

REVUE DE
LINGUISTIQUE
FRANÇAISE
DIACHRONIQUE

7
2017

DIACHRONIQUES

LES ÉTATS ANCIENS
DES LANGUES À L'HEURE
DU NUMÉRIQUE

Gómez Rabal – 979-10-231-2162-9



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 FASCIOLO

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 MASIÀ

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

Ana Gómez Rabal

CSIC

Institution Milá y Fontanals (Barcelone, Espagne)

La réalisation de tout projet lexicographique peut susciter l'intérêt plus ou moins vif des spécialistes, mais aussi des personnes intéressées par des matières diverses. Nous voulons présenter ici un projet fondé et réalisé par des philologues (spécialistes de la langue latine, et en particulier médiévale) dont le but dépasse le cercle de la philologie *stricto sensu*. Il s'agit d'un dictionnaire, le *Glossarium Mediae Latinitatis Cataloniae ab anno DCCC usque ad annum MC (GMLC)*, résultat de l'étroite collaboration entre deux institutions, le Consejo superior de investigaciones científicas (CSIC) – le « Conseil supérieur de la recherche scientifique » – et l'université de Barcelone, et réalisé dans le centre du CSIC qui répond au nom d'Institution Milá y Fontanals, à Barcelone (Espagne). Ce dictionnaire, remarquons-le, veut intéresser non seulement les philologues, mais aussi les historiens, les juristes et toute personne attirée par le Moyen Âge, d'où le désir de faire de cette œuvre un instrument de référence indispensable tout aussi bien, par exemple, pour le latiniste qui aborde l'étude d'un champ lexical déterminé du latin médiéval hispanique, que pour le romaniste qui examine des problèmes de la grammaire historique ou pour le juriste qui analyse les formes d'élaboration des testaments.

La réalisation d'une ambition lexicographique comme celle qu'incarne le *GMLC* s'inscrit dans un milieu, le milieu européen, où fleurissent de nombreux dictionnaires, soit nationaux soit

régionaux, de latin médiéval et où s'est forgé un projet qui se maintient dans toute sa vigueur, celui d'un dictionnaire européen commun de latin médiéval. Le *GMLC* suit, donc, les voies tracées en Europe par la philologie latine médiévale et la lexicographie et doit affronter, en toute logique, les mêmes défis que d'autres projets jumeaux.

Pour ces deux raisons – parce que le *GMLC* veut être un instrument utile aux médiévistes quel que soit leur champ d'études, et aussi parce que la trajectoire du glossaire est parallèle à celle d'autres projets européens – il nous semble indispensable de donner, tout d'abord, un aperçu rapide de l'histoire du glossaire et de revenir, ensuite, sur les phases de travail actuellement développées.

Présentation

Le *Glossarium Mediae Latinitatis Cataloniae ab anno DCCC usque ad annum MC* prétend, nous l'avons indiqué, offrir aux philologues, historiens, juristes et, en général, à toute personne intéressée par le Haut Moyen Âge, la documentation latine écrite en Catalogne du IX^e au XII^e siècle. C'est ce que souligne le sous-titre de l'édition imprimée du glossaire : *Voces latinas y romances documentadas en fuentes catalanas del año 800 al 1100*¹, « Mots latins et romans documentés à partir de sources catalanes de l'année 800 à l'année 1100 ».

Ces « mots latins et romans » sont recueillis, étudiés et systématisés à partir, essentiellement, de la lecture et du dépouillement de documents notariés, à savoir des actes de donation, de dotation, de vente, d'acquisition, des testaments, des serments de fidélité et des litiges, dont la production est une constante très riche dans les territoires correspondant aux domaines linguistiques du catalan à partir du IX^e siècle, pendant le Haut Moyen Âge².

1. Voir Mariano Bassols et Joan Bastardas (1960-1985) et Joan Bastardas (2006).

2. Concernant les éditions de ces chartes et des autres œuvres correspondant à l'époque et aux territoires cités, voir Ana Gómez Rabal (2008).

On doit souligner cette abondance face à la presque inexistence dans la Catalogne de l'époque d'autres types de textes légaux, les textes de création législative, à une seule exception près, celle des textes légaux et des textes de droit coutumier locaux qui s'inspirent du *Liber Iudicum* wisigothique³.

Il faut considérer également cette abondance de chartes juridiques au regard de la rareté de tout autre genre de « littérature » dans le sens le plus large du terme, c'est-à-dire celui qui englobe les belles lettres, les œuvres historiques, philosophiques, les traités de rhétorique ou les livres techniques. Parmi ces rares exceptions avant le milieu du XII^e siècle, nous devons citer les écrits littéraires de l'abbé Oliba (ca. 970-1046),⁴ la *Vita Beati Petri Vrseoli* et l'*Epistola Garsiae monachi Cuxanensis*.⁵ Une autre exception remarquable : les textes techniques et scientifiques écrits par Lupitus, archidiacre de Barcelone, ou attribués à lui⁶.

