

---

## Avancées dans le traitement automatique du grec ancien à l'U.C.L.

L'analyse des textes au service d'une description lexicale de la langue.  
Une description lexicale de la langue au service de l'analyse des textes

**Bastien KINDT**

*Institut orientaliste  
Collège Érasme  
Place Blaise Pascal, 1  
B-1348 Louvain-la-Neuve (Belgique)  
e-mail: kindt@ori.ucl.ac.be*

---

### ABSTRACT.

*The «Research project in Greek lexicology» carried out at the U.C.L. approaches the study of Ancient Greek language by analysing its vocabulary. This research is based on a direct examination of the sources by means of automated tagging. This work generates lexical data which, gathered in an electronic dictionary, will offer an exhaustive and coherent image of Greek vocabulary. The lemmatisation, exercise of regrouping various actualizations of a word under the same lexeme, imposes a reflexion on the standard of examination and on the treatment methods of lexical ambiguities. The inherent problems to Ancient Greek in these two fields, solutions brought and progress made are evoked.*

**KEYWORDS :** Ancient Greek, tagging, corpus, Thesaurus Patrum Graecorum, normalization, ambiguity, unicode, unitex

### RÉSUMÉ.

*Le «Projet de recherche en lexicologie grecque» mené à l'U.C.L. aborde l'étude de la langue grecque ancienne par l'analyse de son lexique. Ces recherches se fondent sur un examen direct des sources par le biais de la lemmatisation automatisée. Ces travaux génèrent des données lexicales qui, rassemblées dans un dictionnaire électronique, offriront du lexique une image exhaustive et cohérente. La lemmatisation, exercice de regroupement des différentes actualisations d'un mot sous une même unité lexicale, impose une réflexion sur la norme de dépouillement et sur les méthodes de levée des ambiguïtés lexicales. Les problèmes inhérents au grec ancien en ces deux domaines, les solutions apportées et les avancées réalisées sont évoqués.*

**MOTS-CLÉS :** Grec ancien, lemmatisation, corpus, Thesaurus Patrum Graecorum, norme, ambiguïté, unicode, unitex

---

### 1. Le Projet de recherche en lexicologie grecque

Le grec est une langue vivante attestée depuis le second millénaire avant notre ère jusqu'à l'époque contemporaine. Par «grec ancien» on entend le grec classique, c'est-à-dire l'ionien-attique. C'est une langue littéraire écrite, issue de l'antiquité classique. Prise comme

modèle par les lettrés des époques postérieures, elle sera le véhicule de la culture jusqu'à la fin de l'époque byzantine. Peu perméable aux innovations propres à la langue parlée qui apparaissent dès les époques hellénistique et romaine, d'abord dans les sources non littéraires, puis progressivement dans les textes littéraires, elle demeure empreinte de classicisme et de purisme, ce qui en assure la pérennité pendant vingt siècles et justifie une étude synchronique, même sur une aussi longue période<sup>1</sup>.

L'étude du grec ancien bénéficie du développement contemporain des outils informatiques en matière de stockage et d'analyse des données linguistiques. On ne peut passer sous silence la base de donnée textuelle du *Thesaurus Linguae Graecae (T.L.G.)*<sup>2</sup>, ni le projet *Perseus*<sup>3</sup>, qui fournit en ligne une version du *Greek-English Lexicon* de H.G. Liddell et R. Scott<sup>4</sup> ainsi qu'un analyseur morphologique, ou les réalisations de *Olive Tree Bible Software*<sup>5</sup>, qui diffuse sur cd-rom les différentes versions annotées de la Bible, pour ne citer que ces quelques exemples. Ces réalisations ont cependant été conçues indépendamment les unes des autres, sans un standard d'encodage ou d'étiquetage qui permettrait de les rassembler. Les *corpus* annotés, sur lesquels des systèmes de traitement automatique seraient applicables, à l'instar de ce qui existe pour les langues modernes, se font donc encore attendre.

L'Institut orientaliste de l'Université catholique de Louvain (Louvain-la-Neuve, Belgique) abrite depuis 1991 un «Projet de recherche en lexicologie grecque» dirigé par le Professeur Bernard Coulie<sup>6</sup>. Les travaux qui y sont menés ont pour objet l'analyse de la langue grecque ancienne et, plus particulièrement, l'étude de son lexique. Le but de ces recherches est de fournir une description exhaustive et cohérente des faits lexicaux observés dans les sources. Cette description s'élabore progressivement grâce à l'exploration systématique des textes par le biais de la lemmatisation, exercice de classement des *formes* rencontrées *in textu* sous un *lemme*, c'est-à-dire une entrée lexicale. Dans le cas d'une langue à flexion, le lemme permet de regrouper sous une même unité les différentes actualisations d'un même mot<sup>7</sup>. Les formes relevant de la flexion nominale ont un lemme au nominatif singulier; les formes conjuguées ont un lemme à la première personne du présent, actif ou moyen selon l'usage propre au verbe.

---

<sup>1</sup> Cfr BROWNING, *Greek*, p. 49-50.

<sup>2</sup> Cfr <http://www.tlg.uci.edu>.

<sup>3</sup> Cfr <http://132.236.125.30/default.htm>.

<sup>4</sup> Les dictionnaires sont cités sous l'abréviation du nom de leur(s) auteur(s). Le *Greek-English Lexicon* = LS; cfr les Références, *infra*.

<sup>5</sup> Cfr <http://www.OliveTree.com>.

<sup>6</sup> L'intitulé complet est «Projet de recherche en lexicologie grecque: élaboration de concordances lemmatisées des auteurs grecs patristiques et byzantins» (cité dorénavant par la simple mention de «Projet»), cfr COULIE, *Lemmatisation* et le site web du projet à l'adresse <http://tpg.fltr.ucl.ac.be>.

<sup>7</sup> Sur la lemmatisation des langues classiques, cfr, par ex., BRUNET, *Qui lemmatise*; MELLET, *Lemmatisation*; MELLET – PURNELLE, *Atouts*; sur les traitements applicables aux données lexicales qui en sont issues, cfr, par ex., ÉVRARD – MELLET, *Méthodes quantitatives*; MELLET, *Réflexions*; MOREUX, *Méthodes quantitatives*.

Pour les concepteurs du *Projet*, la lemmatisation reproduit rigoureusement la dichotomie saussurienne établie entre *langue* et *discours*. Les lemmes sont les représentants des formes au niveau de la langue; ils sont des abstractions indispensables à la description. Les formes sont les actualisations des lemmes au niveau du discours; seules réalités tangibles des textes, elles font l'objet de la description. Rassemblés en une base de données relationnelle, les lemmes et les formes constituent donc respectivement la macrostructure et la microstructure<sup>8</sup> d'un dictionnaire électronique propre au *Projet*, le *Dictionnaire Automatique Grec (D.A.G.)*.

