

REVUE DE
LINGUISTIQUE
FRANÇAISE
DIACHRONIQUE

7
2017

DIACHRONIQUES

LES ÉTATS ANCIENS
DES LANGUES À L'HEURE
DU NUMÉRIQUE

Pérez Rodríguez – 979-10-231-2160-5



LES ÉTATS ANCIENS DES LANGUES À L'HEURE DU NUMÉRIQUE

JOËLLE DUCOS

Présentation

ROBERT MARTIN

À propos du *DMF* : réussites et pièges de la lexicographie électronique

SYLVIE BAZIN-TACHELLA & GILLES SOUVAY

De la gestion de la variation en moyen français à son élargissement aux états anciens du français : les développements du lemmatiseur LGeRM

XAVIER-LAURENT SALVADOR, FABRICE ISSAC & MARCO FASCIOLA

Herméneutique des similarités dans le *DFSM* : une expérience

ESTRELLA PÉREZ RODRÍGUEZ

Le *Lexicon Latinitatis Medii Aevi Regni Legionis* (VIII^e siècle-1230) : caractéristiques et quelques exemples (*ventrescas, iera, cumbo, plentum*)

ELISA GUADAGNINI

La lexicographie de l'italien médiéval et les corpus de l'OVI : un bilan provisoire et quelques nouvelles perspectives

ANA GÓMEZ RABAL

Le latin médiéval du *Glossarium Mediae Latinitatis Cataloniae* : un projet lexicographique dans un contexte européen

MICHÈLE GOYENS & CÉLINE SZECEL

Autorité du latin et transparence constructionnelle : le sort des néologismes médiévaux dans le domaine médical

CÉLINE GUILLOT, SERGE HEIDEN & ALEXEI LAVRENTIEV

Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique

GÉRARD PETIT

Terminographie diachronique : le cas de la terminologie médiévale française

RAMON MASÍÀ

Numérisation et traitement de textes mathématiques grecs : méthodes, problèmes et résultats

EARL JEFFREY RICHARDS

À la recherche des communautés discursives au Moyen Âge : un regard numérique sur la connectivité dans la culture vernaculaire et le rôle des traductions dans l'évolution de la prose en moyen français



LES ÉTATS ANCIENS DES LANGUES
À L'HEURE DU NUMÉRIQUE

Les états anciens
des langues
à l'heure du numérique



Les PUPS, désormais SUP, sont un service général
de la faculté des Lettres de Sorbonne Université.

© Presses de l'université Paris-Sorbonne, 2018

© Sorbonne Université Presses, 2021

Diachroniques n° 7

ISBN papier : 979-10-231-0581-0

PDF complet – 979-10-231-2155-1

TIRÉS À PART EN PDF :

Ducos – 979-10-231-2156-8

Martin – 979-10-231-2157-5

Bazin-Tacchella & Souvay – 979-10-231-2158-2

Salvador, Issac & Fasciolo – 979-10-231-2159-9

Pérez Rodríguez – 979-10-231-2160-5

Guadagnini – 979-10-231-2161-2

Gómez Rabal – 979-10-231-2162-9

Goyens & Szecel – 979-10-231-2163-6

Guillot, Heiden & Lavrentiev – 979-10-231-2164-3

Petit – 979-10-231-2165-0

Masià – 979-10-231-2166-7

Richards – 979-10-231-2167-4

Maquette initiale : Compo-Méca (64990 Mouguerre)

Réalisation : Emmanuel Marc Dubois/3d2s

SUP

Maison de la Recherche

Sorbonne Université

28, rue Serpente

75006 Paris

Tél. (33) 01 53 10 57 60

sup@sorbonne-universite.fr

sup.sorbonne-universite.fr

Le *Lexicon Latinitatis Medii Aevi Regni Legionis* (VIII^e siècle-1230) : caractéristiques et quelques exemples (*ventrescas, iera, cumbo, plentum*)

Estrella Pérez Rodríguez
Université de Valladolid

Une grande part de notre connaissance du passé nous a été léguée par l'intermédiaire des mots. C'est à travers des textes écrits à des époques particulières que nous pouvons connaître de nombreux faits qui sont survenus alors, et le Moyen Âge ne fait pas exception à cette règle de fait. Mais, pour que les textes constituent des témoins fiables de l'époque de leur rédaction, encore faut-il les comprendre de manière appropriée. Un instrument peut, sans nul doute, aider leur lecteur contemporain au regard de cette exigence : il lui faut disposer d'un dictionnaire de la langue dans laquelle ces textes sont rédigés.

En 1995, le professeur Maurilio Pérez (Université de León) a décidé d'entreprendre la dure et immense tâche d'élaborer un tel dictionnaire : le *Lexicon Latinitatis Medii Aevi Regni Legionis*, ou *LELMAL*¹, projet que je codirige actuellement. La première remise de ce travail a été publiée par la maison d'édition Brepols dans sa collection « Corpus Christianorum ». Il s'agit encore d'un *Lexicon Imperfectum*, car il comprend en l'état 3 020 entrées lexicographiques, ce qui représente sans doute le tiers, à peu près, de l'ensemble lexical établi par nos sources.

1. À la fin du présent article, le lecteur voudra bien trouver la liste de toutes les abréviations utilisées, ainsi qu'une bibliographie complète. Ce travail s'inscrit dans le cadre des projets FFI2015-64340-P (MINECO/FEDER) et VA027U14 (JCyL).

Cette contribution répond à un double objectif : en premier lieu, je souhaite exposer les caractéristiques fondamentales de ce dictionnaire. En second lieu, je développerai et commenterai quelques exemples intéressants qui démontrent l'importance de l'étude lexicographique afin de mieux connaître l'histoire de la langue d'un territoire donné.

Caractéristiques du dictionnaire

Sources et principes du travail lexicographique

Le *LELMAL* s'élabore actuellement à partir d'un corpus de sources fermé. Ce corpus est formé par les textes écrits principalement en langue latine sur le territoire du royaume des Asturies et de León, depuis la mer Cantabrique au Nord, jusqu'à Coria (Cáceres) au Sud. Ces textes s'inscrivent dans une période chronologique qui s'étend du VIII^e siècle, époque à laquelle les plus anciens d'entre eux ont été composés, jusqu'à 1230, année de l'union définitive des royaumes de León et de Castille. Pour des raisons linguistiques et également pratiques, la zone galicienne a été exclue du corpus. Celui-ci rassemble deux types de textes : d'un côté les chartes, qui ont un contenu juridique et sont très nombreuses (nous manipulons concrètement 47 cartulaires, qui contiennent un total de presque 10 000 actes) ; de l'autre côté les chroniques, qui sont au nombre de sept, et incluent le *Poème d'Almería*, qui fait partie intégrante de l'une d'entre elles.

Pour des raisons pratiques, nous travaillons sur des éditions publiées. Ce choix implique d'emblée un inconvénient réel, dont nous sommes et avons toujours été conscients, et qui tient à la différence de qualité de ces éditions. Aussi vérifions-nous à l'occasion par nous-mêmes quelques formes précises, en recourant à l'examen des actes originaux.

Depuis son origine, le travail a été organisé selon quatre principes fondateurs. En premier lieu, il requérait une équipe, cette équipe devant être aussi interdisciplinaire que possible, et associer des philologues latinistes, romanistes, germanistes, arabisants, et des historiens. Depuis le commencement, nous

avons également voulu que le travail se réalise avec l'aide inestimable des moyens informatiques, afin d'être assurés que nous disposions bien de toutes les occurrences de chaque terme dans les sources étudiées. S'agissant des textes historiographiques, la difficulté était résolue d'avance grâce aux concordances complètes des chroniques latines médiévales de l'Espagne déjà publiées (López Pereira *et al.*, 1993). En revanche, rien de semblable n'existait pour les chartes. En ce qui les concerne, le travail a commencé avec leur numérisation et la révision des textes ainsi obtenus afin d'éliminer les erreurs et les éléments étrangers au texte lui-même. Ainsi, autant de fichiers .txt que de cartulaires formant notre corpus ont été créés, et c'est avec eux que nous avons établi notre base de données. Nous travaillons sur cette base de données au moyen de deux programmes informatiques.

Le premier d'entre eux est le logiciel gratuit éditeur de textes ConTEXT, qui permet de réaliser des recherches dans toute la base de données, bien que l'on puisse seulement y rechercher des formes concrètes, par exemple le mot *arbusta* :

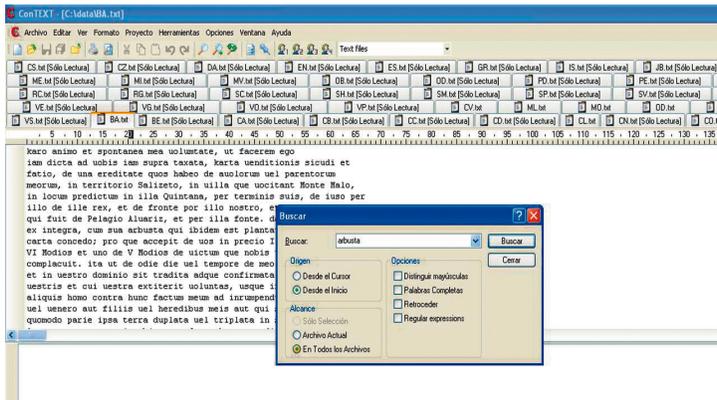


Fig. 1. Recherche du mot *arbusta* dans ConTEXT

Les résultats sont visibles dans un contexte restreint, et l'on peut aussi accéder au texte complet dans lequel l'une des occurrences observées apparaît :

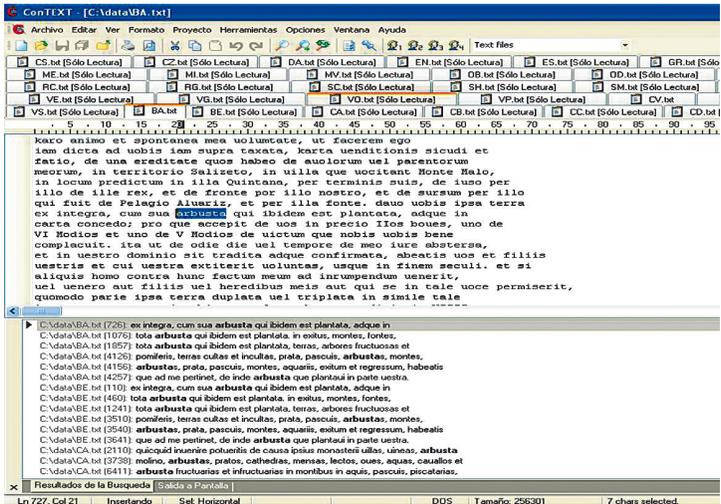


Fig. 2. Résultats de la recherche de *arbusa*

Le second programme utilisé est Microsoft Visual FoxPro, qui permet d'obtenir les concordances avec toutes les occurrences de chacune des formes incluses dans la totalité de la base de données. Pour les réaliser, la liste alphabétique de toutes les formes existantes dans la base de données est un prérequis nécessaire, car il nous faut indiquer, à partir de cette liste, le mot antérieur à la première forme que nous voulons rechercher, et la dernière forme qui nous intéresse. Ainsi pour obtenir les concordances du mot *cautio*, *-onis*, qui apparaît sous trois formes seulement dans nos sources (*cautione*, *cautionem* et *cautionis*), nous devons renseigner le mot qui précède la première de ces formes, c'est-à-dire *cauti*, et la dernière concernée (*cautionis*) :

