

REVUE DE
LINGUISTIQUE
FRANÇAISE
DIACHRONIQUE

7
2017

DIACHRONIQUES

LES ÉTATS ANCIENS
DES LANGUES À L'HEURE
DU NUMÉRIQUE

Petit – 979-10-231-2165-0



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 & Szeceł – 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

Terminographie diachronique : le cas de la terminologie médiévale française

Gérard Petit

EA 4509, Sens Texte Informatique Histoire

Université de Paris Nanterre

Problématiques

Une approche diachronique du lexique se heurte à au moins deux problèmes cruciaux :

- celui de la définition de la diachronie en question (T-1). L'intervalle couvert conditionnera la complexité des résultats obtenus ;
- celui du réglage théorique de l'approche en elle-même. À partir de quel modèle de représentation (morphologique et sémantique notamment) l'investigation est-elle menée ? Celui qui est contemporain de la recherche elle-même ou bien un autre, qui serait spécifique à la diachronie envisagée ?

Le projet CréaLSscience (ANR-10-CREA-0007) a été l'occasion de remettre à plat un certain nombre d'acquis théoriques et méthodologiques, que leur évidence impose au chercheur, mais dont l'influence sur la nature des résultats ne doit pas être sous-estimée. D'abord en situant son objet, la langue médiévale, sur une synchronie isolée, passée (T-1), mais qui couvre à elle seule une véritable diachronie. À cet effet, il privilégie la cartographie en T-1 plutôt que l'évolution vers T₀, laquelle se déduira de la distance entre les deux configurations. Ensuite, en tentant de restituer autant que faire se peut l'état des représentations conceptuelles en T-1, telles qu'on peut supposer qu'elles étaient configurées et exprimées alors. De ce fait il renonce à les exposer à travers le filtre des connaissances

en To. Ainsi, il se démarque fondamentalement des perspectives diachroniques adoptées usuellement dans les recherches linguistiques. Mais il intègre en même temps l'incertitude et l'insécurité comme données heuristiques primordiales. Enfin, le projet CréaLScience vise la production d'un dictionnaire électronique représentant les données de l'investigation. Ce dictionnaire, le *Dictionnaire du français scientifique médiéval (DFSM)* est l'enjeu d'un développement d'outils (théoriques, méthodologiques, informatiques) dédiés à une investigation terminologique diachronique.

Le *DFSM* est un ouvrage inédit à plusieurs titres :

- il se dote d'un objet explicitement terminologique portant sur une synchronie non explorée jusqu'ici par la terminologie générale ;
- il se démarque des principaux dictionnaires consacrés à la langue du Moyen Âge sur plusieurs points : (i) il ne constitue pas un ouvrage de traduction intralinguale du français médiéval vers le français moderne ; (ii) il ne vise pas un objectif anecdotique, qui reposerait sur des dépouillements partiels et non explicites et qui chercherait à présenter une forme d'exotisme culturel, caractéristique d'un passé¹ ; (iii) il ne cherche pas à décrire des emplois non reconnus au sein de discours spécialisés² ; (iv) il vise une représentation des concepts aussi proche que possible de leur état dans la conscience supposée des locuteurs de T-1 ;
- ses inventaires reposent sur le dépouillement systématique de textes spécialisés³.

1. Sont concernés ici des ouvrages destinés au grand public qui, en tant que tels, ont leur validité, mais qui n'entrent pas dans un projet scientifique. Citons entre autres : <http://www.medieval-moyen-age.net/categorie-125346.html> ; <http://www.castlemaniac.com/lexique-medieval/lexique-medieval.php> ; <http://www.provins-banquet-des-troubadours.fr/vocabulaire-medieval>.

2. Pour cette raison, ses inventaires se distinguent pour partie de ceux du *Dictionnaire du moyen français* (en ligne : <http://www.atilf.fr/dmf>).

3. Voir la bibliographie générale du projet.

Compte tenu de ces paramètres, deux orientations lexicographiques majeures ont été adoptées :

- **Dictionnaire « de langue » vs. dictionnaire de corpus.** Opérant sur un état de langue pour lequel aucune compétence vive, nourrie de l'interlocution, n'est disponible, le dictionnaire ne peut puiser ses informations que dans des corpus écrits, et reste donc tributaire de ceux-ci. Deux attitudes sont possibles : (i) soit il cherche à se dégager des corpus pour viser une représentation abstraite, et donc se conformer à son objectif initial (approcher les concepts dans la culture de T-1). Cet objectif n'est tenable qu'en opérant une moyenne sur les informations fournies par les corpus ; (ii) soit il adhère aux corpus, et s'achemine alors vers une forme de philologie. Aucune des deux directions ne peut être tenue pleinement, et ce pour des raisons différentes : l'objectif philologique ne présente aucun intérêt pour une démarche terminologique générale ; l'objectif linguistique, fondé sur des moyennes, ne peut être mené à son terme en toutes circonstances, compte tenu de l'état de langue investigué ;
- **Dictionnaire « de langue » vs. dictionnaire encyclopédique.** Le projet CréaLScience s'inscrit dans la tradition des dictionnaires encyclopédiques, puisque son objet est le vocabulaire technique et scientifique médiéval. Toutefois, visant la description des concepts en T-1, il s'éloigne de l'objectif de principe d'un dictionnaire encyclopédique traditionnel : son but n'est pas la connaissance du monde en lui-même (le réel), mais du monde pensé, représenté à un moment donné de l'histoire (les conceptologies). Par ailleurs, les emplois syntaxiques particuliers (notamment pour les verbes, le passif, l'emploi pronominal, le participe passé⁴) sont recensés et font l'objet d'un traitement dans des divisions spécifiques. Le dictionnaire envisagé est un ouvrage foncièrement terminologique, et s'oriente donc vers l'encyclopédie, tout en reposant sur un socle dont

4. Par ex. *aiguiser*, *apostumer*.

les fondements théoriques et épistémologiques sont explicitement linguistiques.

Cette étude consacrera un premier développement aux problèmes posés par l'établissement de la nomenclature, compte tenu notamment de l'extrême variabilité rencontrée s'agissant de la graphie du lexique dans la langue médiévale. Dans un second temps, nous aborderons la microstructure en concentrant notre attention sur les contraintes qui obèrent la représentation sémantique, l'objectif du *DFSM* étant, rappelons-le, d'approcher au plus près la configuration du concept, tel qu'il pourrait être formulé dans la synchronie visée.

La métalangue de description est le code dans lequel sont formalisés l'ensemble des descripteurs, mais aussi les règles qui président à la description elle-même. Au titre des descripteurs, on retiendra :

- la forme graphique de l'entrée ;
- l'indication de catégorie grammaticale ;
- la définition ;
- l'exemple et sa référence bibliographique ;
- les notes diverses, linguistiques et encyclopédiques, qui accompagnent de droit la définition.