-
3. Le *Liber Iudicum*, ou *Liber Iudiciorum*, est le recueil sur lequel s'est basée la pratique juridique haut-médiévale catalane et qui a servi de modèle pour l'élaboration des textes légaux et des textes de droit coutumier locaux. Voir Josep Maria Font i Rius (1969 et 1983). Sur l'application du *Liber Iudicum* en correspondance avec les canons conciliaires et les capitulaires carolingiens, se référer à Font i Rius (2003, en particulier p. 71-72).
 4. Entre eux, toujours dans un style soigné, on compte une assez riche collection de lettres, dont seulement sept pièces ont été conservées, deux d'entre elles adressées à ses moines de Ripoll, deux à Sancho III de Pampelune (1000-1035), une à Gauzlin, archevêque de Bourges et abbé de Fleury, et une encyclique à d'autres monastères annonçant la mort (en 1020) de son frère le comte Bernard ; et on compte aussi des poèmes, parmi lesquels un éloge historique de Ripoll (ca. 1032), intégrant un résumé de son passé rédigé en hexamètres, des louanges à certains membres de la famille comtale des Oliba, et un poème élogieux également adressé à son ami l'abbé de Fleury. Pour l'édition de ces textes, se référer à Eduard Junyent i Subirà (1992).
 5. La première de ces deux œuvres, composée entre 1075 et 1100, a été éditée par Jean Mabillon (1737). Pour une édition de la deuxième œuvre, écrite entre 1040 et 1046, voir Petrus de Marca (1688).
 6. Ce petit corpus technico-scientifique contient diverses œuvres portant sur un sujet concret, l'astrolabe, et sur d'autres questions de géométrie ou d'arithmétique. Pour l'édition de ces traités, voir José Maria Millàs Vallicrosa (1931) ; en particulier sur l'astrolabe, se référer aux p. 271-275 (Lupitus Archidiaconus, *Prologus in libellum de astrolabio*, ca. 980, manuscrit du XI^e siècle), p. 275-293 (*Sententie astrolabii*, traduction d'un original arabe du X^e siècle attribué, par conjecture, à Lupitus de Barcelone), p. 293-295 (Anonymus, *De mensura astrolapsus*, X^e siècle, manuscrit du XI^e siècle), p. 304-305 (Anonymus, *Regulae de astrolabio*, X^e siècle, manuscrit du XI^e siècle) et p. 308-315 (Anonymus, *De astrolabio*, X^e-XI^e siècles).

Ce sont tous ces textes, qu'ils relèvent d'un type ou d'un autre (juridique, littéraire, scientifique), produits entre les années 800 et 1100, qui ont été dépouillés et insérés dans les fichiers du *GMLC*, et qui forment le corpus des textes servant de base pour l'élaboration du dictionnaire. Mais c'est surtout la documentation qui reproduit les actes juridiques – la documentation notariée – qui concentre l'intérêt des lexicographes du *GMLC*, parce qu'elle se convertit en un témoignage écrit privilégié de la langue latine médiévale; expression du langage juridique, ecclésiastique, institutionnel et curial du Haut Moyen Âge, époque où affleurent des innovations lexicales importantes et des indices très clairs, de nature phonétique ou morpho-syntaxique, d'une langue romane qui commence à percer. Ce latin documentaire présente souvent, même pour un bon latiniste, de grandes difficultés quant à l'interprétation, car deux phénomènes s'entrecroisent: le latin subit l'influence plus ou moins forte de la langue romane et, parallèlement, la langue romane se latinise, et apparaît fréquemment comme le fruit d'un effort de latinisation d'expressions et de formes déjà pleinement romanes. Le *GMLC*, constitué d'articles lexicographiques élaborés surtout à partir des documents notariés que nous venons d'évoquer, cherche à aplanir le chemin ouvert à ceux qui souhaitent affronter les difficultés de cette langue.

Histoire

Le Glossarium Mediae et Infimae Latinitatis de Charles Dufresne, sieur Du Cange, et le Novum Glossarium Mediae Latinitatis

Le Glossarium Mediæ et Infimæ Latinitatis de Charles Dufresne, sieur Du Cange, est assurément le premier parmi les ouvrages de référence que tout médiéviste, quel que soit le domaine de recherche qu'il a choisi, s'est inmanquablement vu dans la nécessité de consulter. Cette œuvre lexicographique essentielle – connue comme « le Du Cange » –, publiée en 1678, et dont les rééditions successives jalonnent les XVIII^e et XIX^e siècles,

constitue un point de départ inéluctable pour toute tentative de compilation minutieuse de la langue latine médiévale.

La nécessité de créer un nouveau dictionnaire du latin du Moyen Âge répondant à des critères scientifiques modernes devint, au début du xx^e siècle, de plus en plus évidente. Ce vœu fut formulé au cours d'un congrès d'histoire tenu à Londres en 1913 ; bientôt le projet se dessina et, en 1920, il reçut l'appui de l'Union académique internationale. Répondant à la désignation officielle de *Novum Glossarium Mediae Latinitatis* (NGML), il reste connu par les philologues et les historiens comme le « nouveau Du Cange ».

Le travail de compilation, de lecture et de dépouillement des textes commença en 1924, simultanément dans six pays : la Belgique, la France, le Royaume-Uni, l'Italie, les Pays-Bas et la Pologne. Chacun d'eux constitua peu à peu son propre fichier, et le premier travail fut rendu public en 1936 avec le premier fascicule (*a-agradior*) du *Latinitatis Italicæ Medii Aevi Lexicon Imperfectum*, dirigé par Francesco Araldi et publié dans l'*Archivum Latinitatis Medii Aevi* (ALMA). En 1953 le comité polonais fit paraître le premier fascicule du *Lexicon Mediae et Infimæ Latinitatis Polonorum*, dirigé par Marian Plezia ; en 1975, le comité britannique publia à son tour la première livraison de son *Dictionary of Medieval Latin from British Sources*, dirigé par Ronald Edward Latham. Dans les années 1930, de nouvelles équipes de travail s'incorporèrent au projet, en Suède, en Hongrie et en Tchécoslovaquie, et les académies de Munich et Berlin, en 1959, entreprirent la publication de leur *Mittellateinisches Wörterbuch*, à la charge d'Otto Prinz. D'autres pays se joignirent à l'entreprise, tels la Finlande, le Danemark, l'Irlande ou la Yougoslavie.

Mais l'apparition en 1957 du premier fascicule du *Novum Glossarium Mediae Latinitatis*, correspondant à la lettre L, à la charge de Franz Blatt, entérina définitivement la marche de cette entreprise commune qui maintient aujourd'hui tout son dynamisme. Franz Blatt précisait dans son « Avis au lecteur » le sens et la portée de cette œuvre :

Le nouveau dictionnaire du latin médiéval ne prétend nullement remplacer le *Glossarium mediae et infimae latinitatis conditum a C. du Fresne domino* Du Cange, celui-ci étant à la fois une encyclopédie et un dictionnaire; nous avons, par conséquent, réduit au maximum les explications d'ordre historique et technique. Le point de vue purement lexicologique prévaut.