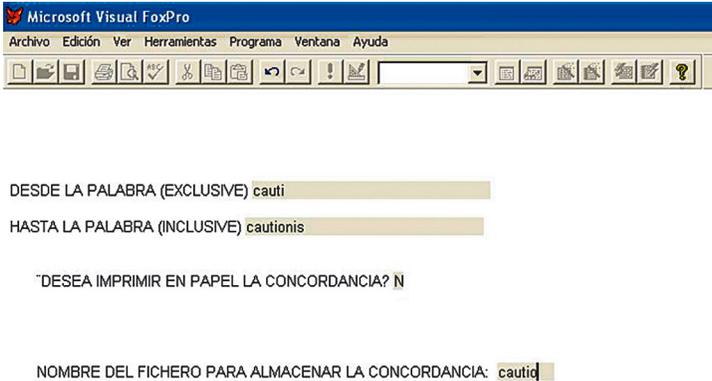


Fig. 3. Recherche des concordances du mot *cautio,-onis* avec Microsoft Visual FoxPro

Par la suite, le programme conçoit la concordance et la conserve dans un fichier auquel nous avons préalablement attribué un nom :

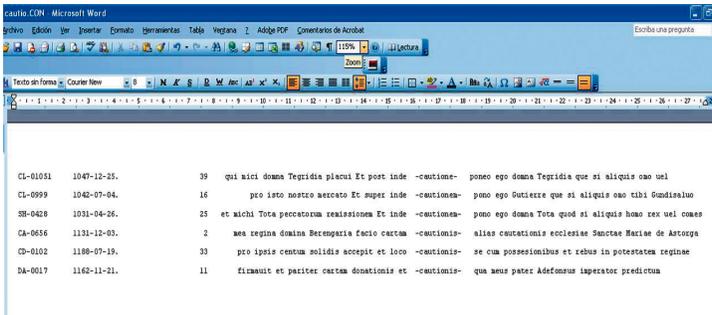


Fig. 4. Liste des concordances de *cautio* obtenues avec Microsoft Visual FoxPro

Dans ces tableaux de concordances, les formes recherchées sont précédées et suivies d'un contexte minimal. Devant ce contexte se trouve la référence de sa localisation : cartulaire, numéro et date du diplôme, numéro de la ligne où se trouve le vocable, etc. Dans l'exemple de concordance présenté ci-dessus, on peut constater, de plus, que deux lettres majuscules sont toujours attribuées à chaque cartulaire du corpus afin de l'identifier : ainsi *CL* correspond au cartulaire de la cathédrale

de León, *SH* à celui du monastère de Sahagun, *CA* à celui de la cathédrale d'Astorga², etc. Les chroniques sont quant à elles abrégées au moyen des trois ou quatre premières lettres de leurs titres (en minuscules): *Alb.* pour la *Crónica Albeldense*, *Adef.* pour la *Chronica Adefonsi imperatoris*, etc.³, et on peut ainsi les distinguer des chartes dès le premier coup d'œil. Ces abréviations sont également utilisées dans le dictionnaire⁴.

La numérisation et le « nettoyage » des textes a constitué une tâche longue et laborieuse, qui s'est achevée en 2001, motif pour lequel la base de données est composée par les chartes publiées jusqu'alors, qui, cela dit, constituent la majorité du corpus existant. Cependant, cela implique que les nouvelles éditions de quelques cartulaires comme ceux de Villaverde de Sandoval (Herrero Jiménez, 2003), d'Eslonza (Ruiz Asencio et Ruiz Albi, 2007) ou de Villanueva de Oscos (Álvarez Castrillón, 2011) ne soient pas utilisées, bien qu'elles soient plus correctes que les anciennes⁵. Une seule exception est à noter: le second volume des chartes d'Otero de las Dueñas, publié en 2005 (Fernández Flórez et Herrero, 2005), a été intégré par la suite.

Nous parvenons ainsi au troisième principe dirigeant le travail: nous avons décidé d'étudier et de réunir dans le *LELMAL* tous les termes qui apparaissent dans les sources sélectionnées, sans aucun genre de limitations (ni en raison de leur origine linguistique, ni à cause de leur ancienneté, ni pour aucun autre motif). Autrement dit nous n'excluons pas les termes du latin classique, ni les termes romans, ni les arabismes, ni les mots de toute autre origine.

Le quatrième principe recteur concerne la dimension lexicographique. Depuis le début de nos travaux, nous n'avons pas suivi le processus traditionnellement appliqué aux dictionnaires

2. Dans cette contribution, nous utiliserons également ces abréviations afin de citer les chartes. À la fin de l'article, le lecteur trouvera la liste des sources notariales mentionnées, avec l'indication des éditions correspondantes.

3. Les mêmes ont été utilisées dans López Pereira *et al.* (1993).

4. Une liste complète de toutes les sources du dictionnaire (et des abréviations qui leur correspondent) se trouve dans le *LELMAL*, p. xviii-xxii.

5. Néanmoins, nous les consultons lorsque cela est nécessaire.

de latin classique ou médiéval, qui consiste à effectuer les entrées lexicographiques par ordre alphabétique. Nous avons décidé, au contraire, de procéder par champs sémantiques ou lexicaux, un système qui présente divers avantages : d'un côté, chaque rédacteur peut se spécialiser dans un ou plusieurs de ces champs ; de l'autre, la comparaison des valeurs et des usages d'un terme avec ceux d'autres termes semblables est facilitée, ce qui garantit d'obtenir une précision et une profondeur plus grandes dans les résultats. Cependant, il faut reconnaître que le concept de champ sémantique a parfois été employé de manière assez souple, si bien que les rédacteurs ont fréquemment plutôt travaillé, en réalité, sur des ensembles de termes qui les intéressaient plus particulièrement pour une raison quelconque.

Caractéristiques externes

Afin que les utilisateurs puissent tirer du *LELMAL* le maximum de profit, il est nécessaire qu'ils connaissent bien ses caractéristiques externes, c'est-à-dire celles de sa structure, car il s'agit d'une œuvre complexe, dans laquelle de multiples éléments ont été pris en compte. Je vais ici résumer une à une les plus importantes de ses particularités externes⁶.

Chaque article lexicographique peut être composé de six parties différentes :

1) *Lemme ou entrée lexicale* : Il apparaît toujours en caractères gras. Comme l'une des caractéristiques de la langue écrite au Moyen Âge est la grande variété graphique avec laquelle les mots sont reportés – et cela devient spécialement évident dans les textes notariaux – nous utilisons comme entrée la forme graphique la plus fréquente dans nos sources : par exemple *añinitas, -atis* ; *castanedo* ; *coniermana, -e*.

2) *Variantes formelles* : Aussitôt après l'entrée, les variantes graphiques du lemme sont indiquées par ordre alphabétique. Les variantes avec une seule occurrence sont signalées au moyen

6. Le *LELMAL* en offre une description plus détaillée (en espagnol, anglais et français), p. vii-lxxvii.

d'un point d'exclamation, qui les précède⁷. Après les variantes, on ajoute les dérivés du terme s'il y en a, et on précise si ce sont des noms propres (toponymes ou anthroponymes⁸).

3) *Étymologie* : La ligne immédiatement inférieure à celle de l'entrée et des variantes est dédiée à l'étymologie du terme, sauf si celui-ci dérive directement de l'indo-européen. L'étymologie s'exprime toujours de façon brève, hormis dans les cas où l'origine du mot est polémique, et où il faut faire place à l'exposé du débat et même à l'indication de la bibliographie appropriée.

4) *Définition* : C'est la partie centrale de l'article lexicographique. En premier lieu, il faut remarquer que notre intention était de ne pas multiplier inutilement les significations des termes, en nous contentant d'énoncer les usages basiques et en réservant les cas problématiques pour les exemples. Toutes les valeurs présentes dans nos sources se recueillent sans exception. Elles sont numérotées et ordonnées lorsque c'est possible, par ordre décroissant d'ancienneté. Celles qui existaient déjà dans l'Antiquité classique, c'est-à-dire dans la période qui s'étend jusqu'au III^e siècle apr. J.-C., sont précédées d'un astérisque. Au moyen de deux barres verticales en caractères gras, on sépare les acceptions concrètes d'une signification plus générale⁹.

Chacune des définitions est suivie de quelques exemples d'emploi du terme. Nous avons veillé à ce que les exemples soient les plus variés possibles, que leur contexte soit suffisant pour une compréhension parfaite, et que toutes les variantes graphiques du terme soient représentées avec eux. L'exemple le plus ancien et le plus récent de chacune des variantes est systématiquement donné. Les exemples se situent selon l'ordre suivant : d'abord ceux provenant des chroniques, par ordre alphabétique de leurs abréviations ; puis ceux provenant des actes, par ordre

7. Par ex., les variantes de *afinitas,-atis* sont : *!adfnitatibus, affinit-*. Celles de *castanedo* : *castacnedo, castagnedo, !castaneto, castaniedo, castanieto, castanned-, castanneto, !kastanedo, !kastanieto*. Celles de *coniermana,-e* : *coermana, !cogermanas, coiermana, congermana, cormana, quarmana, quermana, !quoiermanis*.

8. Par ex., le mot *auellano*. Var. : *!auelano*. Dériv. : *auelaned/t-, auellaned/t-, aulaneda*.

9. Par ex., *afinitas,-atis* a les significations suivantes : 1. *Proximité, voisinage. 2. *Affinité, parenté. || Parent. 3. Limite. De celles-ci, uniquement 1 et 2 sont classiques.

chronologique. Chaque exemple est toujours précédé de sa référence exacte, indiquant la page et la ligne de l'édition où se trouve le vocable étudié, dans le cas des chroniques ; dans le cas des chartes, le cartulaire, le numéro du document et de la ligne concernée, et à la suite la date entre parenthèses. Dans tous les exemples, le vocable analysé apparaît écrit *in extenso* et en caractères gras¹⁰.

Si le vocable a également dans les sources un usage toponymique ou anthroponymique, ce dernier emploi est indiqué après deux barres verticales et est aussi accompagné d'exemples¹¹.

5) *Noms des auteurs*: Le *LELMAL* constitue une œuvre d'équipe, mais n'est pas anonyme pour autant. Chaque article lexicographique est signé des initiales de son (ou de ses) rédacteur(s), situées à la fin de l'article, près de la marge de droite¹².

6) *Notes*: L'une des caractéristiques propres au *LELMAL*, car peu fréquente dans les dictionnaires, est la présence de notes. Les notes sont optionnelles, c'est le rédacteur de l'entrée qui décide de les ajouter ou non, ainsi que de l'étendue de leur texte. Les notes contiennent une information additionnelle : historique, linguistique, bibliographique, précisant la fréquence et le lieu

10. Voir l'exemple développé dans la note 12.

11. Dans le cas de *auellano* (voir note 8), indiqué comme ci-après : « || *Sust. (casi siempre fem.) usado como topónimo con la adición del sufijo -eto/a > -edo/a*: *CL 1306.17 (c.1100)* Id sunt : ... De **Auelanedo**, modio; *CN 4.11 (1120)* Do etiam uobis **Auelaneta** cum omni hereditate sua; *CO 143.33 (1122)* ecclesiam Sancti Vincentii et **Auelanetam**; *SH 1461.5 (1191)* illam hereditatem quam habeo del Escouio de **Auelanedo**; *RC 106.14* dedit ad Corias pro anima sua medietatem de illa uila de **Aulaneda** cum suis pertinenciis; *RC 151.30* cum medietate de Sancto Iohanne de Godan et **Aulaneda** integra ».

12. S'ils sont deux ou davantage, les initiales apparaissent par ordre alphabétique. Voici un exemple complet d'un terme tel qu'il se trouve dans le *LELMAL* :
« **audacter**. *Var.:* !audaciter.