L'ensemble de ces descripteurs est usuellement réparti entre la macrostructure (les entrées d'un dictionnaire) et la microstructure (les constituants de l'article de dictionnaire).

La macrostructure : gérer la variation, frein de la lexicalité

La macrostructure est la liste des entrées d'un dictionnaire. Elle constitue le premier matériau réuni en vue de son élaboration. Fonctionnellement, elle renseigne sur les choix et prélèvements effectués au sein des corpus, le degré de profondeur ou d'étendue de l'investigation. L'établissement de la nomenclature est également l'enjeu d'un réglage morphosyntaxique⁵, capital quand il s'agit de la langue médiévale, concernant :

5. Il touche par ex. le classement alphabétique des caractères accentués quand ils ont valeur discriminante (*tache/tâche ; coupe/coupé*).

- (1) le traitement à appliquer à la variation graphique et à la difficulté, parfois très importante, à dégager un lemme ;
- (2) le mode de classement des séquences polylexicales (noms composés, locutions adjectivales ou verbales) ;
- (3) le traitement à apporter aux emplois morphosyntaxiquement contraints (passif, participe passé, pluriel, emplois substantivés, adjectivés) au sein du dictionnaire : disposent-ils d'une entrée ou bien sont-ils traités comme subdivisions d'un article ? Sur la base de quels critères la décision est-elle prise ?

Le traitement de la variation graphique et/ou phonologique

La langue médiévale se caractérise par une forte variation (notamment graphique et morphologique), laquelle hypothèque parfois l'identification d'un lemme commun à plusieurs occurrences et susceptible de synthétiser l'unité psycholinguistique du signe. Une investigation ciblée portant sur les lettres A et C révèle les configurations suivantes :

Existence d'un terme en ancien français (AF), mais pas en français moderne (FM)

Sont concernés les termes renvoyant soit à un concept sans équivalent en FM (par ex. : *caladre*), soit à un concept disponible en FM mais associé à un autre signifiant (par ex. : *amoustir*, *apostume*, *eaueux*, *aigos*, *aidement*). La variation affecte la structure morphologique de l'unité.

Tableau 1. Variations terminologiques entre FM et AF

Français moderne	Ancien français
<i>humidifier</i>	<i>amoustir</i>
<i>abcès, tumeur</i>	<i>apostume</i>
« oiseau blanc qui a le pouvoir de prédire la mort... »	<i>caladre</i>
<i>aqueux</i>	<i>eaueux, aigos</i>
<i>aide</i>	<i>aidement</i>

Existence d'un terme en AF avec des graphies multiples, dont éventuellement celle attestée en FM (nous soulignons)

Ces variantes peuvent résulter de différences phonologiques affectant plus ou moins largement la structure du signifiant.

Tableau 2. Variations graphiques en AF

Français moderne	Ancien français
<i>courage</i>	<i>corage, coraige, courage</i>
<i>concombre</i>	<i>cucumère, concourde, concombre, coucombre, cucumerus</i>
<i>aloès</i>	<i>alee, aloe, aloee, aloem, aloen, aloü, aloüs</i>

Existence d'un terme en AF avec des graphies multiples, mais non celle attestée en FM

La variation peut affecter trois plans différents de l'unité : graphique (*cyprès/cypré*; *cyclamen/ciclamen*), phonologique (*cyprès/cipre*), morphologique (*camomille/chermière*), de manière indépendante ou cumulée⁶.

Tableau 3. Graphies spécifiques à l'AF

Français moderne	Ancien français
<i>croûte</i>	<i>crouste, croste</i>
<i>cyclamen</i>	<i>ciclamen, ciclam, malum terre, panis porcinus, pain à porc, pain de pourceau</i>
<i>camomille</i>	<i>chermière, camomil</i>
<i>cyprès</i>	<i>cyprus, cypré, cipré, cypres, cippre, cyperi(s)</i>
<i>châtaignier</i>	<i>chastignier, chasteignier, chastaignier</i>
<i>aiguiser</i>	<i>aguiser</i>
<i>haleine</i>	<i>alaine, alainne</i>
<i>ammoniac</i>	<i>amoniac, armoniac, amoniat</i>

La difficulté consiste donc à déterminer la forme de l'entrée. Dans la configuration (1), il n'existe pas d'autre alternative que d'entrer la forme médiévale en vedette, en sélectionnant celle qui sera la moins marquée dialectalement en ancien français, ou la mieux attestée en cas de variantes graphiques. Pour (2) et (3), le choix pourrait se porter sur le FM ou l'AF. Toutefois, partant du principe que le *DFSM* s'adresse à un public médiéviste, mais

6. Les dimensions de cette étude ne nous permettent pas d'entrer ici dans le détail.

aussi à des lecteurs venant de tous horizons, le choix de la forme en FM sera privilégié, ce pour plusieurs raisons :

- le FM peut être valablement tenu comme la langue de consultation des lecteurs non médiévistes ;
- le FM est la langue dans laquelle s’effectue la description dans la microstructure ;
- l’ensemble des variantes recensées dans le *DFSM* ne sont pas connues de tous les médiévistes.

Les articles consacrés à des termes présentant des variantes sont donc traités de manière hiérarchique :

- le lemme de l’entrée est exprimé en FM et institue ainsi un article tutélaire : *croûte*, *cyclamen*, *camomille*, *cyprès*, *châtaigner*, *aiguiser*, *ammoniac*, *haleine* ;
- un renvoi par lien hypertexte à la (ou aux) variante(s) est ajouté sous l’entrée ;
- la (ou les) variante(s) présente(nt) un article vide, constitué uniquement d’un renvoi à l’article tutélaire : *crouste*, *croste* vers *croûte* ; *amoniac*, *armoniac*, *amoniac* vers *ammoniac*, etc.
- lorsque la variation affecte la structure morphologique de l’unité, les différentes formes sont traitées indépendamment, chacune comme vedette d’un article plein avec définition et renvoi synonymique réciproque : *cyclamen* vs. *pain à porc*, *pain de pourceau*.

Variante et nomen

La variation graphique, morphologique et la synonymie affectent des unités exprimées en AF, mais aussi des formants latins apparaissant dans des énoncés produits en AF. Certains de ces formants correspondent à des unités disposant par ailleurs d’une dénomination en AF :

- lat. *malum terre*⁷, *panis porcinus* / fr. *ciclamen* (FM : *cyclamen*) ;
- lat. *cacochia* / fr. *cacochie*⁸ ;

7. Littéralement, « pomme de terre » (sic!).

8. « Dégradation de l’état de santé due à un déséquilibre des humeurs* ».

- lat. *calamus aromaticus* / fr. *calame aromatique*⁹ ;
- lat. *calcatrix* / fr. *hydre* ;
- lat. *capparis* / fr. *capre* (FM: *câpre*).