Et, un peu plus bas :

Le nouveau dictionnaire contient [...] non seulement les néologismes médiévaux, mais aussi les mots et sens classiques encore vivants, le but du dictionnaire étant de donner une description sinon complète, du moins fouillée de la langue latine au Moyen Âge.

Actuellement le Comité Du Cange, dont le siège est à Paris, poursuit la rédaction du *NGML*. Le comité est régi par l'Académie des inscriptions et belles-lettres, par le Centre national de la recherche scientifique (CNRS) et par l'École pratique des hautes études (EPHE, 4^e section de sciences historiques et philologiques), et constitue la section de lexicographie latine de l'Institut de recherche et d'histoire des textes (CNRS)⁷.

Le Glossarium Mediae Latinitatis Cataloniae ab anno DCCC usque ad annum MC

À Barcelone, un groupe de professeurs réunis sous la direction de Mariano Bassols de Climent, et parmi lesquels se trouvait Joan Bastardas, décida au cours de l'année scolaire 1952-1953 de se joindre au projet du *Novum Glossarium (NGML)*. Dans le but de donner davantage de matière au *NGML*, le groupe entama un travail de dépouillement de textes latins médiévaux hispaniques – spécialement, mais non exclusivement catalans – et d'élaboration des fiches correspondantes. Tout cela se fit à partir d'éditions des textes documentaires bien que, lorsque la rigueur scientifique l'exigeait, on eut recours à la lecture directe des documents.

7. Concernant l'histoire, l'organisation et le but scientifique poursuivi par l'équipe de lexicographie latine médiévale, voir en ligne : <http://www.irht.cnrs.fr/fr/recherche/sections/lexicographie-latine>, et : <http://www.aibl.fr/travaux/moyen-age/article/le-bureau-du-cange?lang=fr> [consultés le 21 juin 2017].

Bientôt cependant le groupe envisagea la possibilité de publier, en mettant à profit le travail de dépouillement réalisé et qui continuait à se réaliser, son propre glossaire. C'était l'occasion tout d'abord de recueillir le lexique qui pourrait être exclu d'un dictionnaire plus général, car il était logique qu'entre les rédacteurs du *NGML* primât le critère de choisir des exemples communs à tout l'Occident européen; d'où la perte de certaines nuances quant aux différences et caractères particuliers du latin des divers territoires, où peu à peu apparaissaient des traits socioculturels et linguistiques spécifiques.

D'autre part, proposer aux responsables du *NGML* un corpus non seulement élaboré, mais aussi publié semblait beaucoup plus fructueux que de leur remettre des fiches qui pourraient présenter des difficultés d'interprétation. Autrement dit, on considéra qu'il était souhaitable d'offrir au *NGML* des matériaux révisés par des chercheurs connaissant la langue, l'histoire, les conditions sociales, économiques, institutionnelles, la culture et les coutumes du territoire.

Enfin, au cours de l'année scolaire 1956-1957, époque où l'École de philologie de Barcelone du Conseil supérieur de la recherche scientifique disposa, enfin, de locaux appropriés, il fut décidé d'entreprendre la publication d'un glossaire à partir des sources documentaires catalanes, surtout notariées, et dont les limites chronologiques iraient de l'année 800 à l'année 1100. Les premiers articles furent soumis à des romanistes et latinistes européens qui conseillèrent à l'équipe de poursuivre ses travaux, et celle-ci publia en 1960 le premier fascicule (*a-aragalius*) du *GMLC*.

C'est le professeur Mariano Bassols de Climent qui fut à l'initiative de la création et de la première organisation du *GMLC*, mais l'entreprise et la réalisation des articles lexicographiques ont été, dès le début, de la responsabilité du professeur Joan Bastardas.

Dans les années 1960, le projet initié par l'École de philologie latine de Barcelone paraissait donc parfaitement viable, puisque la documentation latine notariée des siècles compris entre le IX^e et le XII^e est en Catalogne, nous l'avons souligné, abondante, riche et suffisamment variée.

Mais l'intérêt de l'entreprise était aussi conforté par d'autres tentatives, antérieures, d'étude et de systématisation des caractéristiques du latin du Haut Moyen Âge en Catalogne. L'antécédent le plus important, déjà lointain mais source d'une inspiration réelle pour le nouveau projet, était le corpus de matériel lexicographique rassemblé par Josep Balari i Jovany (1844-1904). Les collections de fiches (16 000 fiches) écrites et compilées par Balari i Jovany furent la base sur laquelle ce chercheur composa son œuvre *Orígenes històrics de Catalunya* (*Les Origines historiques de la Catalogne*, Barcelone, 1899). En 1956, le professeur Joaquim Carreras i Artau hérita de ces fiches, et les transmit au professeur Bastardas. Elles sont un point de repère exceptionnel pour l'équipe du *GMLC*.

Réalisation du *Glossarium Mediae Latinitatis Cataloniae*

L'équipe de rédaction dirigée par le professeur Bastardas poursuivit, toujours dans la même orientation, son travail jusqu'à compléter la publication du premier volume en 1985. Ce volume, qui correspond aux lettres A, B, C et D, contient une préface, une introduction générale, une bibliographie et plus de 1 000 articles offrant une ample information linguistique sur le latin médiéval et le catalan pré-littéraire, articles enrichis par des notes à caractère archéologique, historique et culturel.