Adv. en -(i)ter formado sobre el adj. audax,-cis.

1. **Con audacia, valientemente*: *Adef. 197.16* septimo uero die **audacter** uiri bellatores Christiani eruperunt de ciuitate per portas ad occasum solis; *Sil. 121.5* Idem uero qui eum tam **audaciter** percussit,...ab opidanis incolumis receptus est; *CL 442.8 (975)* **audacter** superuia inflati et malitia subducti atque diabolico furore accensi; *CL 1343.16 (1113)* precipio ut omnis quicumque sine uestra iussione intrauerit **audaciter** infra terminos... calumnia quam intus fecerit pariat. *MPG* ».

d'usage du terme, etc. Ces notes sont indiquées dans le texte de l'article lexicographique à l'aide d'une lettre minuscule qui figure en exposant, et sont développées à la fin de l'article. La minuscule peut être répétée si la même note concerne plus d'un élément de l'article.

Pour en finir avec la description des caractéristiques externes du dictionnaire, il faut évoquer ici une catégorie spéciale de termes. Il s'agit des mots inconnus et des « mots fantômes ». On désigne par cette expression des mots sans réalité linguistique, en général issus d'erreurs, qu'elles soient du fait du scribe médiéval ou bien du transcripateur moderne. Pour que le lecteur du dictionnaire puisse immédiatement percevoir qu'il se trouve devant l'un de ces termes, le lemme est présenté en petites majuscules, et l'article obéit à un ordre spécifique: tout de suite après l'entrée arrivent le (ou les) exemple(s) et, ensuite seulement, la définition possible du terme¹³.

Il faut remarquer, de même, qu'à toutes les variantes graphiques des vocables correspond une entrée, classée dans le dictionnaire à la place qui lui revient selon le principe alphabétique, et au niveau de laquelle, au moyen d'un *videatur*, on renvoie à l'article lexicographique qui leur est attribué.

Nous souhaitons par ailleurs que le dictionnaire soit d'un usage pratique et d'un contenu précis, suffisant sans pour autant se révéler prolix; il importe que sa consultation soit aisée. Nous ne l'avons pas conçu comme le résultat final de nombreuses recherches, sinon comme leur matrice, le point de départ de possibles travaux institutionnels, linguistiques, historiques, littéraires, etc. Nous cherchons à ce que le *LELMAL* devienne indispensable aux historiens, utile aux philologues

13. Par ex., « ADEIGNAUIMUS: CL 910.11 (1032) ut faceremus ad uobis kartula uendicionis... de terra nostra propria..., determinata ipsa terra :... quarta parte karrale qui discurre ad Legionem, ubi termini determinauimus et coram testis **adeignauimus**.
Forma fantasma producto de un error, bien del escriba medieval bien del editor moderno, que afecta solo a la letra e, que en realidad debería ser una s, es decir adsignauimus, verbo que aparece en ese mismo contexto en otros diplomas: por ej. CL 922.15 (1033) in karale qui discure ad a Treualiolum, per ubi terminu delimitauimus et coram testis adsignauimus. EPR ».

latinistes et romanistes et précieux pour tous. Cependant, son but premier reste de faciliter la lecture et la compréhension des textes médiévaux rédigés en langue latine du Moyen Âge ou de la période romane.

Le travail de rédaction du *LELMAL* est aujourd'hui encore en cours, avec l'élaboration de nouveaux termes (jusqu'à présent presque 4000) et la correction de la partie déjà publiée. Notre aspiration est de pouvoir achever le dictionnaire du royaume de León d'ici trois à quatre années, et de l'enrichir par la suite avec les textes castillans.

Exemples

Voyons maintenant les exemples. Nous en avons choisi trois relatifs au roman, plus un « mot fantôme ».

Ventrescas

Le premier terme retenu est *uentresca*, un substantif roman dérivé du latin *uenter* (« ventre »), particulièrement intéressant du fait qu'il est à peine attesté en castillan avant le *xviii*^e siècle (Leiva, 1999). À l'époque médiévale, l'examen des sources de la zone castillane ne permet d'isoler que deux attestations seulement du vocable. On peut le lire sous la forme *ventresga* dans le *Fuero de Zurita de los Canes* (ca. 1218-1250) :

*Et sabedera cosa es que el pellegero no a de retener alguna cosa
delas UENTRESGAS, delas pelleias, ni de otras taiaduras*¹⁴

et, sous la forme *ventrisca*, dans les *Cortes de Jerez* de 1268 (Alfonso X¹⁵) :

*la dosena delas gorias delas nutrias seys mrs.; la dosena delas
UENTRISCAS delas nutrias dose mrs.*

Il semble qu'il soit également attesté au Moyen Âge dans la zone catalano-aragonaise : plus précisément, Miguel Gual

14. Ureña y Smenjaud (1911). Texte trouvé dans Real Academia Española (RAE), Banco de datos (CORDE) : *Corpus diacrónico del español*, en ligne : <http://www.rae.es> [consulté le 03 février 2015].

15. Avec la précédente citation, ce sont les deux seules occurrences médiévales de ce terme que nous avons trouvées dans Real Academia Española (RAE), en ligne, voir note 14.

Camarena¹⁶ l'a repéré dans les ordonnances de courtiers de Barcelone remontant à 1271 et dans les taxes de « *reua* » de Perpignan datées de 1284¹⁷, où sont évoquées des *ventresques de lú(d)ries* ; et, sous la forme *bentresqua*, il s'utilise en Aragon au xv^e siècle¹⁸. Dans tous ces usages, son signifié est « peau de la zone ventrale des animaux¹⁹ », des loutres, en particulier, dans les exemples catalans. D'après le *DCECH* (s.v. « vientre »), la forme castillane provient du français²⁰. Cependant, en français ancien, on trouve *ventresches/kes* dans un document provenant d'Orléans et daté de 1340, et dans plusieurs textes du xv^e siècle, avec le signifié de « ventrière (pièce du harnais) » ou de « filet du poisson (tranche du flanc)²¹ ». Le dictionnaire de français ancien de Godefroy (s.v. « ventresche »²²) le signale également avec le signifié de « peau du ventre » dans un texte du xiii^e siècle. En italien, il est présent fréquemment dès le xiv^e siècle²³. Il semble donc que, dans toutes ces zones géographiques, y compris la région castillane, ce substantif soit documenté depuis le xiii^e siècle. Ce panorama, brièvement décrit ici, peut être complété grâce au travail réalisé dans le cadre du *LELMAL*. Le vocable se trouve en effet attesté dans les sources du *LELMAL* sous la forme *uentrescas*²⁴. Nous disposons de deux occurrences dans autant

16. Gual Camarena (1968), s.v. « lúdría ».

17. Les textes se trouvent dans Gual Camarena (1968), p. 126-135 et 142-147, respectivement.

18. Sesma Muñoz et Líbano Zumalacárregui (1982), s.v. « bentresqua ».

19. Voir Corominas, *Diccionari etimològic i complementari de la llengua catalana*, s.v. « ventre ».

20. Leiva (1999, p. 156) propose, d'un côté, que le castillan aurait introduit ce vocable comme un emprunt du catalan ou de l'italien et, de l'autre, affirme que le castillan *ventrecha* dérive de l'influence du français *ventresche*.

21. D'après le *DMF* [consulté le 03 février 2015].

22. [consulté le 03 février 2015]. Nous n'avons trouvé le terme dans aucun des autres dictionnaires de français ancien inclus dans : <http://www.micmap.org/dicfro/dictionaries>.

23. D'après Leiva (1999, p. 149). Dans les bases de données qui incluent l'*Opera del vocabolario italiano* nous n'avons trouvé qu'une seule fois *ventresche*, dans un texte daté de 1234, le *Libro di Mattasalà di Spinello* (« *penello dele ventresche di madona Moschada* »). En ligne : <http://www.oivi.cnr.it/Il-Vocabolario.html> [consulté le 13 février 2015]. Le terme n'est toutefois pas mentionné dans le *TLIO*.

24. Nous ne l'avons pas localisé dans les textes latins de la Galicie : *Corpus Documentale Latinum Gallaeciae (CODOLGA)*, versión 10 (2013, en ligne : <http://corpus.cirp.es/codolga> [consulté le 13 février 2015]), ni dans ceux de la Catalogne : *Corpus Documentale Latinum Cataloniae (CODOLCAT)*, versión 3 (2014, en ligne : <http://gmlc.imf.csic.es/>

de chartes du XIII^e siècle, originaires de deux monastères des actuelles provinces de León et de Zamora, respectivement :

CD 172.6 (1202) *Verum tamen circa ordinis susceptionem haec erit dispensatio, quod neque cocullam nec uellum coetur suscipere donec ipsa uellit, sed lineis et pannis albis ac pardis, et pelibus agninis et eis quae uulgo dicuntur UENTRESCAS ;*

SM 123.33 (1221) *Et uxori uestre unum mantum et una garnacha de stanforte de quatuor in quatuor annos et mantum habeat pennam albam de UENTRESCAS et garnacham similiter.*

Le premier texte évoque une dispense accordée aux religieuses du monastère de Carracedo au moment de prendre le voile : on ne les oblige pas à se vêtir de la cape à capuche ni du voile tant qu'elles ne le désirent pas, elles peuvent choisir de porter des vêtements de lin ou d'étoffe blancs ou de couleur foncée et des peaux d'agneaux, ou encore celles que l'on appelle vulgairement « *uentrescas* ». Le second extrait est tiré d'un document testamentaire léguant à une femme un manteau et un autre vêtement, une garnache d'estanfort : tous deux sont taillés dans une peau blanche de « *uentrescas* ».

Le problème que posent ces deux textes tient au fait qu'ils sont uniquement conservés dans des copies datant du XVIII^e siècle. Il est impossible de savoir avec une totale certitude si le mot qui nous intéresse provient bien de l'original, mais c'est toutefois probable parce que, comme nous avons pu l'observer, il existe d'autres témoins du terme dans les textes de la péninsule Ibérique du même siècle. S'il en était ainsi, nous serions devant les attestations les plus anciennes provenant de ce territoire. Dans les textes de León, le vocable se rapporte à des vêtements féminins doublés avec des peaux d'animaux et a la même signification que dans les autres textes péninsulaires antérieurement cités : « peau de la zone ventrale d'un animal », laquelle, comme le soulignent certains de ces textes, a une plus grande valeur que la peau du cou²⁵. Ici également, il s'utilise au

codolcat [consulté le 13 février 2015]), ni non plus dans le *Corpus de documentos españoles anteriores a 1700 (CODEA), versión 2011* (en ligne : <http://demos.bitext.com/codea> [consulté le 13 février 2015]).