Ces formants latins sont la trace d'une dénomination antérieure, laquelle perdure alors qu'une concurrence avec le français s'est installée. La relation du latin au français médiéval peut être de l'ordre de l'adaptation morphosyntaxique (*cacochia/cacochie*, *calamus aromaticus/calame aromatique*) ou de la traduction (éventuellement à partir d'un autre modèle): *panis porcinus/pain à porc*; *calcatrix/hydre*¹⁰. Ces formants fonctionnent dans le discours scientifique de manière analogue aux noms (étiquettes latines que les classifications ultérieures utiliseront, aux xvii^e et xviii^e siècles, pour systématiser les taxinomies).

Pour tout binôme associant un terme français à un nomen, que des variantes graphiques existent ou non pour l'un et/ou l'autre, les deux unités disposent chacune d'un article plein, avec définition(s) et exemple(s), du fait qu'elles sont morphologiquement distinctes. Certaines variantes présentent un caractère hybride, car elles mêlent le latin et le français. Tel est le cas de *bol armenicum*, variante de *bol armenic* (« Argile que l'on fait venir d'Arménie, et qui se caractérise par sa couleur rouge et sa grande viscosité »). Du fait de leur divergence morphologique, les deux formes sont assimilées à des synonymes car elles ne constituent ni des noms à part entière, ni de simples variantes graphiques. Leur existence met néanmoins en évidence un phénomène dont la terminologie médiévale se fait le témoin : la distinction entre statuts sémiotiques (dénomination, variante, nomen, synonyme) n'est pas étanche, mais se négocie sur la base d'une continuité, d'où l'existence d'unités qui n'appartiennent pleinement à aucune catégorie faute d'en présenter toutes les caractéristiques.

9. « Petit arbre* qui pousse en Inde, dans des sols humides et qui a une tige creuse, ses feuilles ressemblant à celles de l'iris, roseau odorant ».

10. La place nous manque pour traiter le cas de *malum terre*, littéralement « pomme de terre ».

D'autres formants latins ne disposent pas d'équivalent français: par ex. *carpobalsamum*. Leur effectif est moins nombreux que le précédent, les auteurs médiévaux cherchant à franciser les dénominations afin d'homogénéiser l'expression de leur discours. Ces termes latins « orphelins » sont traités comme des dénominations à part entière: ils figurent en entrée d'article, sont définis et illustrés d'exemples. Ces unités présentent toutefois un régime qui rend délicate l'assignation d'un statut sémiotique fixe: dénomination, variante ou nomen. Ainsi, pour poursuivre avec le terme *carpobalsamum*:

Fruit d'un arbrisseau, le baumier, charnu et à un seul noyau blanc osseux, d'un gris rougeâtre, d'une saveur légèrement amère et aromatique.

Cette unité fonctionne dans les discours comme nomen associé à *baume*. Or, sa définition précise qu'il identifie un référent spécifique à une sous-catégorie de celui-ci. *Carpobalsamum* fonctionne donc comme une dénomination à part entière et se positionne comme hyponyme de *baume*. Ce type de difficulté est révélateur de l'ambivalence sémiotique qui affecte certains termes de la langue médiévale.

Entrée et sous-entrée

Une tradition lexicographique empiriquement instituée depuis le ^{xvii}e siècle affecte l'entrée aux dénominations correspondant au gabarit du mot graphique¹¹. Les unités polylexicales et autres locutions sont traitées comme des sous-entrées: elles figurent à l'intérieur de l'article, dans une division spécifique. Ces sous-entrées disposent de leur propre lemme, sont définies et illustrées d'exemples. Bref, elles reçoivent (presque) le même traitement que les entrées à part entière.

Le *DFS* comporte des sous-entrées, qui reçoivent le même traitement que celui appliqué dans les autres dictionnaires. Ainsi, sous l'entrée *ail*, figure le nom composé *ail sauvage*:

11. Le mot graphique peut se définir comme une chaîne de caractères séparée par des espaces. Le trait d'union vient abolir l'espace et réduire au format du mot graphique des suites polylexicales (par ex. *château-fort*). Les mots à trait d'union figurent traditionnellement en entrée d'article dans les dictionnaires.

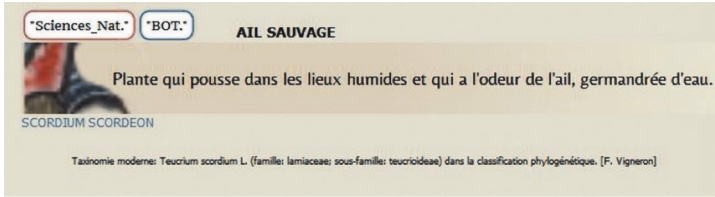


Fig. 1. Ail sauvage (DFSM - capture d'écran)

qui dispose de ses propres noms, différents de celui d'*ail* (*allium*).

Le champ réservé traditionnellement à l'expression de la sous-entrée est également affecté aux emplois syntaxiques particuliers de l'entrée qui justifient une définition à part entière, associée à une relation référentielle spécifique. Ainsi le verbe *consolider* connaît-il trois emplois dans le domaine médical :

- transitif : « Assurer la cicatrisation d'une plaie, la consolidation d'une fracture » ;
- intransitif : « Assurer sa propre cicatrisation en parlant d'une plaie, sa propre réparation en parlant d'une fracture » ;
- pronominal : « Se fermer, *en parlant d'une plaie ou d'un ulcère* », ce dernier disposant d'un périmètre sémantique différent (nous soulignons).

Pour prendre un autre exemple, le verbe *aggraver* possède en AF une acception médicale spécifique, au passif : « Être attaqué, mis en danger, en parlant du corps et en particulier des vertus* qui l'animent », qui est renseignée en sous-entrée.

Fonctionnellement, le champ de la sous-entrée accueille du matériau qui marque un écart relativement au régime prédictible ou lexicalisé de l'entrée, soit au plan morphologique, soit au plan syntagmatique. Il constitue le site d'accueil des formes ne présentant pas la sémiotique intègre de l'entrée, mais ne justifiant pas pour autant leur éviction du champ de l'unité psycholinguistique du signe.

Les signes diacritiques

La nomenclature est la liste des entrées d'un dictionnaire. C'est aussi un ensemble de règles qui président à son organisation. À ce titre se pose régulièrement la question du traitement des lettres portant des signes diacritiques (accents, tréma, cédille) et notamment de l'ordre de leur classement relativement aux caractères « simples » dont ils s'écartent. Les dictionnaires choisissent usuellement de les classer soit avant, soit après ceux-ci dans l'ordre alphabétique de la nomenclature.