La lemmatisation est automatisée. En pratique, un automate compare les données lexicales du texte traité à celles du *D.A.G.* À l'issue de ce processus, une forme reconnue reçoit le lemme qui lui correspond, une forme faisant l'objet de plusieurs analyses reçoit plusieurs propositions de lemme. Aucun lemme n'est attribué à une forme non reconnue. Le lemmatiseur confirme, complète ou corrige les relations entre formes et lemmes établies par la machine. Il attribue ensuite un lemme aux formes non reconnues, soit en établissant un lien avec un lemme existant déjà dans le *D.A.G.*, soit en formulant un nouveau lemme. La multiplication des comparaisons nourrit le dictionnaire de données lexicales nouvelles. L'automatisation, loin de rendre inutile l'intervention des philologues, permet cependant d'accélérer sensiblement les analyses, d'assurer l'exhaustivité des dépouillements et la cohérence des classements et des regroupements. Le *D.A.G.* pourra ainsi offrir, *in fine*, l'image attendue du lexique.

Les hellénistes ont déjà à leur disposition de nombreux dictionnaires généraux ou spécialisés adaptés à tous les types de sources pour les différentes époques de l'histoire de la langue grecque<sup>9</sup>. Ce fait ne peut pourtant pas servir d'objection. N'étant pas fondés sur l'opposition entre langue et discours, ces outils traditionnels ne suffisent pas pour rencontrer, à eux seuls, les objectifs du *Projet*. Ils rendent compte des usages attestés dans l'ensemble des textes retenus, c'est-à-dire uniquement dans ce que leur auteur aura retenu comme *corpus*. Leur macrostructure présente dès lors à la fois trop d'éléments et pas assez:

1. Les dictionnaires traditionnels fournissent une description de toutes les variations formelles d'une forme rencontrées dans le *corpus*. La macrostructure devient donc pléthorique car elle retient comme entrée lexicale :

- non seulement les formes dialectales d'un même mot: le dictionnaire d'A. Bailly enregistre les entrées ἄλιος (B 79) et ἱάλιος (B 897-898), formes respectivement dorienne et ionio-attique du même mot signifiant «soleil» et que le *D.A.G.* regroupera sous la forme classique ionio-attique, ἱάλιος; le même dictionnaire prévoit deux entrées κοῦρις (B 1127), l'une adjectivale «qui tond», l'autre nominale «crabe». Cette dernière n'est pourtant qu'une forme dialectale dorienne du mot καρίς «crabe»<sup>10</sup>. Le *D.A.G.* ne retiendra donc qu'une entrée pour κοῦρις, l'adjectif, et fera figurer le substantif homographe sous καρίς;

<sup>8</sup> Au sens de REY-DEBOVE, *Étude*, p. 21, 114 et 148; GAUDIN – GUESPIN, *Initiation*, p. 110-112.

<sup>9</sup> Cfr par exemple la page relative au dictionnaire sur le site de la *Bibliotheca Classica Selecta* à l'adresse <http://bcs.fltr.ucl.ac.be/DicLanAGr.html#DicLan2>.

<sup>10</sup> Cfr LS 987; C 498, s.v. καρίς.

- mais aussi des formes fléchies: le même dictionnaire (B 1623) réserve pour le mot *πρίων* les entrées 1 *πρίων* pour «la scie», 2 *πρίων* pour «le scieur», 3 *πρίων* pour «le marchand» et enfin 4 *πρίων* pour «part. prés. de *πρίω* 1», la forme conjuguée d'un verbe;

- et des mots considérés, explicitement ou implicitement, comme fautifs: le dictionnaire de E.A. Sophocles présente une entrée *σπουδικόν* (S 1005) «corrupt for *δεσποτικόν*».

2. De plus, pour l'ensemble des textes retenus, le dépouillement n'est jamais systématique ni complet. De nombreux mots sont passés sous silence, même s'ils apparaissent dans le texte d'un auteur retenu pour la confection du dictionnaire. Ne décrivant que les usages rencontrés, ils évitent de plus toute généralisation, omettant sans scrupule tous les termes qui existent manifestement dans la langue mais qui ne s'actualisent pas dans leur *corpus* de référence forcément incomplet. Si *λυμαίνω* fait l'objet d'une entrée car il est présent dans les sources retenues, *περιλυμαίνω*, pourtant attesté dans d'autres sources, n'est ni dans le *Greek-English Lexicon* ni dans le *Dictionnaire Grec Français*; les dictionnaires présentent ici une première carence.

3. D'un dictionnaire à l'autre, des usages identiques répertoriés dans les mêmes sources sont représentés différemment: le *Dictionnaire Grec Français* et le *Greek-English Lexicon* fournissent deux représentations lexicographiques sensiblement différentes de l'entrée *ἀλέκτωρ*:

- B 74	1 ἀλέκτωρ [...] 1 (à priv.) «vierge»
	2 (à cop.) «éroux»
	2 ἀλέκτωρ [...] «coq» [...] (cfr ἀλέκω?)
- LS 62	ἀλέκτωρ (A) [...] I. «cock» II. «husband»
	ἀλέκτωρ (B) [...] (à priv., λέγω) = ἄλεκτρος

Aucun des deux dictionnaires ne justifie le regroupement du mot *ἀλέκτωρ* «éroux» avec *ἀλέκτωρ* «vierge» ou avec *ἀλέκτωρ* «coq»; deuxième type de carence.

Tous les dictionnaires traditionnels ont leur mérite. Ils ont aussi des limites bien connues<sup>11</sup>. Les dictionnaires de la langue grecque, issus d'une longue tradition lexicographique, n'échappent pas à la règle, même s'ils ont rendu de grands services à des générations d'hellénistes. Le recours à un dictionnaire électronique, le *D.A.G.*, rend possible la rationalisation de la nomenclature. Toutes les informations relatives aux mots, conservées dans la microstructure, restent accessibles par des interrogations pertinentes. De plus, à l'inverse des dictionnaires «papiers» traditionnels, la matière d'un dictionnaire électronique est susceptible d'être augmentée et révisée.

À ce jour, le *D.A.G.* comprend 174.758 formes différentes classées sous 33.874 lemmes. Les analyses portent sur un *corpus* de 4.284.380 occurrences, tous textes confondus (cfr tableau 1). La majorité des œuvres traitées relève de la patristique

<sup>11</sup> Cfr par ex. CORBIN, *Morphologie*, p. 22; REY-DEBOVE, *Étude*, p. 64.

(2.787.998 occurrences) et de l'historiographie byzantine (803.094 occurrences)<sup>12</sup>. La nature fortement classicisante de la langue de ces sources permet cependant d'appliquer le processus de comparaison à des textes d'époque classique, tout en conservant des taux raisonnables de reconnaissance des formes (cfr tableau 3, *infra*). Les concordances lemmatisées publiées dans le *Thesaurus Patrum Graecorum* (*T.P.G.*), une sous-série du *Corpus Christianorum* diffusée par Brepols Publishers, constituent sans doute la partie la plus visible du *Projet*. Mais ces réalisations constituent déjà autant de «produits finis», le centre du *Projet* étant le dictionnaire en lui-même.

