôlables entre tous ces éléments, la problématique étant d'ailleurs compliquée par l'existence de plusieurs copies d'un même texte.

Il y a également une catégorie de textes, dont il existe des exemples dans les deux groupes susmentionnés (littéraire ~ administratif) : les textes que j'appelle « à multiple origine », c'est-à-dire où en plus d'une différence scriptologique/dialectale entre texte et copie, le manuscrit lui-même témoigne d'une origine mixte. La communication prendra quelques exemples (littéraires et non-littéraires) de ce phénomène dans le but d'un examen de la fonctionnalité des descripteurs géolinguistiques pour les textes de ce type. Pour un dictionnaire de l'anglo-normand (donc : différentiel) ce problème est particulièrement compliqué car il soulève souvent la question de l'appartenance d'un texte à l'anglo-normand.

Liste des participants

Bazin Sylvie (Université Nancy2)	sylvie.bazin@univ-nancy2.fr
Bertrand Olivier (Polytechnique Paris)	olivier.bertrand@polytechnique.fr
Del Mancino William (Nancy)	wdelmanc@atilf.fr
Gerner Hiltrud (Nancy)	hiltrud.gerner@atilf.fr
Glessgen Martin-Dietrich (Zurich)	glessgen@rom.uzh.ch
Guillot Céline (ENS Lyon)	celine.guillot@ens-lsh.fr
Ingham Richard (Birmingham)	ringham04@yahoo.co.uk
Kunstmann Pierre (Ottawa)	pierrekunstmann@sympatico.ca
Laidlaw James (Edimbourg)	james.laidlaw@ed.ac.uk
Lavrentiev Alexei (ENS Lyon)	alexei.lavrentev@ens-lsh.fr
Lodge Anthony (Saint-Andrews)	rall@st-and.ac.uk
Mansfield Charly (Edimbourg)	charlie.mansfield@ed.ac.uk
Mazziotta Nicolas (Liège)	nicolas.mazziotta@ulg.ac.be
Murray Sarah Jane (Waco, Texas)	sj_murray@baylor.edu
Prévost Sophie (Paris)	sophie.prevost@ens.fr
Romanova Natasha (Liverpool)	n.romanova@liverpool.ac.uk
Schøsler Lene (Copenhague)	schoesl@hum.ku.dk
Souvay Gilles (Nancy)	gilles.souvay@atilf.fr
Städler Thomas (Heidelberg)	thomas.staedtler@urz.uni-heidelberg.de
Stein Achim (Stuttgart)	achim.stein@ling.uni-stuttgart.de
Stumpf Willy (Nancy)	stumpf@atilf.fr
Trotter David (Aberystwyth)	dt@aber.ac.uk
Völker Harald (Zurich)	hvoelker@rom.uzh.ch

