

J. LEEMANS & L. JOCQUÉ, *Corpus Christianorum 1953-2003:*
Xenium natalicium, Turnhout, 2003, p. 168-172

CORPVS CHRISTIANORVM THESAVRVS PATRV M GRAECORVM

Depuis plusieurs décennies déjà, les développements des techniques de calculs statistiques, de calculs de fréquences et de tris sont appliqués aux sources écrites; si les langues modernes ont été les premières à en bénéficier, les langues anciennes ont rapidement suivi le mouvement. C'est ainsi, par exemple, que les sources latines font l'objet de traitements informatiques depuis de nombreuses années, et que le *Corpus Christianorum* leur a accordé une large place dans le cadre des *Thesauri Patrum Latinorum* et des *Instrumenta Lexicologica Latina*. Ce sont précisément les démarches novatrices, dans le domaine du latin, du Centre de Traitement Électronique des Documents (CETEDOC) de l'Université catholique de Louvain, sous la direction du professeur Paul Tombeur, qui ont ouvert la voie à un traitement similaire des sources grecques et permis la création d'une série intitulée *Thesaurus Patrum Graecorum*.

Développé au départ sur les bases des travaux du CETEDOC sur les textes latins, le traitement des sources grecques se voulait avant tout lexical, de type morpho-sémantique: il s'agissait, dans un texte donné, de classer toutes les formes du texte, de regrouper les formes identiques, d'attribuer à chacune d'entre elles un lemme ou « entrée de dictionnaire », et de lever les ambiguïtés résultant des homographies (formes semblables mais relevant des deux ou plusieurs lemmes possibles). Les résultats étaient présentés dans des volumes proposant aux lecteurs une série d'instruments, allant de la concordance lemmatisée elle-même (les formes présentées dans leurs contextes textuels), aux index inverses, aux listes de lemmes et de formes et à diverses listes de fréquence. La taille des textes traités empêchant de publier l'ensemble de ces outils sur papier, le choix de la microfiche s'est imposé comme le moyen le plus commode, et le plus stable dans la durée. Pour la réalisation de l'ensemble de ces outils, le CETEDOC a adapté au grec les logiciels développés pour le latin.

Quelques options spécifiques ont cependant été prises du côté du grec, dont deux méritent d'être signalées. La première fut de chercher à traiter des corpus d'auteurs complets, au détriment d'index d'œuvres séparées, ce qui a rendu inutile la création d'une série qui aurait servi

de parallèle aux *Instrumenta Lexicologica Latina*, puisque les résultats étaient destinés à paraître dans les seuls *Thesauri*. La seconde option fut de ne préparer et de ne publier que des instruments lemmatisés, sans passer par l'équivalent des volumes de la « Series A. Formae » du latin. Ces options ont entraîné deux conséquences.

La première conséquence était que la volonté de traiter des corpus complets rendait plus important encore le choix des auteurs. Pour des raisons qui tiennent aux projets en cours à l'Université catholique de Louvain, le premier auteur analysé fut Grégoire de Nazianze, un auteur grec chrétien, Père de l'Église du iv^e s. (sur cet auteur et le projet qui le concerne, voir la présentation du *Corpus Nazianzenum*). Ce choix a naturellement orienté le développement du grec au CETEDOC vers le domaine de la patristique, et cela d'autant plus que d'autres centres d'analyse statistique des langues anciennes concentraient leurs efforts sur des œuvres de la littérature grecque classique. Mais les Pères grecs sont aussi des auteurs représentatifs de la société byzantine: ils utilisent les mêmes mots, les mêmes cadres conceptuels et le même bagage culturel que les auteurs byzantins profanes; tous ont en commun la volonté de s'inscrire dans l'héritage de l'Antiquité classique, et cette volonté confère à leur langue et à leur lexique des caractères bien particuliers, faits de conservatisme et d'innovations. C'est pourquoi les efforts d'analyse des textes grecs se sont ouverts à des œuvres non patristiques, en particulier aux œuvres des historiens et chroniqueurs byzantins. Pour éviter une trop grande dispersion, deux axes ont été suivis de manière prioritaire mais non exclusive: d'une part, les œuvres de l'ensemble des Pères Cappadociens (Grégoire de Nazianze, Basile de Césarée, Grégoire de Nysse [en préparation], Amphiloque d'Iconium, Astère d'Amasée et Firmus de Césarée), et, d'autre part, les historiens du vi^e s. (Procopé de Césarée [publié], Agathias le Scholastique, Ménandre Protecteur, Théophylacte Simocatta [en préparation], etc.). Ces deux axes permettront des analyses non seulement sur des corpus d'auteurs, mais aussi sur des corpus de groupes d'auteurs.