Chantiers		Occurrences	Lemmes	Formes
12	ThAstFirm (Firmus) (IV <sup>e</sup> -V <sup>e</sup> s.)	4.181	1.130	1.894
9	ThPsNon, <i>Comm.</i> (VI <sup>e</sup> s.)	20.248	2.818	5.579
6	ThPsNonPan, <i>Paraph.</i> (V <sup>e</sup> -VI <sup>e</sup> s.)	23.391	3.389	7.356
7	ThConcOec ( <i>varia</i> )	27.011	2.986	6.638
18	ThMenPro (VI <sup>e</sup> s.)	29.990		
3	ThLeoCons (VI <sup>e</sup> s.)	30.819	3.679	8.329
5	ThAmlc (IV <sup>e</sup> s.)	38.740	4.158	9.989
12	ThAstFirm (Astérius) (IV <sup>e</sup> -V <sup>e</sup> s.)	48.174	5.479	13.799
17	ThAgMyr (VI <sup>e</sup> s.)	66.659		
20	ThTheoSim (VI <sup>e</sup> -VII <sup>e</sup> s.)	78.874		
8	ThTheoConf (VIII <sup>e</sup> -IX <sup>e</sup> s.)	131.948	8.692	22.448
2	ThGrNaz, II (IV <sup>e</sup> s.)	153.723	9.854	31.304
1	ThGrNaz, I (IV <sup>e</sup> s.)	268.327	10.506	41.617
10	ThProcCaes (VI <sup>e</sup> s.)	292.552	7.982	28.997
16	ThPhotCons (IX <sup>e</sup> s.)	347.182		
13-14	ThBasCaes (IV <sup>e</sup> s.)	707.853	14.843	68.867
22	ThAthAlex* (IV <sup>e</sup> s.)	760.000		
19	ThGrNys* (IV <sup>e</sup> s.)	807.000		
Autres		313.398		
Total général		4.284.380		
Total Cappadociens		2.027.998		
Total historiens**		803.094		

Tableau 1: nombre d'occurrences, de lemmes et de formes dans les différents chantiers réalisés ou en cours de traitement

Depuis leur origine, ces travaux sont le fruit d'une collaboration avec le «Centre de traitement électronique des documents» (CETEDOC), créé et dirigé à l'U.C.L. par le Professeur Paul Tombeur. À l'éméritat de ce dernier, le centre a été rebaptisé «Centre de

<sup>12</sup> Les numéros renvoient à la liste des titres parus et en préparation fournie en annexe. La rubrique *Autres* rassemble les données chiffrées relatives aux textes étudiés sans avoir été finalisés par un *Thesaurus* ou ayant fait l'objet d'une concordance publiée dans une autre collection que le *T.P.G.* Le nombre de lemmes et de formes n'est pas mentionné pour les chantiers en cours de traitement. L'astérisque (\*) signale des données provisoires; deux astérisques (\*\*) indiquent que, pour les historiens, les œuvres autres qu'historiographiques (lettres, traités, poèmes) ne sont pas prises en compte dans l'établissement du nombre d'occurrences.

traitement automatique du langage» (CENTAL) et placé sous la houlette du Professeur Cédric Fairon<sup>13</sup>.

Sous l'ancienne direction, l'équipe, accoutumée au traitement des textes latins, a transmis au *Projet* son expérience de la normalisation des formes et de leur étiquetage. Ces deux opérations sont donc automatisées, tout comme les dépouillements et les comptages. Les formes ambiguës sont ensuite traitées une à une par le lemmatiseur. Un codage et une police propres au Centre ont été mis au point pour traiter les textes en grec. Si le traitement de masse se fait sur des données transcodées, les interfaces mises à la disposition des utilisateurs présentent un affichage en grec. Les données finies sont diffusées dans les *Thesauri* en partie sur papier, en partie sur microfiches. Le *D.A.G.*, créé à l'instar du *Dictionnaire Automatique Latin (D.A.L.)*, est constitué de trois tables de données, celle des formes, celles des lemmes, et une dernière table fournissant le lien numérique de un à plusieurs entre les formes et les lemmes.

Actuellement, le CENTAL apporte au *Projet* son expertise en matière de traitement automatique des langues modernes. L'automatisation portera bientôt sur la levée des ambiguïtés lexicales<sup>14</sup>. Les données seront diffusées sur un support électronique, en l'occurrence le WEB. Les traitements informatiques s'appliquent à des données converties en UNICODE, codage qui permet un affichage en grec à toutes les étapes du travail. Les lemmes du *D.A.G.* ont été pourvus d'informations relatives à leur classe morpho-syntaxique et d'une analyse dérivationnelle. L'adjonction d'une analyse morphologique complète est à l'étude. Le dictionnaire évolue ainsi vers un dictionnaire du type DELAF, c'est-à-dire un dictionnaire mis au format des outils conçus au «Laboratoire d'automatique Documentaire et Linguistique» (L.A.D.L.)<sup>15</sup>. Les versions expérimentales de ces applications sont déjà opérationnelles.

## **2. La question de la norme. L'analyse des textes au service d'une description de la langue: réflexions et principes posés par les concepteurs du *Projet***

Le *Projet* tire toute sa matière de l'analyse des textes; il s'agit de répertorier et de classer toutes les réalités lexicales observées. En grec, les actualisations des lemmes peuvent revêtir un grand nombre de variations formelles; le *D.A.G.* fournit 186 formes différentes du verbe εἶμι «être», 238 formes différentes de γίγνομαι. Comme dans les autres langues, ces variations sont conditionnées par de nombreux facteurs, susceptibles de combiner leurs effets: la langue grecque repose sur un système flexionnel; elle possède une variété dialectale (certains dialectes, attachés à un genre littéraire spécifique, ont perduré dans l'histoire de la langue); elle présente des variations phonétiques. La lemmatisation des mots homographes et des mots rares soulève des questions auxquelles il a fallu apporter une réponse. Il ne s'agira pas d'effectuer un relevé complet des variations formelles rencontrées

---

<sup>13</sup> Une description du laboratoire et de ses activités est disponible sur la toile à l'adresse <http://cental.fltr.ucl.ac.be>.

<sup>14</sup> Cfr 3.1., *infra*.

<sup>15</sup> Cfr <http://infolingu.univ-mlv.fr>.

ni d'écrire une grammaire. Le propos se limitera à illustrer un fait: l'enjeu de la lemmatisation réside dans le partage qui sera établi entre les variations qui justifient, pour une série de formes, la création d'un nouveau lemme et celles qui ne le justifient pas. C'est la question de la norme qui détermine ce qui, dans la langue, apparaît comme systématisable et productif, et ce qui ne l'est pas.