La graphie des entrées privilégiant le FM pour favoriser l'ergonomie de la consultation, la distribution et le classement des caractères accentués ne répond pas à une propriété de la langue médiévale, qui s'écrit sans accent : l'accent aigu attesté sur le « e » en finale de mot est une commodité moderne permettant de distinguer des participes de formes homonymes ou d'éviter des interprétations fautives. Si le « ç » est attesté, il dispose d'un régime différent de celui qui est le sien en FM :

- (1) il peut se rencontrer devant une voyelle autre que « e » et « i » à l'intérieur d'un mot, comme en FM : *commençait* (FM « *commençait* ») ;
- (2) toujours devant une de ces voyelles, il peut apparaître comme variante de « s » ou de « ss », par ailleurs attestés en AF (*ç sucre*, pour *sucre* ; *cuiçon*, pour *cuisson*) ;
- (3) il peut doubler un « s » (*sçavoir*, pour *savoir*) ;
- (4) comme on le voit en (2), le « ç » peut figurer à l'initiale d'un mot, position qui n'est plus la sienne en FM. La nomenclature du *DFSM* associe les deux logiques diacritiques, celles du FM et de l'AF. Ainsi, pour reprendre un exemple cité plus haut, le dictionnaire comprendra deux entrées : *sucre* et *ç sucre*, la seconde n'étant qu'une variante de la première, ou encore une entrée *çoire* (synonyme de *pois chiche*). Les lemmes comportant des caractères accentués relèvent donc uniquement d'une intervention en FM. Ajoutons que, dans le but d'éviter une confusion dans la prononciation, certains lemmes de termes spécifiquement médiévaux ont

été accentués (*atempèrement* à la nomenclature devient *atemprement* dans les citations).

Dans le déroulé alphabétique de la nomenclature du *DFSM* les caractères portant des signes diacritiques n'entrent pas en concurrence avec leurs équivalents nus. Pour cette raison, ils sont traités à l'identique de ceux-ci (nous soulignons) :

cedewale > **cèdre** > *ciguë* > *ceindre*
cane > *canel* > *canele* > **çanesson** > *canette*
capel > **céphalée** > *cephalica*
codling > **cœ** > **coégalité** > **coéquation** > *coer* > *cœur* > *coffin*
coingnier > **çoire** > *cois*

Actuellement, compte tenu de l'état de la nomenclature, aucune concurrence n'apparaît entre caractères accentués et non accentués justifiant une décision relative à l'ordre d'apparition des lemmes dans la succession alphabétique. L'orientation terminologique, donc domaniaire du dictionnaire n'est probablement pas étrangère au fait que la question de la précellence d'une lettre sur une autre ne se pose pas dans la synchronie considérée.

La microstructure

La microstructure d'un dictionnaire comprend l'ensemble des informations ayant en charge la description de l'entrée. Nous examinerons ici l'une de ses composantes principales : la définition.

La définition : la diachronie contre l'anachronisme

La définition est à la fois une opération sémantique et la séquence linguistique qui la concrétise. En terminographie, elle possède également une fonction signalétique, car elle justifie le rattachement d'un terme à un domaine.

Nous ne reviendrons pas sur la difficulté de principe liée à l'activité de définition, *a fortiori* lorsqu'elle est terminologique, et qui repose sur le postulat (indémontrable!) que le sens est une entité close, objectivable dans une périphrase, que l'on peut segmenter en éléments simples sur le même modèle que

les sons d'une langue¹². La question devient plus aiguë dès que l'on aborde le lexique en diachronie et que l'on cherche à représenter le concept dans sa configuration T-1. La définition lexicale ou terminographique n'est plus alors qu'une hypothèse, dont les contours sont au moins partiellement informés par les représentations de son producteur en To¹³.

Approcher le concept au plus près de son état dans la synchronie T-1 implique de reproduire l'état de la culture qui pensait le réel, tel qu'elle le pensait. C'est également projeter une perspective sur la terminologie et admettre que les concepts ne reflètent pas un réel en soi (tel que le postuleraient une onomasiologie ou une ontologie modernes), intemporel et insensible aux variations de l'espace, mais sont des représentations ancrées dans la culture de leur temps, de leur espace géographique et surtout dans les mots qui les véhiculent. Pour cette raison, la terminologie est abordée sous un angle sémasiologique. Elle repose sur le principe que la définition d'un concept doit passer par des termes en accord avec la conscience de l'époque et du lieu.

Le risque rencontré par la terminographie est alors celui de l'anachronisme. Celui-ci peut avoir deux sources : la configuration du concept lui-même ou bien le choix des définissants. Le *DFSM* tient compte de ces deux ordres de contraintes, bien qu'ils répondent à des logiques différentes. La structuration du concept, envisagé comme chaînage de traits structuré sur un modèle logique par incluants et différences spécifiques, peut être attestée dans le corpus par des énoncés définitoires (nous soulignons) :

Boterel **est un ver envenimé qui** habite en terre et en lieu
moiste si come dit Plinius ou XXIIe chapitre de son XVIIIe livre.
(Jean Corbechon, *Livre des propriétés des choses*, XVIII, 15,
fol. 299 rb.)

12. Nous préciserons seulement que la définition, qu'elle soit lexicographique ou non, reflète davantage la construction du modèle de représentation sémantique dont elle est l'application que les données du réel extralinguistique.

13. Or, approcher le concept en T-1 impose que l'on fasse abstraction de ses propres représentations sémantiques en To.

Apotaine¹⁴ est un poisson qui est apellez chevfl[uvi]el, por ce que il naist ou fluve dou Nil; et son dos et ses crins et sa voiz est come decevals. Ses ongles sont fendues come de buef et danz come sengler et la coe retorte, [et mangue] ble[s] [de champ] ou il vet a recolons por les aguaiz des homes. (Brunet Latin, *Trésor*, I, 135, p. 240)

Les énoncés dénominatifs constituent le second type de matériau investigué en priorité. Ils présentent une forme alternative des énoncés définitoires. Ils permettent entre autre de situer les termes au sein de champs notionnels, ou bien les variantes d'un même terme entre elles (nous soulignons) :

La voie de respirer on l'**appelle** trachee artere et le chief on l'**appelle** epiglote, la voie de la viande on l'**appelle** meri ou ysophagus et la voye moyenne entre ces deux conduis, on l'**appelle** gorgeron et la languette qui est sur ces deux conduis, on l'**appelle** la vule ou luette. (GORDON, *Prat.*, c.1450-1500, IV, 1)

Persicaire, c'est une herbe qui a les feules qui ressemblent à feules de de pechier; l'en l'**appelle** currago ou currage. [...] Aucuns l'appellent sanguinaire. (Sec. Sal., éd. Camus, xv^e s., p. 102)

Il est possible de maîtriser le choix des définissants en se référant aux connaissances disponibles actuellement sur l'AF et sur les sciences et techniques au Moyen Âge. Toutefois, la configuration d'ensemble du concept reste sujette à discussion si elle ne se trouve pas attestée dans un énoncé. Souvent, comme c'est le cas pour le sens de *cyclamen* (cf. plus haut), elle résulte de recoupements d'énoncés. Pour cette raison elle se présente davantage comme une proposition que comme le reflet strict et exhaustif d'un état de conscience¹⁵.