25. Voir Gual Camarena (1968), s.v. « lúdría », p. 353.

pluriel. Le premier exemple cité nous apprend, de plus, qu'il s'agit d'un mot de la langue parlée, au moyen de l'expression *uulgo dicuntur*. Ce fait, ainsi que la date précoce de cette occurrence, antérieure encore à celle des témoignages catalans et français connus, semble indiquer une origine patrimoniale²⁶. De plus, à en juger d'après les données dont nous disposons, l'usage du terme dans l'Espagne médiévale a été faible et, après le xv^e siècle, il a cessé d'être utilisé. Cependant, il se réintroduit au xvi^e siècle par l'influence italienne, avec une nouvelle signification²⁷.

leras

Le second exemple retenu est un mot relatif à la façon de mesurer les terres. Il s'agit du terme *ieras*, que nous avons uniquement trouvé dans une charte provenant du monastère cistercien de Moreruela (Zamora) :

MO 39.20 (1195) *In alio loco damus uobis tres terras que fuerunt Pelagii Citiz. Et in alio loco quantum ad nos pertinet de Pelagio Velloso et unam terram que fuit uxoris Pelagii Saluadoriz. Et*

-
26. Sa dérivation de la forme diminutive du latin vulgaire **uentriscula*, telle que la proposait Meyer-Lübke (*REW* 9211), paraît impossible, mais peut-être pas celle de *uenter* avec le suffixe pan-roman *-esco*, lequel, bien que sans trop de profusion, est également présent en castillan médiéval, surtout dans des adjectifs : par ex. *grezisco*, *morisco*, les deux témoignés dans nos textes, mais aussi le substantif *parentesco*, dont nos sources présentent également un cas (CL 1391.4 [1129]). Au sujet de ce suffixe, voir Malkiel (1972).
27. Le terme, d'après Leiva (1999, p. 141-142), s'emploie dans deux œuvres castillanes du xvi^e siècle de forte influence italienne : *Retrato de la Loçana Andaluza* (1524) de F. Delicado et le livre de cuisine *Libro de Arte de Cocina* (1599) de D. Grando, avec le signifié de « ventre du porc » (voir aussi Leiva, 2001, p. 544-546). Et au xviii^e siècle, Martín Sarmiento (*Lettre sur les thons*, 1757) désigne le terme comme italien ou portugais, et déjà référant au thon : « *Ya hemos llegado al Atun hecho. De el se preparan dos alimentos uno de imus venter latino Hypogastrion Griego, y Atun de hijada Castellano. A esto los Italianos llaman ventresca, y los Portugueses ventrisca* » (de Salas et García Solá, 1876, p. 157). La situation que nous percevons par rapport au terme *ventresca* dans la zone castillane concorde avec ce qu'observe Malkiel (1972) pour le suffixe *-esco* à l'échelle européenne : il s'emploie dans les formes médiévales du français, du provençal, du catalan, de l'espagnol et du portugais, mais à partir de la Renaissance et en raison de la grande influence italienne que subit tout l'Occident, il revient dans ces langues comme un italianisme. Une telle situation tend à confirmer, de même, le fait que le terme se trouvait dans les chartes léonaises originales, et qu'il ne s'agissait pas d'une introduction des copistes du xviii^e siècle, époque à laquelle, comme l'indique le témoignage de Martín Sarmiento cité plus haut, ce substantif s'utilisait avec un autre signifié. Sous la forme *ventrecha*, ce vocable est indiqué dans le *DRAE*, pour la première fois dans son édition de 1817, avec le signifié de « ventre des poissons ».

in alio loco XV IERAS et unam cupam. Hoc itaque concambium damus uobis pro hereditate quam a uobis accepimus.

Le LHP (s.v. « iera ») affirme que *iera* dérive du terme du latin tardif *hera* < *area* et a la signification de « friche », inspirée de celle proposée par Du Cange pour (*h*)*era* (s.v.) : « le champ ou le lieu qui n'est ni cultivé ni labouré ». Cependant, une telle signification ne convient pas à notre texte, car *iera* y apparaît accompagné d'un adjectif numéral, alors que *area* / *era* ne s'utilise jamais avec un adjectif semblable dans les sources dont nous disposons²⁸. Ce qui est affirmé dans le DCECH au sujet du substantif roman *jera* (s.v.) nous semble plus approprié. Il y est décrit comme un vocable léonais dérivé du collectif latin *diaria* (« labeur, salaire journalier ») avec le signifié de « surface de terre labourée en une journée par une paire de bœufs, labeur d'une journée », bien que ses auteurs ne le trouvent pas attesté avant 1627. Il s'agit d'un substantif employé jusqu'à nos jours aux Asturies, dans les régions de León, Salamanque, Estrémadure et Zamora²⁹, territoire qui appartient au royaume de León. Zamora est aussi la zone de provenance de la charte à laquelle nous nous reportons. En galicien, il existe encore *geira*³⁰, également avec le signifié de « mesure agraire : arpent de terre », un signifié similaire à celui que, croyons-nous, prend *ieras* dans son unique occurrence parmi les sources léonaises médiévales, que nous pourrions définir comme : « arpent de terre : mesure de terre équivalente à la surface labourée en une journée par une paire de bœufs³¹ ». Notre texte indique que sont livrés « une terre qui a appartenu à la femme de Pelayo Salvadorez, et dans un autre endroit 15 arpents de terre, et un tonneau³² ».

28. Voir LELMAL, s.v. « area,-e ».

29. Le Men, s.v. « jera », reprend ses usages dans toutes ces provinces, ainsi qu'en Galicie et au Portugal.

30. Pour le Portugal, voir aussi Santa Rosa de Viterbo, s.v. « geira », où il est défini comme « labeur ».

31. Le Men cite un exemple très similaire au nôtre, originaire d'Olivenza (Estrémadure) : « *He comprado 20 geras de olivar* ».

32. Il faut comparer le texte avec cet autre, où un autre terme de mesure agraire est utilisé à la place de *ieras* : IS 198.30 (1214) *insuper* (sc. *concedo*) *quatuor aranzadas de uinea et unam cupam de XX modis*.

Combo et recombo

L'adjectif *combo* peut aussi nous informer sur la langue parlée dans la zone étudiée. Il apparaît uniquement dans deux exemples tirés des sources asturiennes, concrètement du cartulaire de la cathédrale d'Oviedo :

CO 23.46 (926) *per terminos suos: per Selia, per regum de Argandoe et per pando et per forkata de illa Ornia et per petra ficta et per petra longa et per forkata et per arbor COMBO, per serra de Ossilis*; CO 88.14 (1084) *per suos terminos pernominatos: per flumen Seliam et per Telluam, quę est integra mea, et per pinnam Forcatam et per arborem COMBUM et per busto Arnales et per Riumuode usque in Seliam.*

Aucun original de ces deux textes n'est conservé, nous ne disposons que de copies incluses dans le problématique *Liber Testamentorum* de l'évêque Pelayo³³, réalisé dans le premier tiers du XII^e siècle. Plus précisément, le premier d'entre eux est tiré d'un acte du roi Ramiro II en faveur de l'église d'Oviedo qui présente, paraît-il, de nombreux traits caractéristiques de la documentation de Ramiro III (961-985), ce qui laisse à penser que le copiste a utilisé comme modèle une donation dudit roi³⁴. Une copie de ce document est également conservée dans la *Regla Colorada*³⁵, laquelle diffère seulement quant à certains éléments graphiques qui n'affectent pas notre vocable.

Le deuxième texte, en revanche, est uniquement conservé dans la copie du *Liber Testamentorum*. Entre les deux documents, il y a, vraisemblablement, un peu plus d'un siècle et demi d'écart, mais l'adjectif apparaît dans les deux cas dans un contexte identique: il qualifie le substantif *arbor* à l'intérieur de la démarcation des limites d'une propriété. Étant donnée la grande probabilité que la partie du texte où se trouve l'adjectif soit une interpolation ou un ajout du copiste du *Liber*, nous ne pouvons pas retrouver ce terme au-delà de la date de composition du *Liber*, c'est à dire dans les années 1120.

33. Valdés (2000), p. 522.27 et 616.6, respectivement.

34. Comme l'affirme Fernández Conde (1971), p. 185-189. Voir aussi Valdés (2000), p. 129-130.

35. Rodríguez Díaz (1995). Ce cartulaire a été réalisé à Oviedo en 1384.

En ce qui concerne l'origine du mot, il faut dire qu'il n'existait pas dans le latin de Rome. Cet adjectif appartient à la famille du substantif *comba* (« convexité, concavité »), terme d'origine celtique³⁶. Ce nom est considéré dans le *DCECH* (s.v. « *comba* ») comme un vocable dialectal, léonais ou mozarabe, probablement apparenté avec le terme gallo-latin *cumba* > français *combe* (« vallée étroite et profonde, petite vallée encaissée³⁷ »), mot qui se trouve attesté à la fin du ^{xiii} siècle en Gaule et qui a laissé une trace en occitan (*comba*) et en catalan (*coma*). Avec la même signification, le substantif *cumba* s'utilise fréquemment dans les diplômes latins médiévaux de Catalogne : on l'y trouve attesté pour la première fois en 844³⁸, et aussi en Angleterre au début du ^{xiii} siècle³⁹; en revanche, on ne trouve pas l'emploi adjectival. D'après les auteurs du *DCECH*, l'adjectif *combo* est beaucoup plus « incertain » car, selon leurs informations, il n'est pas signalé avant la fin du ^{xviii} siècle, où il apparaît dans un dictionnaire, et n'est pas, non plus, confirmé par les dialectes. De ce fait, il leur semble que « si l'adjectif *combo* n'est pas un mot fantôme, déduit par les lexicographes de *comba* et de *combado*, il serait, tout au plus, un adjectif dérivé secondairement de ces mots ». Cette opinion ne peut plus être tenue face aux exemples tirés des sources asturiennes. Conformément à ceux-ci, nous pouvons affirmer que l'adjectif existe au moins depuis le commencement du ^{xii} siècle, et que ce doit être, en effet, un terme spécifique et exclusif de la langue parlée dans la zone asturienne⁴⁰, spécialement utilisé pour désigner les arbres au tronc courbé de façon très exagérée.

Le substantif *comba* est également attesté dans notre corpus, bien que, dans ce cas, uniquement sous la forme d'un nom propre, toponymique et patronymique :

36. De la racine celtique *kumbo- (« courbé ») ; en gallois *cwmm* (« vallée »). Voir à ce sujet la bibliographie citée dans le *GMLC*, s.v. « *cumba* », n. 2.

37. Définition du *Dictionnaire du moyen français* (1330-1500), 2012, en ligne : <http://www.atilf.fr/dmf> [consulté le 03 février 2015].

38. *GMLC*, s.v. « *cumba* ». Néanmoins, la première attestation de ce substantif dans le reste de l'Espagne est datée, suivant le *DCECH*, de 1573.

39. *DMLBS*, s.v. « *cumba* ».

40. Toujours en usage actuellement, puisqu'il figure dans le *DLA*, s.v. « *cumbu*, -a, -o ».

RC 97.15 *dedit ei iure hereditario illos montes et de Arganza et de Porcinero et de CONUA*; RC 98.8,9 *De Sancto Mamete de CONUA. Sancto Mamete de CONUA fuit de Zalone*; IS 222.23 (1225) *Guilielmus de la CONBAS*.

Le toponyme se trouve employé de manière similaire uniquement dans la documentation asturienne, ce qui confirme le caractère local de l'adjectif. Son dérivé *recombo*, présent dans un unique document, encore conservé exclusivement dans la copie du *Liber Testamentorum* (587.12), le confirme à nouveau :

CO 117.62 (1100) *Ex alia parte per riuum quem dicunt Bullera ad sursum et per quoto Pennino et per illa aquilera et per Penna de Rege et per illa uerruga et per illo trabe et per illa spelunca et per arbor RECOMBO et per illo scuio quem dicunt Pede de Mula*.

Comme on peut l'observer, il s'emploie dans le même contexte et avec un signifié identique au terme simple. Il semble clair que ces mots sont deux vocables de la langue parlée exclusivement dans la zone asturienne du royaume, qui s'utilisent ici dans un souci de précision propre aux exigences de la profession notariale⁴¹.