Au cours des années 1990, les possibilités qu'offrirent les nouveaux moyens techniques ouvrirent la voie à l'élaboration d'un fichier informatisé. Ce nouveau pas fut jugé opportun et utile. En vue de la rénovation du programme – continuer le travail de dépouillement des documents et de rédaction des articles et parallèlement commencer le travail d'informatisation du fichier manuel constitué par 50 000 fiches papier – les accords établis entre l'Institution Milá y Fontanals du CSIC et l'université de Barcelone et, depuis 1985, entre ces deux centres et l'Institut d'études catalanes se sont renforcés. Il faut ajouter à cela les subventions ministérielles obtenues sous l'impulsion de Pere J. Quetglas Nicolau, professeur du Département de

philologie latine de l'université de Barcelone et successeur du professeur Bastardas comme directeur de l'équipe.

État actuel du projet

La réalisation du *GMLC* est actuellement structurée autour de deux phases de travail. La première correspond à sa rédaction, et la seconde, à sa numérisation.

Pour cette seconde phase, la direction de l'équipe a tout d'abord pensé, nous l'avons dit, à informatiser le fichier manuel, mais dès les années 2001-2002 le groupe de travail prit une décision différente et s'assigna un autre but : celui de numériser la base documentaire constituée par les éditions des textes, sans négliger les critères de qualité dans la sélection des éditions. On projette de créer ainsi, sur un support digital, un fichier riche, inédit, ouvert aux variations et aux élargissements (exigés, sans doute, par le dépouillement de nouvelles publications documentaires), cohérent, intelligible et facile à manier. Le fait de numériser permet, en outre, la création et l'utilisation de concordances – ou index alphabétiques de mots, d'unités lexicales, d'un texte avec leur contexte significatif – qui confirment, complètent et élargissent l'information recueillie dans les fiches sur lesquelles travaillent les rédacteurs du *GMLC*. Le corpus de textes corrigés à partir d'éditions et de transcriptions est constitué actuellement par plus de 23 000 documents.

Les deux phases de travail – rédaction et numérisation – sont complémentaires et indissociables. Ce tout s'est matérialisé, d'une façon encore très partielle, par la publication du fascicule 11, qui correspond à la lettre F (2001); d'une façon un peu plus complète, par la rédaction et la publication du fascicule 12, qui correspond à la lettre G (2006); d'une façon beaucoup plus systématique, par la préparation de la réédition des lettres A-D (prêtes pour publication papier en 2010) et, surtout, par la préparation des articles du fascicule E⁸.

8. Rédaction achevée en 2016.

Mais, en ce qui concerne cette phase de numérisation, nous devons souligner surtout que l'équipe est en train de convertir le fichier numérique, d'usage interne jusqu'en 2011, en un fichier accessible à toute personne intéressée par la documentation latine produite tout au long du Haut Moyen Âge en Catalogne. En tant que coordinatrice de ce projet, et sous la direction du professeur Quetglas, je me suis spécialement chargée de diriger une partie des efforts de l'équipe vers le but annoncé, c'est-à-dire le dépouillement, l'édition et la publication de notre corpus de textes; pour le dire autrement, vers la publication d'une base de données lexicale de consultation publique, le *Corpus Documentale Latinum Cataloniae* (CODOLCAT)⁹.

Ce projet de création d'un nouveau service de consultation externe de la base de données lexicale numérisée a obéi au désir d'utiliser les avantages des nouvelles technologies et s'inscrit dans la lignée d'un autre projet espagnol: le *Corpus Documentale Latinum Gallaeciae* (CODOLGA), dirigé par le docteur José Eduardo López Pereira, professeur à l'université de la Corogne¹⁰. Ce projet pionnier a servi d'inspiration et de modèle pour la conception du CODOLCAT. Le CODOLCAT est une plate-forme multilingue qui offre la possibilité de procéder à des requêtes simples ou d'affiner la recherche par type de producteur du document, par localisation, selon un éventail chronologique, etc.; les résultats sont affichés sous la forme de concordances. Nous avons publié cinq versions du CODOLCAT. La première d'entre elles (v. 1, 2012) permettait l'accès à environ 1000 actes, correspondant à 4 cartulaires; la deuxième version (v. 2, 2013) était constituée par environ 2 400 actes, correspondant à 10 cartulaires; la troisième version (v. 3, 2014), par environ 3 600 actes, correspondant à 14 cartulaires; la quatrième (v. 4, 2015) met à la portée du lecteur 5 200 actes, correspondant à 17 cartulaires; la cinquième (v. 5, 2016), 6 300 actes de 21 cartulaires. Notre intention est de continuer à publier une version annuelle avec un incrément approximatif de 1 000 chartes, au minimum, par an.

9. En ligne : <http://gmlc.imf.csic.es/codolcat>.

10. En ligne : <http://corpus.cirp.es/codolga> (v. 12, 2015) [consulté le 18 octobre 2016].

Il existe actuellement en Espagne un troisième projet d'élaboration et de publication d'un corpus numérisé suivant le même modèle : il s'agit du *Corpus Documentale Latinum Valencie* (CODOLVA), sur la documentation du royaume de Valence et, au Portugal, du *Corpus Documentale Latinum Portucalese* (CODOLPOR), de l'université de Lisbonne. Il existe d'autre part en Espagne une autre équipe de lexicographie médiolatine, qui s'occupe de la documentation des royaumes de León et de Castille et a déjà publié son *Lexicon Latinitatis Medii Aevi Regni Legionis imperfectum (s. VIII-1230)*, sous la direction des professeurs Maurilio Pérez González (université de León) et Estrella Pérez Rodríguez (université de Valladolid).

Pour l'équipe du *GMLC*, la continuation de la préparation et de la publication du glossaire ainsi que la poursuite de l'édition et de la publication annuelle du CODOLCAT constitueront toujours deux objectifs inséparables.