Comme on peut s'y attendre, le périmètre sémantique des termes médiévaux n'est pas isomorphe de celui de leurs équivalents modernes. Deux grands faisceaux d'explications peuvent être avancés pour justifier de ce constat.

14. *Apotaine* : hippopotame. Pour la conscience médiévale l'hippopotame est un poisson !

15. À supposer qu'un tel objectif soit tenable, et même qu'il ait un sens. Les définitions lexicales et lexicographiques en synchronie contemporaine sont-elles autre chose que de simples propositions ?

En premier lieu, l'évolution du monde réel et des représentations qui l'accompagnent, laquelle aboutira à ce que certaines dénominations voient leur représentation sémantique changer. Ainsi, on ne définit plus aujourd'hui la chirurgie comme on le faisait au Moyen Âge, la pratique et la profession qui l'exerce ayant évolué. Ainsi, si l'on peut définir en To la chirurgie, et le nom *chirurgie*, par :

Partie de l'art médical qui consiste à faire avec la main ou à l'aide d'instruments certaines opérations sur le corps de l'homme. (Wiktionnaire)

A specialty in which manual or operative procedures are used in the treatment of disease, injuries, or deformities. (Termscience¹⁶)

Partie de la thérapeutique qui met en œuvre des procédés manuels et l'usage d'instruments, et qui groupe elle-même diverses spécialités selon les organes ou appareils intéressés (chirurgie thoracique), les buts recherchés (chirurgie réparatrice), etc. (*Trésor de la langue française*)

en T-1 sa signification sera circonscrite à :

Pratique de l'art de guérir les maladies et les maux du corps par des incisions, cautérisations et remplacement des os, et autres opérations à l'aide d'instruments.

L'autre grand faisceau de causes tire ses références non pas de la mutation des pratiques mais de celle des savoirs eux-mêmes et des représentations qui y sont associées. Ainsi l'une des différences majeures entre les représentations médiévale et moderne tient-elle à la coupure épistémologique engagée à la Renaissance et qui a entraîné, jusqu'au XVIII^e siècle, une refonte des principes théoriques et méthodologiques des sciences et des techniques. Pour cette raison certains domaines de connaissances sont à repenser totalement, comme la botanique (et l'ensemble des sciences de la nature), pour être abordés dans leur configuration médiévale. Ainsi, le terme *cyclamen* se définit-il aujourd'hui par :

Plante **herbacée** à gros **tubercule** d'où naissent des racines, à feuilles en cœur, d'un vert sombre moucheté de blanc et

16. En ligne : <http://www.termssciences.fr/-/Index/Rechercher/Rapide/Naviguer/Arbre/>.

d'un rouge pourpre sur leur face inférieure, à fleurs solitaires pendantes, blanches, rosées ou violines, surplombant un long **pédoncule** recourbé, présentant une **corolle** à cinq **pétales** renversés, tordus sur eux-mêmes, et donnant naissance à un fruit en **capsule** arrondie. (*Trésor de la langue française*)

Plante vivace tuberculeuse de la famille des Primulacées, selon la classification classique et selon la classification phylogénétique (APG III) (Anciennement Myrsinacées APG II). Le cyclamen n'a pas de parenté nette avec les autres primulacées, quoiqu'il ressemble aux *Dodecatheon* d'Amérique du Nord par ses pétales renversés. (Wikipedia)

Définitions que l'on peut compléter par la taxinomie botanique :

Règne *Plantae*, Sous-règne *Tracheobionta*, Division *Magnoliophyta*, Classe *Magnoliopsida*, Sous-classe *Dilleniidae*, Ordre *Primulales*, Famille *Primulaceae*, Genre *Cyclamen* L., 1753.

Ces connaissances témoignent de savoirs modernes (nous soulignons ci-dessus), lesquels n'étaient pas disponibles au Moyen Âge. Dans le *DFSM*, le terme *cyclamen* se définira par :

Plante des bois dont la racine est en forme de pomme et qui servait notamment de nourriture pour les porcs.

Cette définition est conforme aux informations livrées par les corpus. L'exemple de *cyclamen* ci-dessus illustre une évolution conceptuelle entre T-1 et To, laquelle se traduit par un accroissement, une précision plus grande et une systématisation des informations classifiantes à la période moderne. En l'occurrence, elles expriment les propriétés intrinsèques du référent. Mais pour certains termes¹⁷, la définition impose une prise en compte de la signification spécifique des définissants en T-1, comme le montrent les exemples d'*écrevisse*, de *boterel*, d'*apostume*, d'*humeur* et d'*armoise* (ci-dessous). Sont marqués d'un astérisque à droite les définissants présentant une pertinence pour la conceptualisation du défini en T-1, mais dont

17. Aucune quantification ni proportion n'est actuellement possible, mais les études menées sur les lettres A, B et C montrent que, comme on pouvait s'y attendre, l'immense majorité du vocabulaire est affectée.

l'interprétation doit tenir compte de la signification dans cette même synchronie :

écrevisse: Petit poisson* de mer protégé par une enveloppe rigide, qui se déplace en reculant.

boterel: Ver* venimeux aux yeux rouges qui fréquente les lieux humides et subit une mue, crapaud.

abeille: Mouche* qui fabrique le miel et la cire.

apostume: Toute enflure, grosseur causée par une corruption* des humeurs*.

bègue: Personne dont la parole est corrompue* comme par un tremblement.

armoïse: Herbe* chaude* et sèche* avec des feuilles plus grandes et grasses que l'absinthe*, de longues tiges et de petites fleurs blanches odorantes en été¹⁸.

Pour ne prendre que deux exemples, la classification médiévale des animaux ne s'opérait pas sur la base de propriétés anatomiques, comme c'est le cas aujourd'hui, mais sur la prise en compte du milieu de vie dominant. Pour cette raison, tous les animaux vivant dans un milieu aqueux sont considérés comme des poissons, y compris la baleine et l'hippopotame! Sur un modèle homologue les végétaux sont classés en agronomie selon leur localisation par rapport au niveau du sol: ceux qui, comme le cyclamen, poussent sous terre¹⁹ sont des *racines*; ceux qui croissent au-dessus de la surface de la terre sont des *fruits*; ceux qui, comme le poireau, occupent une position intermédiaire, sont des *herbes*.