### 2.1 Variations justifiant la création d'un lemme

Le lemme doit refléter le caractère systématisable – dans le sens de l'intégration au système de la langue – et productif d'une série de formes. Des mots apparentés mais relevant de paradigmes morphosyntaxiques différents justifient la création de deux lemmes distincts. Toutes les classes morpho-syntaxiques sont concernées:

- des substantifs: ἀντέρεισμα vs ἀντέρεισις, ὀλόλυγή vs ὀλόλυγμα vs ὀλόλυγμός;
- des anthroponymes: Τατζάτιος vs Τατζάτης, Ἄρτασήρας vs Ἄρτασήριος, Ἑκατόμνωσ vs Ἑκατόμνος;
- des surnoms: Τριφύλης vs Τριφύλιος;
- des ethniques: Πεπουζηνοί vs Πεπουζιανοί vs Πεπουζίται;
- des toponymes: Γράσση vs Χρᾶσις; Νισίβη vs Νίσιβις; Σαρδώ vs Σαρδανία vs Σάρδος; Τομέα vs Τόμις; Τύανος vs Τύανα;
- des adjectifs: ἄσωμος vs ἄσώματος; βλακώδης vs βλακικός; ἰσαίος vs ἴσος; Μοντανός vs Μοντανιστής; Εὐτυχηταί vs Εὐτυχιανός vs Εὐτυχιανιστής vs Εὐτυχιανικός;
- des adverbes: ἀστενάκτως vs ἀστενακτί;
- des verbes: ἀνταιχμαλωτίζω vs ἀνταιχμαλωτεύω; στερέω vs στερίσκω.

Deux lemmes s'imposent quand une forme déclinable s'oppose à une forme indéclinable: ἄββᾶς vs ἄββᾶ, πάππας vs πάππα, ῥαββίς vs ῥαββί (pour des substantifs); Δανίηλος vs Δανίηλ, Ἰσαμὸς vs Ἰσαμ, Οὔμαρος vs Οὔμαρ, Ζαδέσπρας vs Ζαδεσπράτης vs Ζαδεσπράμ (pour des anthroponymes); Ἀργάριζος vs Ἀργαριζίν (pour un toponyme).

La distinction de deux lemmes est également de mise quand l'évolution sémantique d'un mot est accompagnée d'un changement du paradigme morpho-syntaxique: ἄνθος (ὅ) «bruant» (oiseau) vs ἄνθος (τό) «ce qui croît, ce qui pousse, une pousse [...], fleur».

La création d'un lemme repose aussi sur son caractère productif dans la langue. Par mesure d'économie, la forme στομαχῶντες, attestée une seule fois, n'a pas été lemmatisée sous \*στομαχᾶω, inconnu des sources, mais sous στομαχέω. Par contre, les lemmes ἐκτρώσκω et σκεδαννύω, qui apparaissent comme productifs, côtoient ἐκτιτρώσκω et σκεδάννυμι dans la macrostructure du *D.A.G.* Dans de tels cas, le classement des formes et la graphie retenue pour les lemmes sont motivés dans des «index des lemmes particuliers» publiés dans les volumes de concordance du *T.P.G.*

### 2.2 Variations ne justifiant pas la création d'un lemme

On peut citer, par exemple:

- la contraction, amalgame de deux voyelles ou d'une voyelle et d'une diphtongue: les lemmes revêtent une forme dans laquelle les voyelles ne sont pas contractées. Des formes

ioniennes non contractées n'imposent pas un autre lemme: ἄδονται *s.l.* αἰίδω; φιλοῦντας *s.l.* φιλέω;

- l'assimilation, extension du timbre d'une consonne sur la consonne qui précède: les lemmes revêtent une forme dans laquelle les consonnes consécutives sont assimilées; la variation des graphies dans les sources ne suffit pas à poser des lemmes différents: ἐνεχείρισα *s.l.* ἐγχειρίζω; κατόψεται *s.l.* καθοράω;

- la psilose, chute de l'aspiration d'un mot: ἀνδανέεις *s.l.* ἀνδανάω. La psilose est fréquente en dialecte ionien, sans justifier pour autant un lemme propre pour les formes qui en sont marquées;

- l'aphérèse, chute d'éléments phoniques à l'initiale d'un mot: ὄστι (pour ἐστί) *s.l.* εἰμί; μαῦρον *s.l.* ἀμαυρός. En grec byzantin, la chute d'un sigma initial peut perturber sensiblement la reconnaissance de certaines formes: ἔνατον *s.l.* σενᾶτον; Ἐρουίου (pour Σερούιου) *s.l.* Σέρουιος, cette forme sans *sigma* initial n'est pas répertoriée dans les dictionnaires;

- la syncope, chute d'éléments phoniques à l'intérieur d'un mot: ἐμπριζομένου *s.l.* ἐμπυρίζω. Dans la langue grecque d'époque byzantine, la syncope affecte fréquemment les mots empruntés au latin. Le lemme revêt une forme pleine si cette dernière est attestée dans les sources: Κάτλος *s.l.* Κάτουλος; βάκλων *s.l.* βάκυλον; mais σταύλων *s.l.* στάβλον;

- l'élision, chute d'éléments phoniques en finale d'un mot: γράμματ' *s.l.* γράμμα;

- la métathèse, échange de la quantité de voyelles successives d'un mot: λεωφόρος *s.l.* λαοφόρος;

Les variations présentées peuvent combiner leurs effets. Pour une forme πιτύαν, le dictionnaire d'A. Bailly présente trois entrées:

B 1561 πίτυα, ας (ή) c. πυτία [...];

B 1699 πυετία, ας (ή) [...] (cfr πύος).

B 1708 πυτία, ας (ή) [...] (v. πυετία);

Il s'agit pourtant, d'un point de vue lexicologique, du même mot. Les formes du type πυτία étant issues de la contraction de πυετία, et πιτύα s'expliquant par métathèse vocalique<sup>16</sup>. Aucune de ces trois graphies n'ayant développé une série productive, le *D.A.G.* n'a qu'un seul lemme, πυετία.

### 2.3 La lemmatisation des mots homographes

De nombreuses formes grecques sont susceptibles de recevoir plusieurs analyses et donc de se voir attribuer plusieurs lemmes. La forme νέων est ainsi enregistrée dans le *D.A.G.* sous les cinq «adresses lexicales»: νέος «nouveau» (adjectif), νέω (νήχω) «nager» (verbe), νέω (νήμα) «filer» (verbe), νέω (νηέω) «entasser» (verbe), et Νέων «Néon» (anthroponyme, la forme νέων pouvant recevoir une majuscule en tête de phrase). Le

<sup>16</sup> Cfr C 956, s.v. πύος.

processus de comparaison des formes d'un texte aux données lexicales du *D.A.G.* ne peut renvoyer à trois lemmes νέω orthographiés de la même manière. Les concepteurs du *Projet* ont donc choisi de distinguer les cas d'homographie des lemmes par l'adjonction de spécifications. La formulation de ces dernières répond à une syntaxe qu'il serait trop long de décrire ici en détail. Le tableau 2 illustre cependant quelques oppositions établies entre des lemmes qui, sans les spécifications, seraient homographes.