Défis pour le *GMLC* et pour la lexicographie latine médiévale

L'équipe du *GMLC* prépare la publication numérique des fascicules déjà rédigés¹¹. L'accès à cette publication s'effectuera librement par internet. La préparation de cette publication constitue un enjeu concret pour notre groupe en raison de la compatibilité des deux outils que nous voulons offrir : la base de données lexicale (le CODOLCAT) et le dictionnaire lui-même. Voyons les points sur lesquels va se structurer cette compatibilité recherchée :

1. Dès le service de requête du dictionnaire, toutes les données contenues dans le CODOLCAT (lexicales ou autres) doivent être récupérables ; et, à l'inverse, toute l'information balisée et structurée – donc indexée – que contient le *GMLC* devra être récupérable quand l'utilisateur interrogera le CODOLCAT.
2. La plate-forme sera flexible : nous sommes en train de créer des formulaires d'introduction des données destinés aux

11. La préparation de l'édition numérique a été commencée à partir des articles lexicographiques revus pour la réédition papier des lettres A-D.

rédacteurs. Ces formulaires rendront plus aisées l'édition et la publication des articles de l'œuvre.

3. S'agissant de l'apparition des exemples, un dictionnaire est forcément sélectif. Dans le CODOLCAT, au contraire, le lecteur pourra accéder à tous les exemples.
4. Il sera possible de réaliser des requêtes et des recherches qui ne seront pas strictement lexicales. Quand nous introduisons les données dans le CODOLCAT, nous recueillons beaucoup d'informations sur les archives ou les bibliothèques où se conservent les chartes et les manuscrits, sur le type d'acte juridique concerné, sur l'écriture, etc. Il sera possible de répondre à des questions posées par des historiens, à des interrogations prosopographiques, géographiques, etc.

Mais notre objectif va plus loin encore : il veut dépasser cette cohérence acquise entre la base de données et le dictionnaire. Notre souhait, notre ambition est d'obtenir la compatibilité avec les autres dictionnaires de latin médiéval et de pouvoir offrir – comme les créateurs du *Novum Glossarium* l'exprimaient – un moyen de consultation commune pour le latin médiéval dans toute sa diversité, géographique, stylistique, même chronologique¹². Il ne s'agit nullement de créer un serveur unique, mais une plateforme commune où chaque groupe déposerait les données qu'il déciderait lui-même de partager. La technologie nous offre maintenant cette possibilité. Le besoin, l'exigence même d'une collaboration entre les équipes – et cela malgré les différences dans les buts déjà atteints et malgré la diversité des méthodes et des moyens employés – correspondent désormais à un désir objectivement réalisable¹³.

12. Si le *Novum Glossarium* a pour limites chronologiques les années 800 et 1200, les divers projets européens de lexicographie latine du Moyen Âge correspondant au domaine linguistique d'une langue (romane ou non romane) ont des marges chronologiques très différentes : ainsi, pour ne donner que deux exemples, le *Dictionary of Medieval Latin from British Sources*, de l'année 400 à l'année 1450 ; le CODOLVA, de l'année 1200 à l'année 1350.

13. Cet article a été rédigé au sein du projet « Informatización del *Glossarium Mediae Latinitatis Cataloniae* (8) » : « Ampliación y desarrollo de la base de datos *Corpus Documentale Latinum Cataloniae* (2) » (FF12016-77831-(2-1-P)), financé par le Ministère espagnol de l'Économie.

Références bibliographiques

Textes

- BASSOLS, Mariano et BASTARDAS, Joan (dir.), *Glossarium Mediae Latinitatis*, t. I (A-D), Barcelona, CSIC/Universidad de Barcelona, 1960-1985.
- BASTARDAS, Joan (dir.), *Glossarium Mediae Latinitatis Cataloniae*, fasc. 11 (F) et fasc. 12 (G), Barcelona, CSIC, 2006.
- FONT I RIUS, Josep Maria, *Cartas de població y franquicia de Cataluña*, 2 t., vol. I: *Estudio. Diplomatario. Presentación monográfico-local e índices*, Madrid/Barcelona, CSIC, 1969.
- , *Cartas de població y franquicia de Cataluña*, t. II, *Estudio. Apéndice al vol. I*, Madrid/Barcelona, CSIC, 1983.
- , « L'escola jurídica de Barcelona », dans ALTURO, Jesús, BELLÉS, Joan, FONT I RIUS, Josep Maria et al., *Liber Iudicum popularis. Ordenat pel jutge Bonsom de Barcelona*, Barcelona, Generalitat de Catalunya, 2003, t. I, p. 67-100.
- GÓMEZ RABAL, Ana, « En torno a las ediciones de la documentación latina catalana altomedieval », *Archivum Latinitatis Medii Aevi (Bulletin Du Cange)*, n° 66, 2008, p. 355-366.
- JUNYENT I SUBIRÀ, Eduard, *Diplomatari i escrits literaris de l'abat i bisbe Oliba* (éd. A.M. Mundo), Barcelona, Institut d'Estudis Catalans, 1992.
- MABILLON, Jean, *Acta Sanctorum Ordinis Sancti Benedicti in saeculorum classes distributa (Saeculum V)*, Venetiae, 1737, vol. VII, p. 851-860.
- MILLÀS VALLICROSA, José Maria, *Assaig d'història de les idees Físiques i matemàtiques a la Catalunya medieval*, Barcelona, 1931.
- PÉREZ GONZÁLEZ, Maurillo (dir.), *Lexicon Latinitatis Medii Aevi Regni Legionis imperfectum (s. VIII-1230)*, Turnhout, Brepols, 2010.
- PETRUS DE MARCA, *Marca Hispanica sive limes Hispanicus, hoc est geographica et historica descriptio Cataloniae, Ruscinonis et circumjacentium populorum ab annu 817 ad annum 1258*, Paris, 1688, ap. 222, cd. 1072-82.

Ressources en ligne

Corpus Documentale Latinum Cataloniae (CODOLCAT), QUETGLAS, Pere J. (dir.), GÓMEZ RABAL, Ana (coord. éd.). En ligne : <http://gmlc.imf.csic.es/codolcat> (V.S, 2016).