Restituer le concept dans sa configuration T-1 implique un certain nombre de contraintes, dont :

- celles portant sur le choix et l'emploi des définissants (cf. paragraphes précédents);
- celles pesant sur le contrôle de la métalangue de description en To. Elles affectent le schéma de la définition, le libellé de clauses définitionnelles de sorte à conférer à l'ouvrage une

18. *Chaud* et *sec* renvoient à la théorie des quatre qualités fondamentales, issue de la *Physique* d'Aristote.

19. Ou sont estimés comme tels.

valeur heuristique non pas seulement quant à la connaissance (en To) de la terminologie médiévale, de sa structuration sémantique, de sa polysémie, de sa construction morphologique ou polylexicale, mais aussi quant à son organisation en tant qu'ensemble de représentations culturelles hiérarchisées qui procèdent d'une conception du monde en même temps qu'elles la véhiculent et la font ou non évoluer.

C'est ce second faisceau de contraintes qui sera envisagé dans les paragraphes qui suivent.

Le schéma de définition : la logique des classes

La sémiotique d'un dictionnaire diachronique peut être perçue (et conçue) selon deux modalités :

- (1) l'une, intralinguale, confronte une langue ou un état de langue avec un(e) autre. Elle viserait à exprimer en français moderne ce qui est énoncé et conçu en langue médiévale ;
- (2) une seconde modalité, extralinguale, confronte le lexique et le monde, le texte de dictionnaire étant alors autosuffisant. Il n'a besoin d'aucun point de référence pour être interprété.

Une logique intralinguale repose sur plusieurs prérequis. Un premier prérequis considère que les états de langue sont suffisamment conjoints pour que l'un (To) fournisse l'interprétant de l'autre (T-1). Dans ce cas, l'état de langue To véhicule l'ensemble des paramètres (orthographiques, sémantiques, syntaxiques, culturels) nécessaires à l'interprétation de T-1. Un second prérequis, découlant du précédent, favorise dans ce cas l'adoption du modèle du dictionnaire bilingue : la traduction intralinguale de T-1 vers To s'opère dans le même schéma sémiotique que celui opposant deux langues différentes, l'une inconnue, et l'autre connue. Dans ce cas, le modèle de définition retenu est synonymique : un terme de la langue A (en l'occurrence T-1) est traduit par un terme en langue B (To), qui est son équivalent.

La définition par synonymie repose sur un postulat, difficilement défendable en soi et indéfendable dans le cas de la terminologie médiévale, selon lequel le concept associé au

terme serait un invariant diachronique. L'exploitation des corpus à la base du *DFSM* confirme que les concepts sont ancrés dans la langue et la culture de leur temps. On aboutit là à une forme de paradoxe dans la mesure où les référentiels terminographiques produits ces dernières décennies reposent précisément sur l'invariance du concept. À décharge, il convient de préciser qu'ils portent sur la synchronie contemporaine. Les dictionnaires spécialisés produits durant les siècles passés (notamment celui de Thomas Corneille et l'*Encyclopédie* de Diderot et d'Alembert) s'appuyaient sur le même principe car ils opéraient toujours en synchronie contemporaine. La terminologie diachronique est très jeune. Lorsqu'elle a débouché sur des référentiels, ceux-ci se sont inspirés de modèles qui n'intègrent pas la variation diachronique du sens, car ils se fondent sur la permanence²⁰ du réel et calquent le modèle du dictionnaire bilingue :

oestre, n. m. : « grosse mouche, taon » ;
oeuchine, n. f. : « atelier, officine », en général ; en particulier, « atelier de foulon, de teinturier, de brasseur²¹ » ;
cyclamen, subst. masc. : « Cyclamen » ;
apostume, subst. masc. : Au propre, « Tumeur, abcès ». Au fig., « Blessure qui s'envenime²² ».

Soit ils définissent le terme d'AF par son équivalent en FM (*oestre*, *cyclamen*, *oeuchine*, *apostume* [sens propre]), soit ils le glosent quand le sens en AF ne dispose pas d'équivalent en FM (sens figuré d'*apostume*). Ces dictionnaires sont des objets hybrides, qui ne réservent la définition logique qu'à l'écart ontologique²³ (et non pas sémantique) entre les synchronies T-1 et To.

Le *DFSM* procède différemment. Optant pour une perspective sémasiologique, il ne considère pas la représentation du réel comme identique en T-1 et To. Pour cette raison, quelle que soit l'entrée, la définition qui lui est appliquée procède sur le modèle logique, par incluants et différences spécifiques, que

20. Souvent dans une représentation très approximative de celui-ci.

21. *Oestre* et *oeuchine*, extraits du dictionnaire de Godefroy.

22. *Cyclamen* et *apostume*, extraits du *DMF*.

23. Le terme renvoie à une réalité spécifique de T-1, orpheline en To.

ce mot soit inconnu en FM (*caladre*), qu'il soit connu avec la même signification²⁴ (*calcul*) ou avec une signification différente (*abominable*, *abonder*) :

caladre : Oiseau blanc qui a le pouvoir de prédire la mort et de détruire certaines maladies par le regard.

abominable : Qui provoque des nausées.

abonder : Être en grande quantité ou en excès, en parlant d'humeurs*, de superfluités ou encore du lait maternel.

calcul : Opération portant sur des nombres, des caractères numériques, ou des objets représentant des nombres et destinée à obtenir d'autres nombres.

Il peut se produire, localement, qu'un terme soit défini par un synonyme, mais uniquement si celui-ci présente une signification spécifique à la langue médiévale :

accidents de l'âme : Passions*

mais le fait n'est pas une généralité à l'échelle du dictionnaire.

Adopter un format de définition logique impose que l'on se penche sur la distinction entre propriétés intrinsèques (PI) et propriétés extrinsèques (PE) du référent. Les PI décrivent l'entité sous son angle concret, perceptuel : dimensions, poids, texture, silhouette, couleur, aspect général, etc. Les PE l'envisagent sous son angle fonctionnel : usage, agents impliqués, résultats attendus, modalités d'action, etc. Une définition par PI est prédisposée à caractériser les entités concrètes, mais peut également s'appliquer à certains adjectifs ou à des verbes associés à des effectuations perceptuelles. Les PE, quant à elles, sont réservées à la seule caractérisation des artefacts, concrets ou abstraits. Pour cette raison, la définition d'un végétal différera selon qu'elle sera indexée en botanique, en agronomie ou en diététique. Dans ces deux derniers domaines, les PE jouent un rôle fondamental dans la caractérisation du référent, alors

24. D'après l'examen des lettres A, B et C du *DFSM*, les unités présentant une signification identique en FM et en AF sont en nombre extrêmement restreint. Par ailleurs, nous restons très prudent sur ce constat d'isomorphisme dans la mesure où il serait certainement contredit par une approche convoquant des niveaux de spécialité plus élevés.

qu'elles sont absentes de la description d'une espèce naturelle²⁵. Ce type de caractérisation convient particulièrement aux noms²⁶ et à certains verbes, dont la contrepartie référentielle peut être décrite en termes perceptuels (*marcher, courir, s'agiter*) et orientée vers une finalité.