ἀγάλακτος (χωρίς) «giving no milk»	vs	ἀγάλακτος (ἄμα) «όμογάλακτος»
ἀγάστωρ (χωρίς) «qui est à jeûn»	vs	ἀγάστωρ (ἄμα) «né du même sein»
ἀλέκτωρ (χωρίς) «vierge»	vs	ἀλέκτωρ (ἄμα) «ἔρουx» <sup>17</sup>
ἄλοχος (χωρίς) «qui n'a pas enfanté»	vs	ἄλοχος (ἄμα) «compagne de couche»
ἀνάξιος (ἀναξ) «à la manière d'un roi»	vs	ἀνάξιος (ἄξιος) «indigne»
ἀναξία (ἀναξ) «prouvoir»	vs	ἀναξία (ἄξιος) «indignité»
Ναυατιανός, ἡ, ὄν «novatien»	vs	Ναυατιανός (ὄ) «Novatianus»
Νεστοριανός, ἡ, ὄν «nestorien»	vs	Νεστοριανός (ὄ) «Nestorianus»
Πέρσης (ὄ) «Persès»	vs	Πέρσης (Πέρσις) «perse»
Σύρος (ὄ) «Syros»		Σύρος (Συρία) «syrien»
Περσικός (ὄ) «Persicos»	vs	Περσικός (Πέρσις) «perse»
βάτος (βάτου) «ronce»	vs	βάτος (μέτρον) «mesure de 50 setiers» <sup>18</sup>
κόρος (κοῦρος) «jeune garçon»	vs	κόρος (μέτρον) «mesure de 6 médimnes attiques» <sup>19</sup>

Tableau 2: distinction des lemmes homographes par les spécifications

## 2.4 La lemmatisation des mots rares

Le caractère productif et systématisable d'une forme peu attestée ou d'un *hapax* est plus difficile à déterminer car leurs attestations dans les sources font, par définition, défaut. La lemmatisation de ces formes se fonde alors sur le caractère prédictible des flexifs et sur la productivité des morphèmes constitutifs du mot. Le lemme ἀγαπητρίς est déduit de l'unique forme ἀγαπητρίδων sur base de la productivité du suffixe d'agent masculin en -τ(η)ρ- et du suffixe féminin en -ιδ-, comme αὐλητήρ / αὐλητρίς. Le lemme συγκαταδέχομαι est déduit de l'unique forme συγκαταδέχεται sur base de la prédictibilité du flexif (-ε-ται → -ο-μαι), ainsi que sur la productivité du v. δέχομαι tant en dérivation qu'en composition et celle de la double préfixation en συγ-κατα-. Les lemmes pour lesquels une forme de départ au nominatif (pour les noms et les adjectifs) ou à la première personne du singulier (pour les verbes) n'est pas attestée dans les sources sont énumérés dans les «index des lemmes particuliers» déjà cités ci-dessus.

## 2.5 Conclusion

Dans le cadre du *Projet*, la lemmatisation est toujours un exercice au niveau de la langue. La graphie d'un lemme n'est pas forcément celle qui apparaît comme la plus

<sup>17</sup> Il faut y joindre ἀλέκτωρ (ἀλέξω) «coq». B 74 et LS 62 dégrouper autrement ces trois termes, cfr *supra*, p. 15.

<sup>18</sup> Il faut y joindre βάτος (ἰχθύς) «sorte de raie».

<sup>19</sup> Il faut y joindre κόρος (κορέννυμι) «satiété» vs κόρος (κορέω) «ordure».

fréquente dans les textes d'un auteur dont le lexique est en cours d'analyse, mais celle qui est la plus productive dans l'ensemble des sources. De la même manière, les concordances lemmatisées publiées font état des lemmes munis des spécifications qui les accompagnent dans le *D.A.G.*, même si un seul des différents lexèmes qu'elles permettent de distinguer s'actualise de fait dans le *corpus* étudié. Enfin, la rareté d'un terme et, éventuellement, sa qualité d'*hapax*, ne se définissent pas par le petit nombre d'occurrences de ce mot dans un *corpus* traité, mais sont déterminés à la suite d'un examen approfondi des sources. Les dictionnaires «papiers» traditionnels et les bases de données textuelles retrouvent ici toute leur utilité. Pour les philologues, ces outils fournissent autant d'accès aux sources disponibles, délimitant ainsi «l'ensemble lexical de référence»<sup>20</sup>, c'est-à-dire toutes les manifestations accessibles de la langue, de toutes natures, sur tous supports et de toutes les époques. La lemmatisation de textes historiographiques byzantins ne dispense pas d'avoir à consulter des textes épigraphiques d'époque classique. Les formes ἐπιδημιουργοί et ἐπιδημιουργῶν ne sont attestées qu'au pluriel dans les sources et dans les dictionnaires. Une inscription dorienne atteste cependant la forme ἐπιδαμιουργόν justifiant, au niveau de la langue, la création d'un lemme singulier ἐπιδημιουργός. Le nominatif du toponyme Ἄνισα, attesté dans les sources littéraires sous les formes Ἄνίσοις ou Ἄνιήσοις, apparaît sur une monnaie.

### 3. Perspectives. Une description de la langue au service de l'analyse des textes: les aspirations des concepteurs du *Projet*

Au fil des analyses successives, progressivement et régulièrement, le *D.A.G.* s'accroît. L'exploration de nouveaux textes enrichit le dictionnaire de formes nouvelles et de lemmes nouveaux. Ces derniers sont définis selon une norme stable mais susceptible d'être perfectionnée ou corrigée selon les réalités lexicales rencontrées. L'ensemble des données répond donc chaque jour davantage aux aspirations attendues d'exhaustivité et de cohérence. Les philologues du *Projet* s'efforcent aussi à diversifier la nature des informations qui accompagnent les lemmes et les formes. Ces progrès quantitatifs et qualitatifs servent directement la description du lexique et ouvrent des perspectives nouvelles pour l'analyse des textes. La levée automatisée des ambiguïtés lexicales et l'application de traitements lexicométriques en font partie. La première est partiellement acquise, la seconde relève encore du rêve.

#### 3.1 L'ambiguïté lexicale

Telle qu'elle est actuellement conçue, la lemmatisation procède forme à forme sans tenir compte du contexte. À l'issue du processus de comparaison des formes d'un texte aux données lexicales d'un dictionnaire de référence, en l'occurrence le *D.A.G.*, le mot ἀγαθοῖς reçoit donc deux attributions de lemme: ἀγαθός, l'adjectif «beau», et ἀγαθῶ, le verbe «faire du bien à...». La forme ἀγαθοῖς constitue un cas d'ambiguïté lexicale analogue à celui du mot «marche» en français qui peut être un nom ou un verbe. Comme le montre le

---

<sup>20</sup> Pour reprendre l'expression de FAMERIE, *Appien*, p. 255-259.

tableau 3, la proportion du vocabulaire recevant plus d'une proposition de lemme dans les chantiers achevés ou en cours atteint les 10%, voire davantage. La levée manuelle des ambiguïtés réclame donc, de la part des philologues, des efforts considérables.