Si les historiens ont trouvé place, aux côtés des Pères, dans une collection intitulée *Thesaurus Patrum Graecorum*, c'est en raison de leur ap-

partenance commune à un monde chrétien, de tradition classique et de langue grecque. Derrière cette uniformité apparente se cache toutefois une grande diversité ! Dans le domaine de la langue, le monde byzantin est caractérisé par un phénomène de diglossie, c'est-à-dire d'écart croissant entre une langue littéraire, fidèle à ses modèles classiques, et une langue parlée, lieu d'adaptations, d'influences et de créations de toutes sortes. Comme il est naturel, des usages de la langue parlée finissent par entrer, de manière plus ou moins organisée, dans la langue littéraire ; plus les auteurs sont tardifs, plus leurs œuvres sont perméables à ces innovations de la langue parlée. L'analyse lexicale des textes doit rendre compte de ce phénomène, et la lemmatisation doit pouvoir les intégrer.

C'est la seconde conséquence des options du projet grec. Désireux de rendre compte des réalités lexicales de textes aussi divers que des œuvres patristiques et des œuvres historiographiques, dont les dates de composition partent du iv^e siècle après Jésus-Christ pour avancer dans le temps jusqu'à la fin de l'époque byzantine, le système d'analyse nécessitait un travail de réflexion approfondie sur la nature des lemmes, sur la nature des néologismes, des formes populaires, des emprunts, etc. Des innovations morphologiques dans la déclinaison ou dans la conjugaison impliquent-elles la création de lemmes nouveaux ? Quels lemmes définir pour des mots résultant d'emprunts à des langues étrangères et qui apparaissent sous plusieurs formes possibles ? Telles sont quelques questions auxquelles le projet devait apporter des réponses. Ce qui sous-tend l'ensemble de l'analyse des formes, c'est la conception du rapport de la forme à son lemme, et l'ensemble des lemmes doit proposer une représentation aussi exacte que possible de l'étendue du champ lexical de la langue grecque. La liste des lemmes et les liens entre lemmes et formes constituent le « Dictionnaire Automatique Grec » (D.A.G.) ; le D.A.G. est le « cœur » du traitement des textes ; le D.A.G. est lui-même un projet scientifique, en arrière-plan de la publication des *Thesauri*, et c'est lui qui assure la cohérence d'ensemble de la collection.

La conception du dictionnaire comme projet confère à la série *Thesaurus Patrum Graecorum* une identité particulière: elle n'est pas qu'une simple collection de volumes juxtaposant des concordances d'auteurs. Elle est aussi, et surtout, l'expression d'un projet plus vaste, dont chaque volume est un élément et une étape. À terme, les volumes de *Thesauri* sont destinés à présenter au public une image de la richesse du domaine lexical grec: dans cette optique, la série *Thesaurus Patrum Graecorum* recèle une grande originalité parmi les publications relevant de la linguistique computationnelle.

Prof. Dr. Bernard Coulie
Directeur *Thesaurus Patrum Graecorum*

Thesaurus Patrum Graecorum: Université catholique de Louvain
Institut orientaliste
Place Blaise Pascal 1
B-1348 Louvain-la-Neuve (Belgique)
tél. +32 10 473793; fax +32 10 472001
www.corpuschristianorum.org
<http://tpg.fltr.ucl.ac.be>
e-mail: coulie@ori.ucl.ac.be