L'exemple ci-dessous montre la répartition des PE et PI dans la définition d'*ammoniac* :

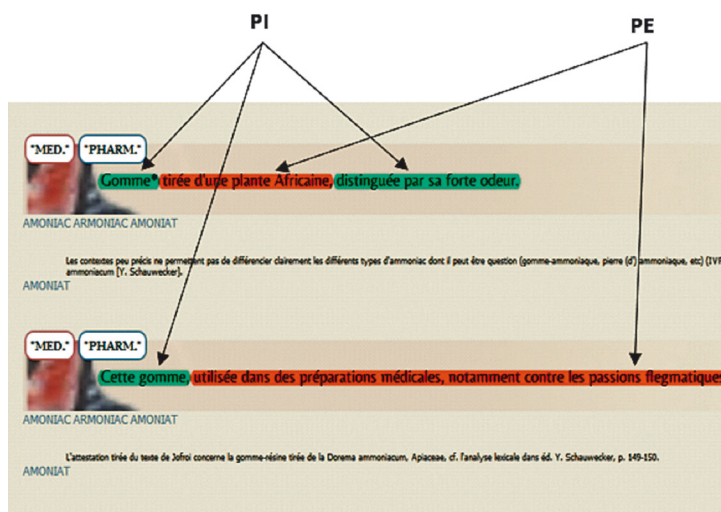


Fig. 2. Répartition des PF et PI dans *ammoniac* (DFSM, capture d'écran)

Les définitions de *caladre* et *calcul* montrent la limite extrême de cette répartition (nous soulignons en gras les PI) :

caladre: **Oiseau blanc** qui a le pouvoir de prédire la mort et de détruire certaines maladies par le regard.

calcul: Opération portant sur des nombres, des caractères numériques, ou des objets représentant des nombres et destinée à obtenir d'autres nombres.

La répartition entre PE et PI permet, pour la nomenclature des noms, d'opérer d'importants dégroupements entre domaines.

25. En tant que telles, les espèces naturelles sont totalement dépourvues de PI.

26. Aux noms concrets, mais aussi aux abstraits qui dénomment des processus (cf. *calcul, opération, amputation...*).

C'est à la définition de justifier le rattachement de tel terme à son domaine. À cet égard elle possède une fonction signalétique. Tout terme se rattache à un domaine: il faut entendre par là qu'il n'existe pas de terme sans domaine²⁷. Ainsi, il n'est guère suffisant de définir *abject*, en Médecine, par « repoussant » seulement. Cette contrainte vaut aussi bien pour des termes dont l'emploi est strictement spécialisé (*alchane*, *bétoine*) que pour ceux qui disposent par ailleurs d'une signification en langue courante ou d'une saillance particulière dans d'autres domaines (*acteur*):

alchane (Médecine): **La poudre** obtenue à partir de cette plante²⁸, en tant qu'elle purifie et soigne la peau, nourrit et teint les ongles et les cheveux.

bétoine (Médecine): **Cette herbe***, en tant qu'elle possède de nombreuses vertus médicinales, et notamment favorise la conception, améliore la vue, soigne les douleurs et les blessures à la tête, favorise la guérison des plaies, traite les écrouelles ou la goutte froide.

acteur (Médecine): Auteur d'un traité de médecine.

Quelques clauses définitionnelles

Les contraintes représentationnelles qui pèsent sur la métalangue de description rendent inévitable l'adoption de formules définitionnelles, lesquelles fonctionnent comme des constructions logico-syntaxiques permettant de traiter des phénomènes identiques, à l'échelle de toute la nomenclature et quels que soient les domaines concernés.

Le démonstratif et la relation anaphorique

Les articles du *DFSM* sont conçus sur le modèle de la polysémie, car le dictionnaire entend privilégier l'unité psycholinguistique du signe, par-delà la contrainte domaniaire. Se pose dès lors la question du principe de classification des sens. Plusieurs axes peuvent être retenus, privilégiant un classement :

27. Ce serait en propre la définition du lexique, aux yeux du terminologue : être un ensemble d'unités non classables dans un domaine, être à soi un non domaine. Nous avons déjà pris position sur ce point (Petit, 1995, 2010).

28. Définition anaphorique, voir plus bas.

- (1) en fonction de la fréquence du sens dans le corpus ;
- (2) historique croissant ;
- (3) par ordre alphabétique de domaine ;
- (4) par génération de sens, notamment dans le cadre d'une polysémie systématique ;
- (5) par types d'emplois : libres vs. contraints, au plan syntaxique ou morphosyntaxique.

Opter pour un modèle unique s'avère une solution peu adaptée compte tenu de l'hétérogénéité du matériau à traiter. Pour cette raison, la configuration sémantique disponible pour chaque unité indique l'axe à privilégier. L'observation des données montre néanmoins que la polysémie systématique fournit un vecteur de classement récurrent, tous domaines confondus. Nombre d'articles présentent donc des définitions anaphoriques, fondées sur l'emploi du démonstratif. Ainsi à l'article *bette* (nous soulignons) :

Herbe* qui pousse en s'étalant au sol, dont les feuilles ressemblent à celles du chou et qui présente une grosse racine.

Cette herbe en tant qu'elle est cultivée pour ses feuilles dont se nourrissent les hommes.

Les feuilles **de cette herbe*** en tant qu'aliment à privilégier dans le régime de santé, en particulier des hommes de complexion chaude.

Les feuilles de **cette herbe***, en tant qu'elles peuvent servir pour confectionner un emplâtre.

La partie dure des feuilles **de cette herbe***, côte de bette.

La détermination syntagmatique : « en parlant de »

Sont concernés les adjectifs et noms dont l'emploi est syntagmatiquement contraint. Ils fonctionnent de manière stable et récurrente comme expansions (arguments appropriés, pour le lexique-grammaire) d'un nom ou d'un paradigme de noms présentant un profil sémantique commun. Le tour « en parlant de » a pour fonction d'exprimer la solidarité syntagmatique entre l'entrée et l'unité terminologique à laquelle elle est liée (nous soulignons) :

commixtion: Union par mélange, **en parlant de** plusieurs ingrédients qui entrent dans une préparation médicale.

L'expression des propriétés extrinsèques : « en tant que »

Cette clause définitionnelle a pour fonction de justifier un dégroupement polysémique sur la base d'une acquisition de PE. Elle est utilisée pour marquer l'articulation d'un concept doté de PE spécifiques sur un autre qui en est *a priori* dépourvu – nom d'espèces naturelles, d'organes, de parties du corps (nous soulignons²⁹):

anis: Herbe* qui ressemble au fenouil.
 Cette herbe* **en tant qu'**elle est cultivée.