Chantiers	Occurrences	Occurrences reconnues	%	Occ. avec prop. de lemme >1	%
ThAstFirm	48.174	44.618	92,62	5.147	10,68
ThAstFirm	4.181	3.807	91,05	402	9,61
ThBasCaes	707.853	664.353	93,85	83.351	11,77
ThPhotCons	347.182	321.147	92,50	38.465	11,07
ThTheoSi	78.791	71.376	90,58	7.972	10,11
ThAgMyr	66.659	60.376	90,57	6.303	10,43
ThMenProt	29.990	27.624	92,11	3.392	12,27

Tableau 3: *proportion des occurrences reconnues et des occurrences ayant reçu plus d'une proposition de lemme*

L'encodage de la classe morpho-syntaxique des lemmes a permis de développer des grammaires locales qui contribuent à lever automatiquement les ambiguïtés à l'aide d'un examen préalable du contexte. Une de ces grammaires informe l'automate de lemmatisation que la séquence <préposition> + <adjectif> + <nom> actualisée dans une phrase doit être retenue contre d'autres séquences possibles si l'on s'en tenait uniquement à l'analyse des formes (par exemple <préposition> + <verbe> + <nom>). La forme ἀγαθός, pourrait apparaître en deuxième position dans chacune des deux séquences citées. Mais en vertu de la règle constitutive de la grammaire utilisée, elle doit être comprise comme un adjectif, et non comme un verbe. La forme ne reçoit donc qu'une seule étiquette, celle du lemme adjectival ἀγαθός.

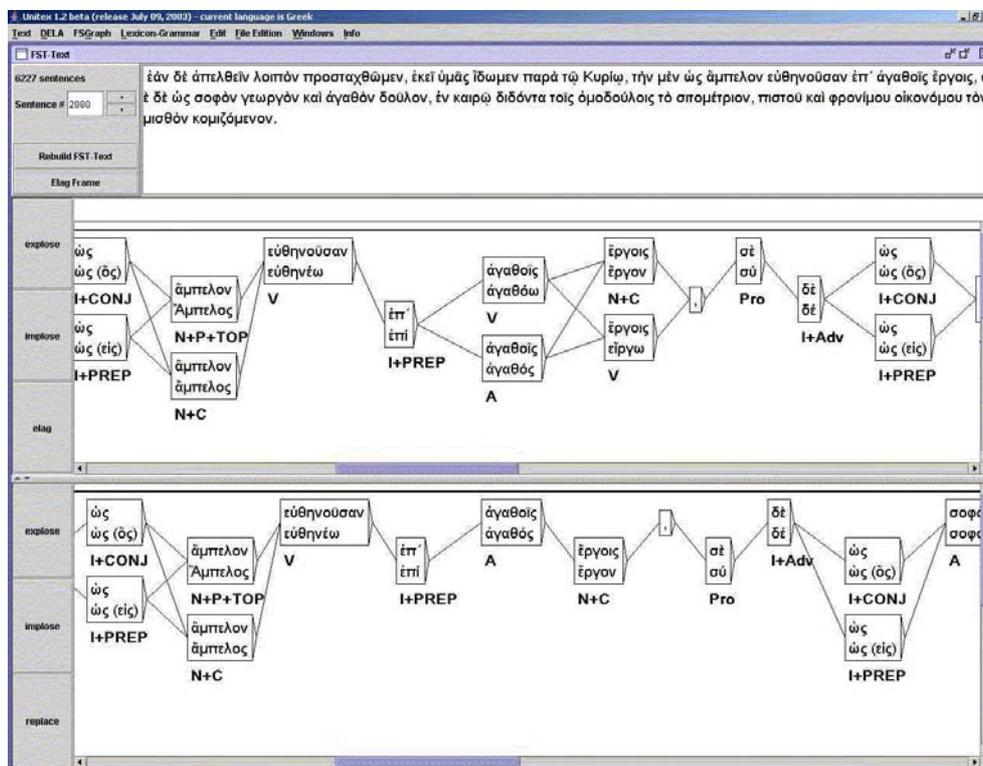


Figure 1: une phrase et les graphes qui lui correspondent avant et après l'application d'une grammaire locale

Une telle opération est possible grâce à l'adaptation du logiciel Unitex aux textes grecs<sup>21</sup>. Comme le montre la figure 1, ce logiciel est capable de construire le graphe d'une phrase. Chaque mot y est représenté dans un cadre appelé «une boîte». Les liens entre les boîtes symbolisent la linéarité de la phrase. Le contenu des boîtes fait état de la forme du mot *in textu*, de son lemme et de son analyse morpho-syntaxique. Si une forme fait l'objet de plusieurs propositions de lemmes, plusieurs boîtes se superposent dans une même étape du graphe. Trois registres composent la figure: le registre supérieur offre le texte grec; le deuxième registre présente le graphe de la phrase avant l'application de la grammaire locale, il y a donc deux boîtes pour ἀγαθοῖς; le registre inférieur fournit le graphe après l'application de la grammaire, les chemins non pertinents ont été effacés. À ce stade du développement du logiciel, l'expérience ne prend en compte que des analyses morpho-syntaxiques. Dans un futur proche, il sera nécessaire de multiplier non seulement les grammaires mais aussi les informations sur lesquelles les grammaires pourront fonder leurs analyses afin de résoudre le plus grand nombre possible d'ambiguïtés. Ce nouveau développement profite directement des travaux antérieurs car les textes déjà lemmatisés sont utilisés comme *corpus* d'entraînement.

<sup>21</sup> Sur Unitex et les grammaires locales, cfr le site du L.A.D.L. cité note 15.

### 3.2 Dépouillements et comptages: vers les traitements lexicométriques

Les dix noms les plus fréquents dans deux *corpus* historiographiques (Procopé de Césarée, VI<sup>e</sup> s. ap. J.-C. et Théophane le Confesseur, IX<sup>e</sup> s. ap. J.-C.) et deux *corpus* patristiques (Grégoire de Nazianze et Basile de Césarée, IV<sup>e</sup> s. ap. J.-C.) ont été sélectionnés. De ces deux séries de dix mots, ont été extraits les trois noms les plus fréquents que les deux historiens présentent en commun et les trois noms les plus fréquents que les pères de l'Église présentent en commun (tableau 4). Pour les textes historiographiques, on relève alors les mots Ῥωμαῖος «romain», βασιλεύς «le roi, l'empereur» et Πέρσης «perse». Les trois noms les plus fréquents communs aux évêques cappadociens sont θεός «dieu», λόγος «parole» et ἄνθρωπος «homme». Des thématiques communes sont donc mises en évidence à l'intérieur du genre littéraire auquel appartiennent les œuvres examinées. Le propos des historiens n'est pas le même que celui des théologiens.