Plentum

Un dernier exemple permet d'illustrer un autre type de difficultés procédant des chartes lorsqu'un travail lexicographique est réalisé à partir de leurs données. Il s'agit de la forme insolite *plentum*, que nous trouvons dans le texte suivant :

CL 704.3 (c.1011) *Nodicia de hereditates et uineas qui sunt in Villa Citi Rege post parte Sancti Iacobi apostoli, fratris Domini. Id sunt: cortes et solares, hereditates et uineas qui fuerunt de domna Sinduara abbatissa siue ab PLENTUM et parentum quomodo et terras suas ganantias ab omni integritate*.

C'est une forme unique dans notre documentation et de surcroît inconnue en latin, aussi bien antique que médiéval, ce qui nous laisse à penser qu'elle relève d'une confusion du scribe médiéval. Il s'agirait d'un de ces mots que nous nommons « fantôme ». Il se trouve dans un contexte évoquant des possessions, dans lequel est également mentionnée l'origine de ces dernières,

41. Dans les sources du *LELMAL* on trouve quelques autres régionalismes, voir Pérez Rodríguez (2014).

avec le génitif *parentum* et avec *ganantias*. Dans les chartes, ces contextes sont habituels, et les expressions suivantes sont communément utilisées :

CA 24.10-11 (923) *donis nostris quos auemus de auibus nostri set de parentibus nostris uel de nostras ganantias*; CA 297.7-8 (1044) *uilla nostra propria que abemus de nostros auios et de nostros parentes et de nostras ganantias*; CR 24.5-6 (1125) *nostra ereditate que abeo de parentorum meorum, sibi de ganantias quomodo de suas comparaturas*; IS 3.20-21 (1043) *ereditates quos uiro meo... abuit de abios et parentibus et illas ganantias*.

Il ne semble pas cependant que *plentum* puisse être le résultat d'une déformation qui affecte *auios/auibus*. Au contraire, nous observons que ce terme rime avec le terme *parentum* qui le suit, ce qui se produit avec une certaine fréquence dans les formules documentaires, et l'altération découle peut-être précisément de cette recherche de l'assonance. Nous estimons, par conséquent, que c'est probablement le résultat d'une adultération de la forme *plenum* ou *plenius*, qui apparaît fréquemment dans tels contextes dans les diplômes asturiens léonais ; parfois, comme dans notre texte, dans l'expression adverbiale *ad plenum/ plenius* (« totalement, complètement, entièrement »), par exemple : SH 1612.20 (1218), « *uobis AD PLENVM restituat* » et – beaucoup plus proche de celle qui inclut *plentum* – CL 405.23 (967), « *terras ab omni integritate quos comparauit..., paludes, carrizales, terras AT PLENIVS habundanter...* » Le sens du texte serait approximativement : « Nouvelle des domaines et vignes de Villacid del Rey, propriété de l'apôtre Santiago, frère du Seigneur. Ce sont des cours et des terrains, des propriétés et des vignes qui ont appartenu à l'abbesse Mère Sinduara au complet, ou bien hérités de ses parents ou bien des terres gagnées par elle dans toute leur intégrité ». C'est l'hypothèse qui nous semble la plus plausible.

Nous pourrions continuer notre démonstration avec davantage d'exemples encore, mais ceux qui ont été exposés dans cette contribution, précisant et complétant considérablement (ou bien corrigeant) les affirmations avancées jusqu'à ce jour

au sujet de ces termes, peuvent, du moins nous l'espérons, se révéler suffisamment représentatifs de l'intérêt des apports rendus possibles par le projet LELMAL.

Références bibliographiques

Dictionnaires

COROMINAS, Joan, *Diccionari etimològic i complementari de la llengua catalana*, 9 vol., Barcelona, Curial Edicions Catalanes, 1983-1991.

DCECH = COROMINAS, Joan et PASCUAL, José Antonio, *Diccionario crítico etimológico castellano e hispánico*, 6 vol., Madrid, Gredos, 1980-1991.

DLLA = *Diccionariu de la llingua asturiana*, Uviéu, Academia de la Llingua Asturiana, 2000.

DMF = *Dictionnaire du moyen français (1330-1500)*.

En ligne : <http://www.atilf.fr/dmf>, 2012.

DMLBS = *Dictionary of Medieval Latin from British Sources*, Oxford, Oxford University Press, 1975-2014.

DRAE = *Diccionario de la lengua española*.

En ligne : <http://www.rae.es>.

DU CANGE, *Glossarium mediae et infimae latinitatis* [Niort, L. Favre, 1883-1888], 5 vol., Graz, Akademische Druck, 1954.

En ligne : <http://ducange.enc.sorbonne.fr/>

GMLC = *Glossarium Mediae Latinitatis Cataloniae. Voces latinas y romances documentadas en fuentes catalanas del año 800 al 1100*, Barcelona, Universidad de Barcelona, 1960-.

GODEFROY, Frédéric, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*, Paris, Vieweg/Bouillon, 1880-1895.

En ligne : <http://micmap.org/dicfro/search/dictionnaire-godefroy>.

LELMAL = PÉREZ GONZÁLEZ, Maurilio et PÉREZ RODRÍGUEZ, Estrella (dir.), *Lexicon Latinitatis Medii Aevi Regni Legionis (s. VIII-1230) Imperfectum. Léxico latinorromance del reino de León (s. VIII-1230)*, Turnhout, Brepols, 2010.

LE MEN, Janick, *Léxico del leonés actual*, 6 vol., León, Centro de Estudios e Investigación San Isidoro, 2002-2012.

LHP = *Léxico Hispánico Primitivo (siglos VIII al XII). Versión primera del Glosario del primitivo léxico iberorrománico*, projet initié et initialement dirigé par Ramón MENÉNDEZ PIDAL, établi par Rafael LAPESA avec la collaboration de Constantino GARCÍA, sous la direction de Manuel SECO, Madrid, 2003.

REW = MEYER-LÜBKE, Wilhelm, *Romanisches etymologisches Wörterbuch*, Heidelberg, Winter, 1972.

SANTA ROSA DE VITERBO, Joaquim, *Elucidário das palavras, termos e frases que em Portugal antigamente se usaram*, 2 vol., éd. Mário FIÚZA, Porto/Lisboa, Livraria civilização, 1966.

TLIO = *Tesoro della Lingua Italiana delle Origini*.
En ligne : <http://tlio.ovi.cnr.it/TLIO>.

Sources citées

CA = CAVERO, Gregoria et MARTÍN, Encarnación (dir.), *Colección documental de la catedral de Astorga*, 2 vol., León, Centro de Estudios e Investigación San Isidoro, 1999-2000.

CD = MARTÍNEZ, Martín (dir.), *Cartulario de Santa María de Carracedo*, t. I, 992-1500, Ponferrada, Instituto de Estudios Bercianos, 1997.

CL = *Colección documental del archivo de la catedral de León (775-1230)*, vol. I, dir. SÁEZ, Emilio; vol. II, dir. SÁEZ, Emilio et SÁEZ, Carlos; vol. III, dir. RUIZ ASENCIO, José M.; vol. IV, dir. RUIZ ASENCIO, José M.; vol. V, dir. FERNÁNDEZ CATÓN, José Maria; vol. VI, dir. FERNÁNDEZ CATÓN, José Maria; León, Centro de Estudios e Investigación San Isidoro, 1987-1990.

CO = LARRAGUETA, Santos García (dir.), *Colección de documentos de la catedral de Oviedo*, Oviedo, CSIC, 1962.

IS = MARTÍN, Encarnación (dir.), *Patrimonio cultural de San Isidoro. Documentos de los siglos X-XIII*, León, Centro de Estudios e Investigación San Isidoro, 1995.

MO = ALFONSO, Isabel (dir.), *La Colonización cisterciense en la meseta del Duero. El dominio de Moreruela (siglos XII-XIV)*, Zamora, Diputación de Zamora, 1986.

RC = GARCÍA LEAL, Alfonso (dir.), *Registro de Corias*, Oviedo, CSIC, 2000.

SH = *Colección diplomática del monasterio de Sahagún (Siglos IX y X)*, vol. I, dir. MÍNGUEZ, José María.; vol. II et III, dir. HERRERO, Marta; vol. IV, dir. FERNÁNDEZ FLÓREZ, José A.; vol. V, dir. FERNÁNDEZ FLÓREZ, José A.; León, Centro de Estudios e Investigación San Isidoro, 1976-1994.

SM = GONZÁLEZ, Ángel Rodríguez (dir.), *El tumbo del monasterio de San Martín de Castañeda*, León, Centro de Estudios e Investigación San Isidoro, 1973.

Autres ouvrages

ÁLVAREZ CASTRILLÓN, José Antonio (dir.), *Colección diplomática del monasterio de Santa María de Villanueva de Oscos (1139-1300)*, Oviedo, Real Instituto de Estudios Asturianos, 2011.

DE SALAS, Javier et GARCÍA SOLÁ, Francisco, *Memoria sobre la industria y legislación de pesca: que comprende desde el año 1870 al 1874*, s.n., Madrid, 1876.

FERNÁNDEZ CONDE, Francisco J., *El Libro de los Testamentos de la catedral de Oviedo*, Roma, Iglesia Nacional Española, 1971.

FERNÁNDEZ FLÓREZ, José Antonio et HERRERO, Marta (dir.), *Colección documental del monasterio de Santa María de Otero de las Dueñas*, t. II, 1109-1300, León, Centro de Estudios e Investigación San Isidoro, 2005.

GUAL CAMARENA, Miguel, *Vocabulario del comercio medieval. Colección de aranceles, aduaneros de la corona de Aragón (siglos XIII y XIV)*, Tarragona, Excelentísima Diputación Provincial, 1968.

HERRERO JÍMENEZ, Mauricio (dir.), *Colección documental del monasterio de Villaverde de Sandoval (1132-1500)*, León, Centro de Estudios e Investigación San Isidoro, 2003.

LEIVA, Francisca, « *Ventresca y sus parientes: ventrisca, ventresco y ventrecha* », *Analecta Malacitana*, nº 22-1, 1999, p. 139-157.

—, *Vocabulario cordobés de la alimentación (s. XV y XVI)*, Córdoba, Universidad de Córdoba: Servicio de publicaciones, 2001.

- LÓPEZ PEREIRA, José E., DÍAZ DE BUSTAMANTE, José Manuel, VÁZQUEZ BUJÁN, Enrique et LAGE COTOS, María Elisa, *Corpus Historiographicum Latinum Hispanum Saeculi VIII-XII: Concordantiae*, Hildesheim, Olms/Weidmann, 1993.
- MALKIEL, Yakov, « The Pan-European Suffix *-esco, -esque* in Stratigraphic Projection », dans VALDMAN, Albert (dir.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, Den Haag/Paris, Mouton, 1972, p. 357-387.
- PÉREZ RODRÍGUEZ, Estrella, « Localismos en el reino asturleonés entre el s. VIII y 1230: algunos ejemplos del léxico agrícola », dans FERNÁNDEZ, Ángel Martínez *et al.* (dir.), *Ágalma. Ofrenda desde la Filología Clásica a Manuel García Teijeiro*, Valladolid, Universidad de Valladolid, 2014, p. 425-431.
- RODRÍGUEZ DÍAZ, Elena (dir.), *El libro de la Regla Colorada de la catedral de Oviedo. Estudio y edición*, Oviedo, Real Instituto de Estudios Asturianos, 1995.
- RUIZ ASENCIO, José Manuel et RUIZ ALBI, Irene (dir.), *Colección documental del monasterio de San Pedro de Eslonza*, León, Centro de Estudios e Investigación San Isidoro, 2007.
- SESMA MUÑOZ, José Ángel et LIBANO, Ángeles, *Léxico del comercio medieval en Aragón (siglo XV)*, Zaragoza, Institución Fernando el Católico, 1982.
- UREÑA Y SMENJAUD, Rafael de (dir.), *El fuero de Zorita de los Canes según el códice 217 de la Biblioteca nacional (siglo XIII al XIV) y sus relaciones con el fuero latino de Cuenca y el romanceado de Alcázar*, Madrid, Establecimiento Tipografico de Fortanet, 1911.
- VALDÉS, José Antonio (dir.), *El Liber Testamentorum Ovetensis. Estudio filológico y edición*, Oviedo, Real Instituto de Estudios Asturianos, 2000.