Corpus Documentale Latinum Gallaeciae (CODOLGA), LÓPEZ PEREIRA, José Eduardo (dir.). En ligne : <http://corpus.cirp.es/codolga> (v. 12, 2015) [consulté le 18 octobre 2016].

ALMUNIA, Prædii rustici species apud Hispanos. Charta Sanctii Regis Aragonum æræ 1132. apud Martinezum in Hist. Pinnatensi lib. 3. cap. 9: *Concedo prædicto Cænobio, ... illam meam Almuniam, vocatam Daymus, quæ afrontat ex una parte cum Torredellas, etc.* Alia Alfonsi VI. Regis æræ 1133. apud Anton. de Yezep in Chronico Ord. S. Benedicti tom. 6: *In quo loco incipit alia via, per quam descendunt usque in viam publicam super Almuniam Regis, etc.* Observantiæ Regni Aragon. lib. 5. tit. de Jure dotium, § 4: *Tamen si Miles vel Infantio habent unam Almuniam vel turrim, censetur una hæreditas cum toto hæreditamento adjuncto illi Almunia, vel turri.* Adde Colmenarezium in Hist. Segobiensi cap. 16. § 4.

In his tribus exemplis, excepto forte postremo, per *Almunia* intelligi potest hortus, ut intelligendus est dubio procul in sequentibus; Testamentum I. Adefhonsi Regis Hispaniæ apud Marten. tom. 1. Collect. Ampliss. col. 546. C: *Et ut hi qui in eadem Ecclesia superscripta permanserint, supplementum aliquod victui habere possint, offero illis villam unam nomine Huleka, et unam Almuniam, quam nos Latine vocamus Ortum, qui est prope illam Ecclesiam S. Servandi.* Et in Archivo S. Victoris Massiliensis armor. Hispan. n. 115: *Almunia Regis, quam nos Latine vocamus Ortum.* Lusitanis *Almuinha*. Conf. S. Rosa de Viterbo Elucidarii tom. 1. pag. 102. et Appendic. pag. 7. A verbo latino *Alimonia* originem trahere hanc vocem scribit et exemplis multis et documentis Lusitan. adductis firmare conatur, ea non tantum *hortum*, sed quodvis prædium haud longe ab urbe situm, significari.]

Fig. 1. *Almunia*: Glossarium Mediae et Infimæ Latinitatis, conditum a Carolo du Fresne, domino Du Cange; nouvelle éd. par L. Favre, t. I-X, Niort, 1883-1887

The screenshot shows a search interface for the word 'almunia'. At the top, there are options for 'consulter un article', 'recherche plein texte', and 'formes exactes'. Below these is a search bar containing the word 'almunia' and a 'Rechercher' button. Underneath the search bar are radio buttons for 'citations latines', 'citations françaises', and 'citations grecques'. The main content area displays the entry for 'ALMUNIA' with its definition and various references. At the bottom, there is a footer with the text '© École des chartes.' and 'ALMUNIA'.

ALMUNIA, Prædii rustici species apud Hispanos. Charta Sanctii Regis Aragonum æræ 1132. apud Martinezum in Hist. Pinnatensi lib. 3. cap. 9 :
 Concedo prædicto Cænobio, ... illam meam Almuniam, vocatam Daymus, quæ afrontat ex una parte cum Torredellas, etc.
 Alia Alfonsi VI. Regis æræ 1133. apud Anton. de Yezep in Chronico Ord. S. Benedicti tom. 6 :
 In quo loco incipit alia via, per quam descendunt usque in viam publicam super Almuniam Regis, etc.
 Observantiæ Regni Aragon. lib. 5. tit. de Jure dotium, § 4 :
 Tamen si Miles vel Infantio habent unam Almuniam vel turrim, censetur una hæreditas cum toto hæreditamento adjuncto illi Almunia, vel turri.
 Adde Colmenarezium in Hist. Segobiensi cap. 16 § 4.

In his tribus exemplis, excepto forte postremo, per *Almunia* intelligi potest hortus, ut intelligendus est dubio procul in sequentibus ; Testamentum I. Adefhonsi Regis Hispania apud Marten. tom. 1. Collect. Ampliss. col. 546. C :
 Et ut hi qui in eadem Ecclesia superscriptæ permanserint, supplementum aliquod victui habere possint, offero illis villam unam nomine Huleka, et unam Almuniam, quam nos Latine vocamus Ortum, qui est prope illam Ecclesiam S. Servandi.
 Et in Archivo S. Victoris Massiliensis armor. Hispan. n. 115 :
 Almunia Regis, quam nos Latine vocamus Ortum.
 Lusitanis *Almuinha*. Conf. S. Rosa de Viterbo Elucidarii tom. 1. pag. 102. et Appendic. pag. 7. A verbo latino *Alimonia* originem trahere hanc vocem scribit et exempla multis et documentis Lusitan. adductis firmare conatur, ea non tantum *hortum*, sed quodvis prædium haud longe ab urbe situm, significari.

Fig. 2. *Almunia*: Du Cange et al., *Glossarium Mediae et Infimæ Latinitatis*, édition numérique, en ligne : <http://ducange.enc.sorbonne.fr/almunia>

Almunia

M16

- Et dimitto filiis meis iterum Arnaldo
 Et Poncio atque Guillelmo almuniam
 quam adquisiui ab Ermengaudo comite
 in termino balagarii in qua pater meus
 dudum obiit. Et almuniam quam adqui-
 siui a guillermo quitardi de chonches
 prope meranges et feuum de rapita
 quem teneo per manum Raimundi quitardi
 de mediano - test. Petri Poncii - 6 id. mayo 8 Lud. 8

Cartulario de la Seo de Urgel.
 Tomo 1 Núm. 65 Folio 35 col. 2 Año 1116

Almunia.