Textes historiographiques				Textes théologiques			
Procopé (292.552 occ.)		Théophane (131.948 occ.)		Grégoire de Nazianze (268.327 occ.)		Basile de Césaré (707.853 occ.)	
Ῥωμαῖος	1.871	βασιλεύς	1.452	θεός	1.782	θεός	5.662
βασιλεύς	1.601	Ῥωμαῖος	547	λόγος	1.590	ἄνθρωπος	2.697
Πέρσης	676	Πέρσης	358	ἄνθρωπος	498	λόγος	1.701

Tableau 4: mots communs dans des corpus de même thématique

Un tel exercice n'est faisable que sur des données lemmatisées et entièrement étiquetées car certaines formes de ces lemmes sont susceptibles de recevoir différentes analyses: θεοῦ peut être une forme nominale, ce qui est le cas ici, ou une forme du verbe θεῶω; Πέρσης peut être un adjectif, comme ici, mais aussi un anthroponyme. Ces données sont tirées des microfiches publiées dans les *Thesauri* des auteurs évoqués. Elles demeurent donc statiques. Les utilisateurs ne peuvent que constater la présence ou l'absence d'une information recherchée dans un *corpus*. Il ne peut en aucun cas les manipuler, à moins de les encoder à nouveau dans un environnement qui lui serait propre.

*A contrario*, quand les données seront diffusées sur support informatique dans un format compatible avec les applications propres à des traitements statistiques, il sera possible de mesurer de nombreuses réalités qu'on ne peut atteindre sur les microfiches. Les applications du *Projet* s'ouvriront alors au calcul des univers sémantiques relatifs à certains mots, à la mesure des spécificités ou des diversités lexicales de parties choisies du *corpus*, ou à l'examen des connexions lexicales entre différents textes. Bref, un produit permettant d'aller beaucoup plus loin dans l'analyse des textes que ne le font les concordances actuelles, les listes statiques de mots, et les tables fréquentielles.

## 4. Conclusion générale

Dans un avenir proche, les différents développements décrits dans ses lignes seront mis en œuvre sur le *corpus* des *Lettres* de Basile de Césarée. Disponible sur la toile électronique, il pourra faire l'objet d'interrogations basées sur les formes, les lemmes et les analyses qui sont attachées à ces derniers. Les résultats s'afficheront en grec, sous forme de concordance. Cet essai constituera la première étape d'une transhumance qui mène de la microfiche au web. Parallèlement, l'exploration des textes se poursuivra afin d'enrichir le

*D.A.G.* et d'affiner les principes de lemmatisation. Car en effet, «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»<sup>22</sup>.

---

<sup>22</sup> Extrait de COULIE, *Thesaurus Patrum Graecorum*, p. 172.

## Références

- B = A. BAILLY, *Dictionnaire Grec Français*, rédigé avec le concours de E. EGGER, éd. revue et corrigée par L. SECHAN et P. CHANTRAINE, 26e éd., Paris, 1963 (réimpr. 2000).
- BROWNING, *Greek* = R. BROWNING, *Medieval and Modern Greek*, 2e éd., Cambridge, 1983<sup>2</sup> (réimpr. 1999).
- BRUNET, *Qui lemmatise* = E. BRUNET, *Qui lemmatise dilemme attise*, dans *Lexicometrica*, 2 (2000), p. 1-19 (= <http://www.cavi.univ-paris3.fr/lexicometrica/article/numero2.htm>).
- C = P. CHANTRAINE, *Dictionnaire étymologique de la langue grecque. Histoire des mots*, Paris, 1968-1984.
- CORBIN, *Morphologie* = D. CORBIN, *Morphologie dérivationnelle et structuration du lexique (Linguistische Arbeiten, 193)*, 2 vol., Tübingen, 1987 (réimpr. [Sens et Structure], Lille, 1991).
- COULIE, *Lemmatisation* = B. COULIE, *La lemmatisation des textes grecs et byzantins: une approche particulière de la langue et des auteurs*, dans *Byzantion*, 66 (1996), p. 35-54.
- COULIE, *Thesaurus Patrum Graecorum* = B. COULIE, *Corpus Christianorum. Thesaurus Patrum Graecorum*, dans *Corpus Christianorum 1953-2003. Xenium Natalicium. Fifty Years of Scholarly Editing*, Turnhout, 2003, p. 169-172.
- DUBUISSON – KINDT, *Les projets* = M. DUBUISSON, B. KINDT, *Les projets informatiques à l'Institut orientaliste* dans *Journée de l'Informatique. FLTR*. 15 septembre 2000, p. 1-3 (cfr <http://Zeus.fltr.ucl.ac.be/journées>)
- ÉVRARD – MELLET, *Méthodes quantitatives* = É. ÉVRARD, S. MELLET, *Les méthodes quantitatives en langues anciennes*, dans *Lalies*, 18 (1998), p. 109-155.
- FAMERIE, *Appien* = É. FAMERIE, *Le latin et le grec d'Appien. Contribution à l'étude du lexique d'un historien grec à Rome (École Pratique des Hautes Études, IV<sup>e</sup> section : Sciences historiques et philologiques, III. Hautes Études du monde Gréco-Romain, 24)*, Genève, 1998.
- GAUDIN – GUESPIN, *Initiation* = F. GAUDIN, L. GUESPIN, *Initiation à la lexicologie française. De la néologie aux dictionnaires (Champs linguistiques. Manuels)*, Bruxelles, 2000.
- GÉRARD – KINDT, *Caractérisation* = R. GÉRARD, B. KINDT, *Un nouveau développement du Projet de recherche en lexicologie grecque à l'Institut orientaliste: la caractérisation morphologique des lemmes du Dictionnaire Automatique Grec*, dans *Bulletin de liaison du Département d'Études grecques, latines et orientales*, 19, avril-mai 2002, p. 3-5 (cfr <http://www.fltr.ucl.ac.be/FLTR/Lexico>).
- GÉRARD – KINDT, *Développement* = R. GÉRARD, B. KINDT, *Un nouveau développement du Projet de recherche en lexicologie grecque à l'Institut orientaliste: la caractérisation morphologique des lemmes du Dictionnaire Automatique Grec*, dans *Bulletin de liaison du Département d'Études grecques, latines et orientales*, 19, avril-mai 2002, p. 3-5 (cfr <http://www.fltr.ucl.ac.be/FLTR/Lexico>).
- GÉRARD – KINDT, *Projet* = R. GÉRARD, B. KINDT, *Le Projet de recherche en lexicologie grecque à l'Institut orientaliste de l'Université catholique de Louvain. La collection du Thesaurus Patrum Graecorum et le Dictionnaire Automatique Grec*, dans *Byzantion*, 73 (2003) [à paraître].
- GÉRARD – KINDT, *Recherches* = R. GÉRARD, B. KINDT, *Recherches en lexicologie grecque à l'Institut orientaliste. La caractérisation morphologique des lemmes du Dictionnaire Automatique Grec (D.A.G.)*, dans *2<sup>e</sup> journée de l'Informatique. FLTR*. 16 avril 2002, p. 1-6 (cfr <http://Zeus.fltr.ucl.ac.be/journées>).