Résumés / Abstracts

Sylvie BAZIN-TACHELLA et Gilles SOUVAY,
De la gestion de la variation en moyen français à
son élargissement aux états anciens du français :
le développement du lemmatiseur LGeRM

Résumé

La langue médiévale ne se livre qu'à travers des témoignages écrits, essentiellement mouvants et variants. Le *Dictionnaire du moyen français*, dès ses débuts, a été confronté à cette difficulté. La lemmatisation des vedettes a été nécessaire pour construire la base de données et un outil, le lemmatiseur LGeRM (acronyme de « Lemmes, Graphies et Règles Morphologiques »), a permis de faire du DMF un dictionnaire véritablement électronique, à la fois dans sa conception et dans sa consultation, deux aspects différents mais liés. C'est lui qui permet d'interroger à partir de la forme rencontrée dans un document. Lors de la recherche d'une entrée dans le dictionnaire, l'analyseur isole un mot – hors contexte – et fournit des hypothèses de lemmes. Il utilise pour cela un lexique et des règles de flexion et de variation graphique. Le lexique est constitué des graphies connues avec leur analyse (graphie, lemme, étiquette). Conçu au départ pour le dictionnaire, le lemmatiseur a pu être intégré dans de nouveaux environnements. Grâce à la lemmatisation d'un texte source encodé en XML/TEI, il est possible de l'interroger par forme, ou par lemme, ou en suivant le texte en continu, ce qui est d'une aide considérable pour mener à bien la préparation d'une édition et la construction d'un glossaire. LGeRM a connu d'autres types de développements, en s'adaptant à la morphologie et aux variations spécifiques d'autres états de langue que celui pour lequel il avait été conçu, ce qui a abouti à la construction de deux lexiques distincts : un lexique LGeRM médiéval, optimisé pour la période 1300-1500 et un lexique LGeRM ^{xvi}^e-^{xvii}^e pour 1550-1700, désormais utilisés par le moteur de recherche de FRANTEXT pour

la recherche par lemme. En accès libre sur demande, LGeRM est devenu un outil d'interrogation des textes anciens, en moyen français (cible du *DMF*) et en amont et en aval de la période (ancien français et français des *xvi^e* et *xvii^e* siècles), complémentaire des outils d'étiquetage morphosyntaxique.

Abstract

Medieval language reveals itself only through diverse and unsettled written accounts. Right from the beginning, the creators of the *Dictionnaire du moyen français (DMF)* have tried to overcome this challenge. The lemmatization of the entries was necessary in order to construct the dictionary's database. The team have also used a lemmatizing tool, LGeRM (*Lemmes Graphies et Règles Morphologiques*), to create an electronic dictionary in both its conception and consultation. When an user researches an entry from the dictionary, the analyzer takes a word out of context and provides hypothesis of lemmas. In order to do this, the analyzer utilizes a lexicon and various rules of inflection and spelling variations. The lexicon is made of known written forms with their analysis (spelling, lemma, tag). The lemmatizer was firstly designed for the dictionary, but is now fit for further use. Thanks to the lemmatization of source texts encoded in XML/TEI, LGeRM can analyze an original text per forms, lemma or even pages which is of significant assistance when preparing a text edition or constructing a glossary. LGeRM has undergone other types of developments, being adapted to the morphology and specific variations of other states of language. Therefore, we now have two distincts LGeRM lexicons; one for the medieval period (1300-1500), and another one for the early-modern period (1550-1700). Both are being used by the FRANTEXT search engine for the research by lemma. LGeRM can thus be used to work on Middle French (the target of the DMF), but also on Old French as well as French of the 16th and 17th Centuries. To finish, this query tool is on open access and complementary to Morphosyntactic taggers.

Ana GÓMEZ RABAL, *Le latin médiéval du Glossarium Mediae Latinitatis Cataloniae: un projet lexicographique dans un contexte européen*

Résumé

Le *Glossarium Mediae Latinitatis Cataloniae* (GMLC), dictionnaire du latin médiéval des territoires correspondant au domaine linguistique du catalan entre le IX^e et le XII^e siècle, est réalisé grâce à la collaboration de la section de lexicographie latine du département d'Études médiévales de l'Institut Milà y Fontanals du CSIC (Consejo superior de investigaciones científicas, à Barcelone) avec le département de Lettres latines de l'université de Barcelone. Les responsables de l'élaboration et de la publication de ce glossaire ont comme objectif scientifique de fournir aux philologues, aux historiens et aux juristes, ainsi qu'à toute personne intéressée par le Moyen Âge, un outil qui rende compréhensible la documentation notariale et les textes littéraires, juridiques et scientifiques latins produits dans les lieux et à l'époque cités, textes qui sont le témoignage écrit non seulement de la langue latine médiévale, mais aussi de la langue romane naissante et dont la lecture est, très souvent, compliquée même pour ceux qui ont une certaine habitude de travailler sur des textes en latin.

Les membres de l'équipe du GMLC travaillent en deux phases indissociables et complémentaires, qui évoluent vers un objectif ultime commun : la publication complète du glossaire. La première phase, la *rédaction*, consiste en la préparation, l'élaboration et la mise à jour des articles du glossaire lui-même. Pour la seconde phase, la *numérisation*, les textes utilisés comme matière première pour l'écriture des articles lexicographiques sont passés au scanner, reconnus et corrigés ; les textes corrigés forment un corpus à usage interne qui sert aussi bien pour la rédaction des articles lexicographiques que pour les recherches parallèles des membres du GMLC. Mais cette deuxième phase a désormais comme objectif le développement et l'expansion du *Corpus Documentale Latinum Cataloniae* (CODOLCAT), base de données lexicale de publication périodique (version 1,

en 2012 ; version 2, en 2013 ; version 3, en 2014 ; version 4, en 2015) qui permet l'accès, de façon libre et gratuite, au corpus textuel utilisé pour écrire le *GMLC* ; ce corpus textuel est traité, dépouillé et réédité lors de son introduction dans le CODOLCAT et, finalement, il est présenté sous forme de concordances.

La progression du travail amène l'équipe du *GMLC* à se confronter au défi de l'édition au format numérique du glossaire lui-même. Comme il en va pour les autres dictionnaires de latin médiéval – pour ceux qui sont en cours de publication autant que pour l'ancien Du Cange –, la publication numérique et en ligne s'impose. Le groupe s'est donc engagé, désormais, dans la préparation du balisage en langage XML des articles déjà rédigés. Le projet de publication en ligne des articles déjà publiés sur papier, et des articles futurs des autres lettres encore à rédiger, doit permettre une diffusion maximale de l'œuvre et rendre service aux chercheurs.

Abstract

The *Glossarium Mediae Latinitatis Cataloniae (GMLC)*, dictionary of Medieval Latin from the territories corresponding to the linguistic area of the Catalan from ninth to twelfth centuries, is realised through the collaboration between two institutions: the Department of Medieval Studies of Milá y Fontanals Institution (CSIC, Barcelona) and the Department of Latin Philology of the University of Barcelona. The developers of the glossary have the scientific purpose of providing philologists, historians and jurists, as well as anyone interested in the Middle Ages, a tool that makes understandable the Latin notarial documentation and the Latin literary, legal and scientific texts produced in the mentioned territories and centuries. All these acts and texts are the written testimony not only of the Medieval Latin language but also of the emerging Romance language, and whose comprehension is very often complicated even for those who have a certain habit of reading and working on texts in Latin.

The *GMLC* team divides and shares their functions between two lines of work, inseparable and complementary, which evolve

towards a common ultimate goal: the complete publication of the glossary. The first line is called *writing* and consists of the preparation, development and updating of glossary articles itself. In the second line of work, called *digitalisation*, the texts used as raw material for writing lexicographical items are passed to the scanner, recognized and corrected; the corrected texts form a corpus to internal utilisation, which is used both for writing lexicographical articles and for parallel searches for the members of the *GMLC*. But this second line of work now aimed at the development and expansion of the *Corpus Documentale Latinum Cataloniae* (CODOLCAT), lexical database of serial publication (version 1, 2012; version 2, 2013; version 3, 2014; version 4, 2015), which provides free access to the textual corpus used to write the *GMLC*, processed, marked, re-edited and presented in form of concordances.

As a result of the increase in the working lines described, the *GMLC* team now faces the challenge of publishing in digital format the glossary itself. Just as for the other teams of Medieval Latin dictionaries – those being published and the old Du Cange as well –, the digital and online publication is essential. So, the *GMLC* group is engaged now in the preparation of XML markup of the articles already drafted. The envisioning of the online digital publishing (of articles published in paper and of articles of letters to write) is strongly encouraged to give the work the maximum dissemination and usefulness.

Michèle GOYENS et Céline SZECEL, Autorité du latin et transparence constructionnelle : le sort des néologismes médiévaux dans le domaine médical

Résumé

Dans cette contribution, nous présentons le projet de recherche *Latin authority and constructional transparency at work: Neologisms in the French medical vocabulary of the Middle Ages and their fate*, subventionné par le Fonds de la recherche de la KU Leuven (OT/14/047). Ce projet étudie les raisons pour lesquelles certains néologismes créés dans le

domaine médical au cours du Moyen Âge existent toujours en français moderne, alors que d'autres ne se maintiennent pas. Notre hypothèse de travail est que des critères morphologiques, et plus particulièrement la transparence constructionnelle, jouent un rôle crucial pour la préservation de ce lexique. En d'autres mots, les termes présentant une relation formelle proche de l'élément latin dont ils sont issus se maintiendraient mieux que des créations françaises originales, c'est-à-dire des dérivés ou des composés réalisés à partir de bases morphologiques françaises. Concrètement, nous esquissons les objectifs du projet et ses hypothèses de travail, avant de présenter le corpus numérisé de textes médicaux du Moyen Âge, comprenant des traductions françaises de textes-sources latins ainsi que des textes directement composés en français. Nous expliquons ensuite les facteurs décisifs pour la survie de ces néologismes : ces critères peuvent être externes ou internes, aussi bien d'ordre général que d'ordre morphologique, ces derniers formant la grille d'analyse pour une base de données morphologique numérique de la terminologie médicale médiévale en français, qui sera mise à la disposition de la communauté scientifique. Nous présentons en dernier lieu le cadre théorique de la morphologie des constructions (Booij, 2010), qui permettra de dégager des corrélations au niveau des structures morphologiques relevées, et terminons par une série de perspectives.

Abstract

This article gives an overview of the research project *Latin authority and constructional transparency at work: Neologisms in the French medical vocabulary of the Middle Ages and their fate*, financed by the Research Fund of the KU Leuven (OT/14/047). This project aims at investigating why certain French neologisms that emerged in the field of medicine during the Middle Ages managed to survive, while others disappeared after some time. Our hypothesis is that morphological criteria, in particular constructional transparency, contribute in a crucial manner to lexical preservation. In other words, terms showing a close formal relation with the Latin equivalent from which they

were borrowed, could stand the test of time better than original French creations, i.e. derivations or compounds on the basis of genuinely French morphemes. In this contribution, we first present the objectives of the project and its working hypotheses, before describing the digitized corpus of medieval medical texts, containing both translations from Latin and texts directly written in French. We then set out the external and internal factors decisive for the survival of these neologisms. With respect to internal factors, a first set of criteria concerns more general linguistic characteristics; a second one, the morphological characteristics of each neologism. Those internal criteria form the guiding principles that will allow us to complete an online morphological database of medieval medical French vocabulary, which will be at the disposal of the scientific community. In a last section, we present the theoretical framework of Construction Morphology (Booij, 2010), which will allow us to extract correlations between morphological structures, before concluding our article with a series of prospects.