L'expression de l'hyponymie fonctionnelle : « tout »

La structure d'une définition logique impose que soit renseigné l'hyperonyme du défini, ou du moins un superordonné si celui-là n'est pas disponible. Lorsque l'accrochage hiérarchique entre concepts s'opère uniquement sur la possession de PE (et non pas d'un complexe de PI et de PE), l'hyperonyme est alors un nœud fonctionnel à l'extension vaste et privé de rattachement catégoriel précis. Il peut alors être occupé par une dénomination occurrence (Petit, 2009), pour peu que ses PE coïncident et fournissent un site d'accrochage valide. Au plan définitionnel, le marqueur *tout* vient manifester l'existence de ce nœud fonctionnel en précisant le choix opéré par le dictionnaire pour le saturer (nous soulignons) :

apostume: **Toute** enflure, grosseur causée par une corruption* des humeurs*.

L'hyponymie : « une des N de »

Dans un registre opposé, la glose *une des N* [*espèce, forme*] *de* indique un rattachement hyponymique en précisant la nature du lien. Ce type de précision a pour effet de marquer la spécificité de l'entrée dans le paradigme de ses cohyponymes :

29. Voir également les définitions 2, 3 et 4 de *bette*, plus haut.

ascite: **L'une des formes d'hydropisie**, caractérisée par son origine essentiellement aqueuse ainsi que par sa localisation au niveau du ventre, et considérée comme la pire.

Conclure un si bref exposé serait une gageure. Nous avons laissé en chemin de nombreuses pistes dont nous évoquerons les deux principales, promises pour un exposé futur :

- l'identification des domaines de connaissance et le rattachement des termes susceptibles de les valider ;
- le traitement informatisé du dictionnaire et son influence sur la présentation de l'information, son codage et l'exploitation heuristique des données ;
- le contrôle de la métalangue de description dans une double perspective : (i) celle de l'harmonisation de la représentation sémantique des données de T-1 avec celles présentes en To ; (ii) celle du codage lexicosyntaxique de l'information en vue de l'exploitation heuristique (et informatisée) des données.

Néanmoins, les quelques réflexions menées ici montrent que la terminographie médiévale reste largement à construire, *a fortiori* si elle débouche sur une exploitation informatique (ce qui est le lot de tous les référentiels terminologiques actuels). Les remarques formulées dans ces pages n'engagent pas seulement le plan méthodologique de la description. En amont, elles mettent en cause la perspective sémiotique à projeter sur la terminologie diachronique, et en particulier sur celle du Moyen Âge.

Références bibliographiques

- BÉJOINT, Henri et THOIRON, Philippe (dir.), *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000.
- BINON, Jean, VERLINDE, Serge, BERTELS, Ann et SELVAT, Thierry, *DAFA = Dictionnaire d'apprentissage du français des affaires*. En ligne : <http://www.projetdafa.net/>.
- CABRÉ, Maria-Teresa, *La Terminologie. Théorie, méthode et applications*, trad. J. HUMBLEY et M. CORMIER, Paris, Armand Colin/ Presses de l'université d'Ottawa, 1998.

- CADIOT, Pierre, « Propriétés extrinsèques en sémantique lexicale », *Journal of French Language Studies*, n° 7, 1997/2, p. 127-146.
- CADIOT, Pierre et LEBAS, Franck (dir.), « La constitution extrinsèque du référent », *Langages*, n° 150, 2003.
- CAMPENHOUDT, Marc, « Le terme : condensation syntaxique et condensation des connaissances en langue spécialisée », 2010. En ligne : http://www.termisti.org/romanica_w.pdf.
- CANDEL, Danielle, « La représentation par domaines des emplois scientifiques et techniques dans quelques dictionnaires de langue », *Langue française*, n° 43, 1979.
- CONDAMINES, Anne, REBEYROLLE, Josette et SOUBEILLE, Annie, « Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus », dans *Actes du colloque Euralex International Congress*, Lorient, Université de Lorient, 2004, p. 547-557.
- CréalScience, *Dictionnaire du français scientifique médiéval*. En ligne : <http://www.crealscience.fr/>.
- JACQUART, Danielle et THOMASSET, Claude (dir.), *Lexique de la langue scientifique (astrologie, mathématiques, médecine...): Matériaux pour le « Dictionnaire du moyen français » (DMF)*, avec la collab. de Sylvie BAZIN-TACHELLA, Jean-Patrice BOUDET, Thérèse CHARMASSON, Joëlle DUCOS, Hervé L'HUILLIER (INALF-CNRS), Paris, Klincksieck, 1997.
- DE BESSÉ, Bruno, « Le domaine », dans BÉJOINT, Henri et THOIRON, Philippe (dir.), *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, p. 182-197.
- DEAF = *Dictionnaire étymologique de l'ancien français*. En ligne : <http://www.deaf-page.de/fr/>.
- DMF = *Dictionnaire du moyen français*. En ligne : <http://www.atilf.fr/dmf/>.
- DURY, Pascaline, « Les noms du pétrole : une approche diachronique de la métonymie onomastique », 2008. En ligne : lexis.univ-lyon3.fr/IMG/pdf/Lexis_1_Dury.pdf.

- GODEFROY, Frédéric, *Dictionnaire de l'ancienne langue française et de tous ses dialectes, du IX^e au XV^e siècle*, Paris, F. Vieweg, 1881.
En ligne : <http://micmap.org/dicfro/>.
- HUMBLEY, John, « La terminologie française du commerce électronique, ou comment faire du neuf avec de l'ancien. Vers une géomorphologie lexicale », dans *Terminologie et plurilinguisme dans l'économie internationale. Actes de la V^e Journée scientifique de REALITER*, Milan, 2009. En ligne : <http://realiter.net/spip.php?article1847>.
- L'HOMME, Marie-Claude, *La Terminologie : principes et techniques*, Montréal, Presses de l'université de Montréal, 2004.
- OTMAN, Gabriel, *Les Représentations sémantiques en terminologie*, Paris, Masson, 1996.
- PETIT, Gérard, « Le traitement des variantes graphiques dans les dictionnaires Larousse et spécifiquement dans *Le Petit Larousse illustré* », dans *Langue française. La variation graphique et les rectifications de l'orthographe française (1990)*, Paris, Larousse, 1995, p. 40-51.
- , *La Dénomination. Approches lexicologique et terminologique*, Louvain, Peeters, coll. « Bibliothèque de l'Information grammaticale », 2010.
- PICTON, Aurélie, *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*, thèse de doctorat, Université de Toulouse 2, 2009.
- REY-DEBOVE, Josette, *Étude linguistique et sémiotique des dictionnaires français contemporains*, Den Haag, Mouton, 1971.

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.ovi.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.ovi.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 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 Szeceł	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