... concessit eidem ecclesie ipsam
 suam almuniam de Herda et
 ortum unum et duas uineas...
 et animalia que laborabant
 prefatam almuniam

R. Ber. 2V n^o 337-
 2 Kl. junio - 23 Lud. junior.
 a. 1160

Fig. 3. et 4. Almunia, fiches de Josep Balari i Jovany.
 Barcelona, Institut Milá y Fontanals, CSIC, bibliothèque de l'équipe du GMLC

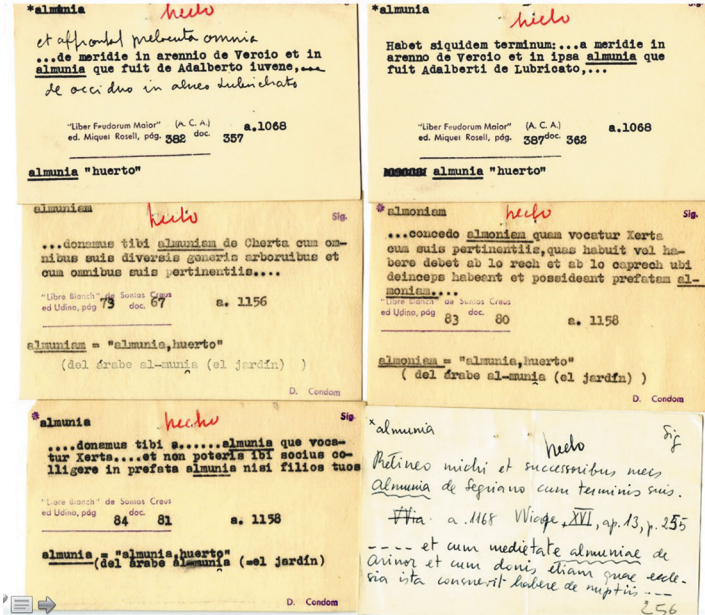


Fig. 5. *Almonia*, fiches du GMLC. Barcelone, Institution Milá y Fontanals, CSIC, bibliothèque de l'équipe du GMLC

almonia, mon- [ár. munya, ár. hisp. 'huerto, jardín vasto'] huerta o granja: 1068 LFeud. I 362, p. 387: habet siquidem terminum... a meridie in arrenio de Vercio et in ipsa almonia que fuit Adalberti de Lubricato. 1068 LFeud. I 357, p. 382: et affrontat preloca omnia... de meridie in arrenio de Vercio et in almonia que fuit de Adalberto iuvene, de occiduo in aluce Lubricato. 1079 LFeud. I 165, p. 174: et transeunt (sc. terminum) ab occidente per transversum per almoniam grossam usque ad terminum comitatus Vrgelli. 1091 CTabernoles 95, f. 50^r: I turrem que uocant Belcarre... cum duabus almonis qui in circuitu eius sunt. Affrontat... de meridie in ipsa almonia de Alfiz et de Filela. 1092 CTabernoles 66, f. 36^r (VViage XII, ap. 1, p. 211-212): insuper addo ipsam almoniam que fuit de Iuceph Caualer, quae est ultra collo de Portella in ipso plano... et sic uadit per summitatem de ipsa serra, et ipsas almonias parulas includit, quae sunt de ipsa almonia iam dicta. 1094 (Urgell) Miret, *Castelló*, ap. VIII, p. 362: Et addo ad hoc donum omnes meschitas que sunt infra muros ciuitatis Balagarii cum omnibus terris et uineis et hortis et arboribus et tendis et almoniis et omnia que illorum sunt uel esse debent. 1116 ACUrgell, *Cart. I* 65, f. 35, col. 2: dimitto filiis meis... almoniam quam adquisiui ab Ermengaudo comite in terminio Balagarii in qua pater meus dudum obiit. Et almoniam quam adquisiui a Guilermo Guitardi de Chonches prope Meranges.

1129 ACUrgell, *Cart. I* 80, f. 40, col. 1: et in castrum de Castellione unam almoniam qui uocant de Auimafada. 1156 CSCreus 67, p. 73: donamus tibi... almoniam de Cherta cum omnibus suis diuersi generis arboribus... et cum omnibus suis ad se pertinentibus aquis. 1158 CSCreus 80, p. 83: concedo almonia quam uocatur Xerta... ab lo rech et ab lo caprech. 1158 CSCreus 81, p. 84: donamus tibi... almonia que uocatur Xerta... ut deinceps habeas et possideas prefatam almoniam... et non poteris ibi socius colligere in prefata almonia nisi filios tuos. 1160 ACA Ramón Berenguer IV, n.º 337: concessit eidem ecclesie ipsam suam almoniam de Ierda et ortum unum... et animalia que laborant prefatam almoniam. 1168 (Lérida) VViage XVI, ap. 13, p. 255: retiño michi... almonia de Segriano. p. 266: cum medietate almoniae de Arinor.

Fig. 6. *Almonia*, première édition du GMLC, fascicule 1, a - aragalius (1960)

almunia, -mon-

[*ar. munya, ar. hisp. 'hort, jardí vast' | 'huerto, jardín vasto' | 'vegetable garden, big garden'*]

horta o granja | huerta o granja | vegetable garden or farm:

- 1068** LFeud. I 362, p. 387: habet siquidem terminum ... a meridie in arenno de Vercio et in ipsa almunia que fuit Adalberti de Lubricato.
- 1068** LFeud. I 357, p. 382: et affrontat prelocuta omnia ... de meridie in arenno de Vercio et in almunia que fuit de Adalberto iuene, de occiduo in alueo Lubricato.
- 1079** LFeud. I 165, p. 174: et transeunt (sc. termini) ab occidente per transuersum per almuniam grossam usque ad terminum comitatus Vrgelli.
- 1091** CTavèrmoles 95, f. 50^r (Soler 47): I turrem que uocant Belcaire ... cum duabus almunis qui in circuito eius sunt. Affrontat ... de meridie in ipsa almunia de Alfiz et de Filela.
- 1092** CTavèrmoles 66, f. 36^r, Soler 49 (VViage XII, ap. 1, pp. 211-212): insuper addo ipsam almuniam quae fuit de Iuceph Caualer, quae est ultra collo de Portella in ipso plano ... et sic uadit per sumitatem de ipsa serra, et ipsas almunias paruulas includit, quae sunt de ipsa almunia iam dicta.
- 1094** (Urgell) Miret, *Castellbó* 8, p. 362: et addo ad hoc donum omnes meschitas que sunt infra muros ciuitatis Balagarii cum omnibus terris et uineis et hortis et arboribus et tendis et almunis et omnia que illorum sunt uel esse debent.
- 1116** ACUrgell, *Cart.* I 65, f. 35, col. 2: dimitto filiis meis ... almuniam quam adquisiui ab Ermengauda comite in terminio Balagarii in qua pater meus dudum oblit. Et almuniam quam adquisiui a Guillermo Guitardi de Chonches prope Meranges.
- 1129** ACUrgell, *Cart.* I 80, f. 40, col. 1: et in castrum de Castellione unam almuniam qui uocant de Auimfadida.
- 1156** CSCreus 67, p. 73: donamus tibi ... almuniam de Cherta cum omnibus suis diuersi generis arboribus ... et cum omnibus suis ad se pertinentibus aquis.
- 1158** CSCreus 80, p. 83: concedo almonia quam uocatur Xerta ... ab lo rech et ab lo caprech.
- 1158** CSCreus 81, p. 84: donamus tibi ... almunia que uocatur Xerta ... ut deinceps habeas et possideas prefatam almuniam et non poteris ibi socius colligere in prefata almunia nisi filios tuos.
- 1160** ACA Ramon Berenguer IV, n. 337: concessit eidem ecclesie ipsam suam almuniam de Ilerda et ortum unum ... et animalia que laborabant prefatam almuniam.
- 1168** (Leida) VViage XVI, ap. 13, p. 255: retineo michi ... almunia de Segriano. *ibid.*, p. 256: : cum medietate almuniae de Arinor.

Fig. 7. *Almunia*, édition numérique du *GMLC* (prototype)

amarello [*cf. esp. amarillo del bajo lat. hisp. amarellus 'amarillento, pálido', dim. de amarurus*]¹ *amarillo*: **1069** ACUrgell, *Cart.* I 407, f. 134, col. 1 (Miret, *BRABL* VI, p. 385): ad ipsa canonica ... dimisit ... unum barril de X quinals simul cum ipso superlito de palio que habebat in ipsa sede et I feltro amarello qui ibidem erat.

¹ *Amarellus no ha dejado rastro en el dominio del catalán.*

Fig. 8. *Amarello*, première édition du *GMLC*, fascicule 1, a - *aragalius* (1960)

amarellus, -a [cf. esp. amarillo < lat. med. hisp. amarellus, 'esgrogueit, pàl·lid' | 'amarillento, pàlido' | 'yellowish, pale', dim. ab amarus]¹ *groc* | amarillo | yellow: **924** DipOsona 283, p. 263 (*donació dels comtes Sunyer i Riquilda a l'altar de S. Salvador del monestir de Ripoll* | *donación de los condes Suñer y Riquilda al altar de S. Salvador del monasterio de Ripoll* | *donation made by the Counts Suñer and Riquilda at the altar of St. Salvador in the monastery of Ripoll*): planeta I colore amarella diocodrina, kappa I uermilia ex diorodono. **998** DipGirona 612, p. 518: ad domum Sancti Petro Gallicantu uacas III et uitulos III et somero I et caldaria I et oues XXX, mapes I, tuallia I, tapitto I et feltro I amarello et capizalo I. **1069** ACUrgell, *Cart.* I 407, f. 134, col. I (ed. Miret, *Aplech*, p. 385): ad ipsa canonica ... dimisit ... unum barril de X quinals simul cum ipso superlito de palio que habebat in ipsa sede et I feltro amarello qui ibidem erat. **1068-1071** Sanahuja, *Àger* 27, p. 348: uestimento ... optimo de oztorino amarel.

¹ Amarellus no ha deixat rastre en el domini del català.

¹ Amarellus no ha deixado rastro en el dominio del catalán.

¹ Amarellus has left no trace in the Catalan area.

Fig. 9. *Amarellus*, deuxième édition du *GMLC* (2010)

amarellus, -a

[cf. esp. amarillo < lat. med. hisp. amarellus, 'esgrogueit, pàl·lid' | 'amarillento, pàlido' | 'yellowish, pale', dim. ab amarus]¹

groc | amarillo | yellow:

924 DipOsona 283, p. 263 (*donació dels comtes Sunyer i Riquilda a l'altar de S. Salvador del monestir de Ripoll* | *donación de los condes Suñer y Riquilda al altar de S. Salvador del monasterio de Ripoll* | *donation made by the Counts Suñer and Riquilda at the altar of St. Salvador in the monastery of Ripoll*): planeta I colore amarella diocodrina, kappa I uermilia ex diorodono.

998 DipGirona 612, p. 518: ad domum Sancti Petro Gallicantu uacas III et uitulos III et somero I et caldaria I et oues XXX, mapes I, tuallia I, tapitto I et feltro I amarello et capizalo I.

1069 ACUrgell, *Cart.* I 407, f. 134, col. I (ed. Miret, *Aplech*, p. 385): ad ipsa canonica ... dimisit ... unum barril de X quinals simul cum ipso superlito de palio que habebat in ipsa sede et I feltro amarello qui ibidem erat.

1068-1071 Sanahuja, *Àger* 27, p. 348: uestimento ... optimo de oztorino amarel.

¹Amarellus no ha deixat rastre en el domini del català.

¹Amarellus no ha deixado rastro en el dominio del catalán.

¹Amarellus has left no trace in the Catalan area.

Fig. 10. *Amarellus*, édition numérique du *GMLC* (prototype)

Résumés / Abstracts

Sylvie BAZIN-TACCHELLA 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 CréaLScience 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 Maigne)
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