Elisa GUADAGNINI, La lexicographie de l'Italien médiéval et les corpus de l'OVI : un bilan provisoire et quelques nouvelles perspectives

Résumé

Ce travail décrit sommairement l'histoire de l'OVI (Opera del vocabolario italiano, CNR - Firenze) et de ses projets : depuis les années 1960, ce centre de recherche travaille à la rédaction d'un vocabulaire de l'ancien italien, le *TLIO* (*Tesoro della Lingua Italiana delle Origini*), et à la constitution d'une base de données textuelles. Le Corpus OVI est aujourd'hui librement consultable sur la toile (en ligne : <http://gattoweb.ovi.cnr.it>). Il recueille plus de 23 millions de mots, et représente une ressource incontournable pour toute étude consacrée à l'italien médiéval. Le *TLIO* compte plus de 30 000 articles : lui aussi publié sur internet (en ligne : <http://tlio.ovi.cnr.it/TLIO/>), il est le principal – et le plus ancien – projet italien de lexicographie électronique.

Abstract

This work outlines the history of OVI (Opera del Vocabolario Italiano, CNR - Firenze) and its projects: since the '60s, this research center is working on compiling a dictionary of old Italian, the *TLIO* (*Tesoro della Lingua Italiana delle Origini*), and on creating a textual database. The Corpus OVI is now freely available on the web (<http://gattoweb.oivi.cnr.it>). It collects more than 23 million words and is an indispensable resource for any study of medieval Italian. The *TLIO* has more than 30,000 items: also being published on the internet (<http://tlio.oivi.cnr.it/TLIO/>), it is the main – and the oldest – Italian project of electronic lexicography.

Céline GUILLOT, Serge HAIDEN et Alexis LAVRENTIEV, Base de français médiéval: une base de références de sources médiévales ouverte et libre au service de la communauté scientifique

Résumé

L'essor actuel de la linguistique diachronique a des répercussions importantes sur le développement de ressources numériques qui soient adaptées à la recherche en langue médiévale et accessibles à une très large communauté. L'enrichissement de ces ressources a en retour une influence très forte sur les objets et les méthodologies utilisés pour l'analyse des données ainsi constituées. C'est cette synergie complexe et les implications méthodologiques qui la sous-tendent que nous tenterons d'illustrer dans cet article, grâce à l'exemple du développement de la *Base de français médiéval*. Nous commencerons par donner un aperçu des possibilités offertes par ce corpus numérique et nous présenterons la double chaîne mise en place pour permettre les recherches : chaîne philologique pour la constitution et la préparation des données textuelles, chaîne analytique pour leur exploitation outillée. Nous montrerons de quelle façon ces deux chaînes s'articulent, et les principes qui fondent leur association en vue d'un développement intégré et communautaire: usage de standards internationaux pour

la représentation des données et pour l'architecture des outils d'analyse, licences *open-source* qui permettent la diffusion, l'enrichissement et la pérennisation des ressources textuelles/logicielles et qui garantissent la reproductibilité des analyses.

Abstract

Current developments in diachronic linguistics have an important impact on the production of digital resources that become more and more adapted to research on the medieval language and accessible to a large academic community. The enrichment of these resources has in turn a very strong influence on the objects and the methodologies used to analyse the data obtained in this process. It is this complex synergy and the methodological implications that underlie it that we will attempt to illustrate in this article through the example of the development of the *Base de Français Médiéval*. We will first give an overview of the possibilities offered by this online corpus and then present the double-fold data analysis workflow: a “philological chain” for the constitution and the preparation of the textual data, and the “analytical chain” for their exploitation powered by linguistic tools. We will show how these two chains interact and the principles that form the basis of their association for integrated and community development: international standards for data representation and for tools architecture, open source licenses that allow the distribution, enrichment and long-term preservation of textual and software resources and that ensure reproducibility of the results of analysis.

Robert MARTIN, À propos du *DMF*

Résumé

Le *DMF* (*Dictionnaire du moyen français*) illustre les bénéfices que procure la lexicographie électronique; il fait prendre conscience aussi de tous les pièges qu'elle comporte: l'instabilité, une complexité informatique de plus en plus difficile à dominer, le risque de l'inexistence dans la durée.

Abstract

Das Mittelfranzösische Wörterbuch *DMF* veranschaulicht die grossen Vorteile der elektronischen Lexikografie; das Werk lässt aber auch verschiedene Schwierigkeiten wahrnehmen: die Unbeständigkeit, eine immer schwerlicher überwindbare informatische Komplexität und schliesslich auf die Dauer die Gefahr der Inexistenz.

Ramon MASIÀ, Numérisation et traitement de textes mathématiques grecs: méthodes, problèmes et résultats

Résumé

Le corpus des textes mathématiques grecs (CTMG) contient un peu plus de cent ouvrages qui ont survécu, totalement ou partiellement, depuis le IV^e siècle av. J.-C. C'est donc un corpus relativement restreint. Notre objectif est de le numériser, puis de le traiter avec les outils créés par la linguistique de corpus. D'une part, cet objectif est réalisable précisément parce que le corpus est de taille réduite, mais aussi parce qu'il ne contient presque pas d'ambiguïtés, le nombre d'occurrences du corpus restant faible et les différences de structure syntaxique peu abondantes. D'autre part, la mathématique grecque est rédigée dans une langue spécifique, que les mathématiciens eux-mêmes maîtrisaient très bien, puisque ce champ de savoir dépend entièrement du style dans lequel il a été écrit. Après avoir procédé à la numérisation des textes, nous avons lemmatisé une grande partie du corpus, puis avons procédé à une analyse comparative de différents textes et auteurs. Au cours de cette première étape, nous avons constaté qu'une telle approche quantitative dans le contexte de l'étude des CTMG était pertinente et nécessaire à la recherche consacrée aux mathématiques grecques.

Abstract

El corpus de los Textos Matemáticos Griegos (CTMG) contiene un poco más de 100 obras y abarca todas las que han sobrevivido, completa o parcialmente, desde el s. IV AC. Se trata, pues, de un

corpus relativement pequeño. Nos hemos planteado el objetivo de digitalizar dicho corpus, así como tratar el corpus digitalizado con las herramientas de la Lingüística de Corpus. Dicho objetivo, por un lado, es factible, precisamente por tratarse de un corpus pequeño, pero también porque presenta pocas ambigüedades, el número de ‘palabras diferentes’ (ocurrencias) del corpus es bajo y las estructuras sintácticas diferentes no són muy abundantes. Además, la Matemática Griega está escrita en un lenguaje muy específico, del cual los matemáticos eran conscientes, ya que en último término, y formalmente, la matemática griega depende completamente del estilo en que se escribió; la matemática griega puede identificarse con esta forma de escribirla. Después de la digitalización de textos, hemos lematizado gran parte del corpus y, posteriormente, hemos hecho análisis comparativos entre diversos textos y autores. En este primer estadio de este proceso de digitalización y análisis, hemos comprobado que este enfoque cuantitativo en el estudio del CTMG es pertinente y necesario para profundizar en la Matemática Griega.

Estrella PÉREZ RODRÍGUEZ, *Le Lexicon Latinitatis Medii Aevi regni Legionis* (VIII^e s.-1230)

Résumé

Le *Lexicon Latinitatis Medii Aevi Regni Legionis*, ou *LELMAL*, est un dictionnaire de latin actuellement élaboré en Espagne à partir d'un corpus formé par les textes écrits principalement en langue latine sur le territoire du Royaume des Asturies et de León entre le VIII^e siècle et 1230. L'objectif principal de cet article réunit deux aspects : en premier lieu, montrer la méthodologie de ce travail lexicographique et les caractéristiques externes fondamentales du dictionnaire ; en second lieu, exposer et commenter quelques exemples intéressants tirés du corpus léonais qui démontrent l'importance de l'étude lexicographique pour mieux connaître l'histoire de la langue d'un territoire. À titre d'exemples, on a choisi quatre romanismes : *uentresca*, à peine attesté en castillan avant le XVIII^e siècle ; *jera*, un mot relatif à la façon de mesurer les terres ; les adjectifs apparentés *combo* et

recombo, seulement attestés dans les sources asturiennes ; et, pour finir, la forme insolite *plentum*, inconnue en latin et résultat vraisemblablement d'une confusion du scribe médiéval (ce que nous appelons un « mot fantôme »).

Abstract

The *Lexicon Latinitatis Medii Aevi Legionis* or *LELMAL* is a Latin dictionary which is being created in Spain from the sources written mainly in Latin in the kingdom of Asturias and León between the 8th century and 1230. The twofold objective of this paper is, on the one hand, to explain the methodology of that lexicographical work and the main external features of the dictionary; on the other hand, to study some interesting examples from the sources of León which can show the important contribution of lexicographical studies to the knowledge of the history of the language of a territory. Five examples have been chosen, four vernacular words: *uentresca*, hardly found in Castilian before the 18th century; *jera*, a word in relation with land measurement, and the related adjectives *combo* and *recombo*, only used in the sources from Asturias; as well as the unique form *plentum*, a ghost-word, as it is called, because it does not exist in Latin and probably originated from a mistake of the medieval scribe.

Gérard PETIT, Terminographie diachronique: le cas de la terminologie médiévale française

Résumé

L'objectif de cet article est de prolonger la réflexion sur la description du lexique et des terminologies en diachronie, mais aussi de présenter un projet lexicographique novateur consacré au français technique et scientifique médiéval: il s'agit de CréalScience. Les présupposés attachés usuellement à la représentation du lexique postulent chez celui-ci une stabilisation des formes, des significations et des régimes syntaxiques. Si une approche en synchronie peut s'appuyer sur la permanence (même relative) des données, il n'en va pas

de même pour une description diachronique, surtout lorsque la synchronie T-1 envisagée – le Moyen Âge – constitue à elle seule une vaste diachronie. Dans cette étude nous montrerons que : (i) les réglages théoriques et méthodologiques préalables à la description sont fondamentalement tributaires de l'écart diachronique entre To et T-1; (ii) la procédure de description, demandant à être adaptée à chaque synchronie passée, ne peut permettre une modélisation de la démarche ou de ses paramètres, sauf sous forme de schémas déclinables; (iii) la notion d'état de langue constitue un objectif pour le chercheur. Elle est néanmoins facteur de risques pour la description qui veut éviter l'anachronisme.

Abstract

The objective of this contribution is to extend the reflection on the description of the lexicon and terminology diachronic, but also to present an innovative lexicographical project devoted to medieval scientific and technical French: CréalScience. Presuppositions usually attached to the lexical representation postulate in this stabilization of forms, meanings and syntactic systems. If an approach in synchrony can rely on permanently (even relative) data, the question arises for a diachronic description, particularly when considered synchrony T-1 – the Middle Ages – is in itself a vast diachronic. In this study we show that: (i) pre-theoretical and methodological adjustments to the description are fundamentally dependent on the diachronic difference between To and T-1; (ii) a description of procedure, asking to be adapted to each past synchrony can enable modeling of the process or its parameters, except as series of patterns; (iii) the concept of state language is an objective for the researcher. Nevertheless, it constitutes a degree of risk for the description aiming to avoid anachronism.