- KINDT, *Lemmatisation* = B. KINDT, *La lemmatisation automatisée au service d'une description lexicale du grec ancien. Propositions pour la formulation des lemmes du Dictionnaire Automatique Grec (D.A.G.). Rapport de recherche présenté en vue de l'obtention du D.E.A. en Philosophie et Lettres (ISLE 3DA), orientation «Philologie et histoire orientales»*, Louvain-la-Neuve, 2003.
- KINDT, *Recherches* = B. KINDT, *Recherches de lexicographie grecque à l'Institut orientaliste*, dans *Bulletin de liaison du Département d'Études grecques, latines et orientales*, 8, septembre 1999, p. 12-13 (cfr <http://www.fltr.ucl.ac.be/FLTR/Lexico>).
- LS = H.G. LIDDELL, R. SCOTT, H.S. JONES, *A Greek-English Lexicon*, 9<sup>e</sup> éd, Oxford, 1940 (réimpr. 1977, avec *Greek-English Lexicon. A Supplement*, ed. E.A. BARBER, Oxford, 1968), à compléter par *Greek-English Lexicon. Revised Supplement*, ed. G.W. GLARE, Oxford, 1996.
- MELLET – PURNELLE, *Atouts* = S. MELLET, G. PURNELLE, *Les atouts multiples de la lemmatisation: l'exemple du latin*, dans *JADT 2002, 6<sup>èmes</sup> journées internationales d'analyse statistique des données textuelles, Saint-Malo, 13-15 mars 2002*, 2 vol., Rennes, 2002 (= <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002>).
- MELLET, *Lemmatisation* = S. MELLET, *Lemmatisation et encodage grammatical: un luxe inutile?*, dans *Lexicometrica*, 3 (2001), p. 1-5 (= <http://www.cavi.univ-paris3.fr/lexicometrica/article/numero3.htm>).
- MELLET, *Réflexions* = S. MELLET, *Quelques réflexions sur l'exploitation statistique des données informatisées*, dans *Les Études Classiques*, 58 (1990), p. 105-113.
- MOREUX, *Méthodes quantitatives* = B. MOREUX, *L'utilisation des méthodes quantitatives en linguistique grecque et latine*, dans *L'Antiquité Classique*, 51 (1982), p. 291-338.
- REY-DEBOVE, *Étude* = J. REY-DEBOVE, *Étude linguistique et sémiotique des dictionnaires français contemporains (Approaches to Semiotics, 13)*, La Haye, Paris, 1971.
- S = E.A. SOPHOCLES, *Greek Lexicon of the Roman and Byzantine Periods (From B.C. 146 to A.D. 1100)*, Cambridge (Massachusetts), 1887 (réimpr. New York, 1957).

**Annexe: Thesaurus Patrum Graecorum**  
**Liste des titres parus et en préparation**

**Titres parus**

1. *Thesaurus Sancti Gregorii Nazianzeni, Orationes, Epistulae, Testamentum* – J. MOSSAY et CETEDOC, 1990.
2. *Thesaurus Sancti Gregorii Nazianzeni, Carmina, Christus Patiens, Vita* – J. MOSSAY, B. COULIE et CETEDOC, 1991.
3. *Thesaurus Leontii Presbyteri Constantinopolitani* – J. NORET et CETEDOC, 1992.
4. *Thesaurus Pseudo-Dionysii Areopagitae, Textus Graecus cum Translationibus Latinis* – M. NASTA et CETEDOC, 1993.
5. *Thesaurus Amphilochei Iconiensis* – B. COULIE et CETEDOC, 1994.
6. *Thesaurus Pseudo-Nonni quondam Panopolitani, Paraphrasis Evangelii S. Ioannis* – B. COULIE, L.F. SHERRY et CETEDOC, 1995.
7. *Thesaurus Conciliorum Oecumenicorum* – B. COULIE et CETEDOC, 1998.
8. *Thesaurus Theophanis Confessoris, Chronographia* – B. COULIE, P. YANNOPOULOS et CETEDOC, 1998.
9. *Thesaurus Pseudo-Nonni, Commentarii in IV Orationes Gregorii Nazianzeni* – B. COULIE, J. NIMMO SMITH et CETEDOC, 1999.
10. *Thesaurus Procopii Caesariensis, De Bellis, Historia Arcana, De Aedificiis* – B. COULIE, B. KINDT et CETEDOC, 2000.
11. *Thesaurus Theophanis Confessoris, Chronographia. Index Nominum* – B. COULIE, P. YANNOPOULOS, B. KINDT et CETEDOC, 2000.
12. *Thesaurus Asterii Amaseni et Firmi Caesariensis, Opera Omnia* – B. COULIE, B. KINDT et CETEDOC, 2001.
13. *Thesaurus Basilii Caesariensis, Opera Omnia. Pars I. Introductio. Enumeratio Lemmatum et Formarum A-I* – B. COULIE, B. KINDT et CETEDOC, 2002.
14. *Thesaurus Basilii Caesariensis, Opera Omnia. Pars II. Enumeratio Lemmatum et Formarum K-Ω* – B. COULIE, B. KINDT et CETEDOC, 2002.

**Titres en préparation**

15. *Thesaurus Sententiarum Sexti, Textus Auctus una cum Sententiis Clitarchi, Sententiis Pythagoricorum et Translatio Latina Rufini Aquileiensis* – B. COULIE, M. DUBUISSON.
16. *Thesaurus Photii Constantinopolitani, Bibliotheca* – B. KINDT, J. SCHAMP.
17. *Thesaurus Agathiae Myrinaei* – B. COULIE.
18. *Thesaurus Menandri Protectoris* – B. COULIE.
19. *Thesaurus Gregorii Nysseni, Opera Omnia* – B. COULIE, B. KINDT.
20. *Thesaurus Theophylacti Simocattae* – A. DE SIENA.
21. *Thesaurus Apocryphorum Veteris Testamenti* – J.-C. HAELEWYCK, B. KINDT.
22. *Thesaurus Athanasii Alexandrini, Opera Omnia* – B. COULIE, P. VAN DEUN.

**Sites WEB relatifs au Projet**

Site du *Projet T.P.G.-D.A.G.*: <http://tpg.fltr.ucl.ac.be>

Site du Centre d'étude sur Grégoire de Nazianze: <http://nazianzos.fltr.ucl.ac.be>

Site du CENTAL: <http://cental.fltr.ucl.ac.be>

Site de Brepols Publishers: <http://www.Brepols.net>

Site du Corpus Christianorum: <http://www.corpuschristianorum.org/index2.html>