Earl Jeffrey RICHARDS, À la recherche des communautés discursives au Moyen Âge: un regard numérique sur la connectivité dans la

culture vernaculaire et le rôle des traductions dans l'évolution de la prose en moyen français

Résumé

Cette communication propose une analyse de l'évolution de la prose médiévale en français avec l'aide de quatre méthodes numériques : la « piste Brepols », la diversité lexicale calculée grâce à AntConc, la stylométrie du logiciel StyloR et la visualisation d'un réseau de communautés discursives grâce au logiciel Gephi.

Est montrée d'abord l'importance de la latinité sous-jacente dans les *Serments* de Strasbourg et la *Cantilène Sainte Eulalie*, en recourant au moteur de recherche de la *Patrologia latina* et de la *Library of Latin Texts* de Brepols, permettant de reconstruire plus précisément l'influence du latin comme substrat ou adstrat dans n'importe quel texte vernaculaire, ce qui implique l'existence d'une communauté discursive dès le IX^e siècle. La survivance des formules légales latines dans les *Serments* semble en effet montrer, mais faiblement, l'existence d'une communauté discursive documentée par des bribes aussi éloquentes que fragmentaires.

Il s'agit ensuite de savoir si les traductions commanditées dans des contextes historiques connus favorisent l'expansion du vocabulaire français. Une analyse de la diversité lexicale au moyen du logiciel concordancier AntConc, à la suite d'une conversion de traductions d'époques diverses en fichiers .txt, permet de calculer les *token/type*-ratio. Les résultats préliminaires suggèrent que la diversité lexicale présentée par les œuvres en prose est nettement plus élevée que celle des œuvres en vers, c'est-à-dire que l'expansion du vocabulaire dépend en premier lieu du choix de la prose par l'auteur. Un autre résultat important est constitué par la différence entre la diversité lexicale des traductions faites pour Philippe le Bel et celle des œuvres composées pour Charles V. Pour expliquer cette différence, les fichiers .txt de plusieurs centaines de textes ont été soumis à une analyse stylométrique StyloR. Ce logiciel combine plusieurs

fonctionnalités basées sur la fréquence des mots, et produit à la suite d'une analyse *bootstrap* un fichier Excel qui sert de base à la visualisation d'un réseau au moyen du logiciel Gephi. La communication se clôt par un commentaire sur cette mise en évidence de communautés discursives à travers trois siècles en France et une comparaison avec la littérature en prose composée en moyen anglais.

Abstract

In this contribution I present an analysis of the rise of prose in medieval French with the help of four digital methods: the “*piste Brepols*” (literally the “Brepols track”: a method which entails translating medieval French expressions into Latin and using this translation in the search engine at the online Brepols Library of Latin Texts), lexical diversity calculated on the on-line concordance program “AntConc” (<http://www.laurenceanthony.net/software/antconc/>), stylometry based on the software “Stylo Package for R”, and the visualization of a network of discursive communities at the internet platform “Gephi”.

It seems important to investigate the lexical and syntactic relationships among these highpoints in order to identify how French prose developed in the late medieval period, especially in order to assess the role of Latin as both substratum and adstratum in the development of both spoken and written French. In the first part of my communication I will briefly show the important of the Latin substratum in the *Strasburg Oaths* and *Eulalie*. Using the *piste Brepols*, the method permits a more precise reconstruction of Latin's influence as adstratum and substratum in many other vernacular texts, implying the existence of a Latin-vernacular interfaces in a discursive community as early as the 9th century. The survival of Latin legal formulae in the *Oaths* suggests, if perhaps only faintly, the existence of such a discursive community documented by scraps that are as eloquent as they are fragmentary.

The next question is ascertaining whether translations commissioned by the royal court in well-known historical

contexts were responsible for lexical expansion in French. To answer this question, I first present calculations of lexical diversity from representative works. I have used the platform AntConc to calculate the token/type ratio as a measure of lexical diversity. Preliminary results suggest that the prose works exhibit a higher lexical diversity than works written in verse: in other words, lexical expansion depended in the first instance on the choice of prose over verse. Another important result of this research was ascertaining the difference between lexical diversity in translations commissioned by Philip the Fair and those commissioned by Charles V. In order to explain these differences, I have performed a stylometric analysis of several hundred medieval French texts (as txt-files) using the StyloR platform. The software, combining several functionalities calculates the statistical differences between authors and produces an Excel-file which can be visualized as a network on the Gephi platform. The contribution ends with a brief commentary on the existence of different discursive communities over a period of three centuries in late medieval France and a comparison with a similar visualization of Middle English prose works.

Xavier-Laurent SALVADOR, Fabrice ISSAC et Marco FASCIOLO, *Herméneutique des similarités dans le DFSM: une expérience*

Résumé

L'avènement de l'informatique a engendré une double révolution pour la dictionnaire. Tout d'abord du point de vue des méthodologies, l'utilisation systématique de corpus numériques pour l'élaboration du *Trésor de la langue française (TLF)* en est un exemple, mais aussi, de manière moins massive cependant, en ce qui concerne les interfaces de consultation proposées aux utilisateurs.

Il existe de nombreux dictionnaires en ligne, de natures très diverses : dictionnaires, glossaires, spécialisés ou non, structurés ou non. Les outils et les ressources proposés ont tous la même forme : une base de données plus ou moins complexe associée à

une interface proposant un ou plusieurs outils de consultation ou de recherche. La grande majorité de ces applications se focalisent sur la mise à disposition de ressources linguistiques plus ou moins structurées. Le processus de constitution est totalement déconnecté du processus de consultation. Le principe – ou scénario – le plus fréquemment rencontré en terme d'interface est un calque, une transposition, plus ou moins réussi de l'utilisation des dictionnaires « papier ». Dans ce schéma l'utilisateur final est paradoxalement oublié et les possibilités offertes par l'ordinateur sous-exploitées, alors que parallèlement la masse d'informations proposée a considérablement augmenté.

Afin de pallier cette absence de *continuum*, nous avons développé un outil dictionnaire appelé Isilex, dont l'objectif est d'assister aussi bien les lexicographes dans l'élaboration du dictionnaire que les utilisateurs finaux pour le consulter. Notre présentation s'appuiera en grande partie sur le projet CréaLScience, dont l'objectif est de construire un dictionnaire du français scientifique médiéval. Nous présenterons les différents modules utilisés par l'ensemble des acteurs, les interfaces et les outils développés spécifiquement.

Abstract

The rise of academic computing has provoked a double revolution in lexical research. From the perspective of methodology, the systematic use of digital corpora in the creation of the *Trésor de la langue française (TLF)* is the first example of this revolution, and secondly as well, though in a less extensive manner, the kinds of interfaces available for readers consulting this on-line dictionary.

There are, of course, many on-line dictionaries, of highly different natures: dictionaries, glossaries, specialized or general. The tools and resources available all follow the same format: a more or less complex databank linked to a graphic user interface with one or many tools for consultation and research. The lion's share of these applications are focused on making more or less structured resources available for consultation.

The most frequently encountered principle or scenario as far as interfaces are concerned follows a transposed format, more or less successful, of hard-copy dictionaries. This format, however, paradoxically forgets the reader while at the same time under-exploiting the possibilities of a web-based environment which has vastly increased the amount of consultable data.

In order to remedy this rupture between hard-copy and on-line web-based dictionaries, we have developed a lexical tool called “Isilex” whose purpose is to help both lexicographers in expanding the dictionary as well as ordinary readers consulting it. Our presentation is based on the larger project CréaLSscience whose goal is to construct a dictionary of medieval scientific French. We present different modules used by both lexicographers and readers and the interfaces and tools specifically developed for them.

COMITÉ SCIENTIFIQUE

Hava BAT-ZEEV SHYLDKROT (Université de Tel Aviv)
Françoise BERLAN (Université Paris-Sorbonne)
Mireille HUCHON (Université Paris-Sorbonne)
Peter KOCH (Universität Tübingen)†
Anthony LODGE (Saint Andrews University)
Christiane MARCHELLO-NIZIA (École normale supérieure-LSH, Lyon)
Robert MARTIN (Université Paris-Sorbonne/Académie des inscriptions
et belles-lettres)
Georges MOLINIÉ (Université Paris-Sorbonne)†
Claude MULLER (Université Bordeaux Montaigne)
Laurence ROSIER (Université Libre de Bruxelles)
Gilles ROUSSINEAU (Université Paris-Sorbonne)
Claude THOMASSET (Université Paris-Sorbonne)

COMITÉ DE RÉDACTION

Claire BADIOU-MONFERRAN (Université de Lorraine)
Michel BANNIARD (Université Toulouse 2-Le Mirail)
Annie BERTIN (Université Paris Ouest Nanterre La Défense)
Claude BURIDANT (Université Strasbourg 2)
Maria COLOMBO-TIMELLI (Université Paris-Sorbonne)
Bernard COMBETTES (Université de Lorraine)
Frédéric DUVAL (École nationale des chartes)
Pierre-Yves DUFEU (Université Aix-Marseille 3)
Amalia RODRIGUEZ-SOMOLINOS (Universidad Complutense de Madrid)
Philippe SELOSSE (Université Lyon 2)
Christine SILVI (Université Paris-Sorbonne)
André THIBAUT (Université Paris-Sorbonne)

COMITÉ ÉDITORIAL

Olivier SOUTET (Université Paris-Sorbonne), Directeur de
la publication
Joëlle DUCOS (Université Paris-Sorbonne-EPHE), Trésorière
Stéphane MARCOTTE (Université Paris-Sorbonne), Secrétaire de rédaction
Thierry PONCHON (Université de Reims Champagne-Ardenne), Secrétaire
de rédaction
Antoine GAUTIER (Université Paris-Sorbonne), Diffusion de la revue

Table des matières

Présentation	
Joëlle Ducos	7
À propos du <i>DMF</i> :	
réussites et pièges de la lexicographie électronique	
Robert Martin	11
De la gestion de la variation en moyen français à son élargissement aux états anciens du français : les développements du lemmatiseur LGeRM	
Sylvie Bazin-Tacchella & Gilles Souvay	25
Herméneutique des similarités dans le <i>DFSM</i> : une expérience	
Xavier-Laurent Salvador, Fabrice Issac & Marco Fasciolo	49
Le <i>Lexicon Latinitatis Medii Aevi Regni Legionis</i> (VIII ^e siècle-1230) : caractéristiques et quelques exemples (<i>ventrescas, iera, cumbo, plentum</i>)	
Estrella Pérez Rodríguez	77
La lexicographie de l'italien médiéval et les corpus de l'OVI : un bilan provisoire et quelques nouvelles perspectives	
Elisa Guadagnini	101
Le latin médiéval du <i>Glossarium Mediae Latinitatis Cataloniae</i> : un projet lexicographique dans un contexte européen	
Ana Gómez Rabal	121
Autorité du latin et transparence constructionnelle : le sort des néologismes médiévaux dans le domaine médical	
Michèle Goyens & Céline Szecl	141
Base de français médiéval : une base de référence de sources médiévales ouverte et libre au service de la communauté scientifique	
Céline Guillot, Serge Heiden & Alexei Lavrentiev	167

Terminographie diachronique : le cas de la terminologie médiévale française Gérard Petit	185
Numérisation et traitement de textes mathématiques grecs : méthodes, problèmes et résultats Ramon Masià	213
À la recherche des communautés discursives au Moyen Âge : un regard numérique sur la connectivité dans la culture vernaculaire et le rôle des traductions dans l'évolution de la prose en moyen français Earl Jeffrey Richards	229
Résumés / Abstracts	249
Comité scientifique	267
Table des matières	269