

**La lemmatisation des sources  
patristiques et byzantines  
au service d'une description lexicale  
du grec ancien**

**Les principes de formulation des lemmes  
du *Dictionnaire Automatique Grec (D.A.G.)***

Bastien Kindt

Extrait de/Abstract from *Byzantion*, 74 (2004), p. 213-272  
(Deuxième édition revue/Second revised edition)



## AVANT-PROPOS

[English version below](#)

Les pages qui suivent constituent une seconde édition, au format ADOBE PDF, de notre article publié en 2004 :

B. KINDT, *La lemmatisation des sources patristiques et byzantines au service d'une description lexicale du grec ancien. Les principes de formulation des lemmes du Dictionnaire Automatique Grec (D.A.G.)*, dans *Byzantion*, 74 (2004), p. 213-272.

Cette contribution sert de base aux travaux réalisés dans le cadre du projet GREgORI (anciennement *Projet de Recherche en Lexicologie Grecque*), mené à l'Institut orientaliste de l'Université catholique de Louvain (cfr <http://www.uclouvain.be/gregori-project>).

Sa mise en ligne la rend directement accessible aux collaborateurs du projet.

Les modifications apportées au texte initial sont colorées en bleu. La pagination originale est reportée en rouge, entre crochets. Les liens Internet cités ont été transformés en liens hypertextes. Des liens hypertextes ont aussi été ajoutés à certains éléments de la bibliographie. Tous ces liens sont actifs au moment de la rédaction de cet avant-propos. L'usage des signets facilitera la navigation dans le document PDF.

Nous tenons à exprimer notre gratitude envers le Professeur Jacques Schamp, actuel Directeur de la revue *Byzantion*, qui nous a autorisé à produire et à diffuser la seconde édition de cet article.

*Pour citer cet article*, veuillez utiliser la référence bibliographique traditionnelle et y joindre l'URL du projet GREgORI.



Bastien KINDT  
Octobre 2015

## FOREWORD

The following pages are a second edition in ADOBE PDF format of our paper published in 2004 :

B. KINDT, *La lemmatisation des sources patristiques et byzantines au service d'une description lexicale du grec ancien. Les principes de formulation des lemmes du Dictionnaire Automatique Grec (D.A.G.), dans Byzantion*, 74 (2004), p. 213-272.

This article is still being used as a basis for works realized in the frame of the GREgORI project (formerly *Projet de Recherche en Lexicologie Grecque*) at the Institut orientaliste of the Université catholique de Louvain (Louvain-la-Neuve, Belgium ; see <http://www.uclouvain.be/gregori-project>).

This online version makes it directly accessible to the collaborators of the project.

Changes to the original text have been colored in blue. The original pagination, in red, has been put between brackets. Internet links cited have been turned into hyperlinks. Some bibliographic items have also been provided hyperlinks. All these links are active at the time of writing this foreword. The use of bookmarks makes the navigation within the PDF document easier.

We would like to express our gratitude to Professor Jacques Schamp, current Director of the *Byzantion* journal, who authorized us to publish and disseminate this second edition of our paper.

*How to quote* : please use the traditional bibliographic citation adding the URL of GREgORI project.



Bastien KINDT  
October 2015

## TABLE DES MATIÈRES

LES PRINCIPES DE FORMULATION DES LEMMES DU DICTIONNAIRE  
AUTOMATIQUE GREC (*D.A.G.*)

BIBLIOGRAPHIE

LISTE DES ABRÉVIATIONS

I. DÉFINITION DU PROJET (§1-22)

1. La lemmatisation en grec ancien
2. Le *D.A.G.*
3. L'enquête lexicographique
4. Les principes *d'économie* et de *fidélité*

II. PRINCIPES FORMELS DE LEMMATISATION (§23-33)

1. Variations formelles justifiant la création d'un lemme
2. Variations formelles ne justifiant pas la création d'un lemme

III. PRINCIPES MORPHO-SYNTAXIQUES ET SÉMANTIQUES :  
DISTINCTION DES LEMMES HOMOGRAPHERS (§34-44)

IV. PRINCIPES DE LEMMATISATION LIÉS À LA CLASSE MORPHO-  
SYNTAXIQUE DES MOTS (§45-89)

1. Les substantifs
2. Les adjectifs
3. Les numéraux
4. Les adverbes
5. Les verbes
6. Les pronoms

V. PRINCIPES LIÉS AU STATUT LEXICAL DES LEMMES (§90-111)

1. Les formes rares
2. Les formes présentant plusieurs graphies et les emprunts
3. Les lemmes défectifs et les formes supplétives
4. Les formes et locutions lexicalisées
5. Les fautes d'édition
6. Les lettres, les phonèmes et les morphèmes

CONCLUSIONS

# LA LEMMATISATION DES SOURCES PATRISTIQUES ET BYZANTINES AU SERVICE D'UNE DESCRIPTION LEXICALE DU GREC ANCIEN

## LES PRINCIPES DE FORMULATION DES LEMMES DU DICTIONNAIRE AUTOMATIQUE GREC (D.A.G.)

[Page 213]

Le *Projet de recherche en lexicologie grecque* (= *Projet*) mené depuis 1991 à l'Institut orientaliste de l'Université catholique de Louvain (Louvain-la-Neuve, Belgique), aborde l'étude du grec ancien par le biais de l'analyse de son vocabulaire, tel qu'il se manifeste dans les textes conservés et édités<sup>1</sup>. Son objectif est de fournir les bases d'un dictionnaire électronique qui, compte tenu des particularités de la langue grecque et de la nature des sources, offrirait du lexique une description exhaustive (assurant un dénombrement complet des réalités lexicales observées) et homogène (assurant un traitement identique des mêmes réalités) ; il s'agit en fait de procurer aux hellénistes un «lexique lexicographique» le plus proche possible d'un «lexique linguistique»<sup>2</sup>. La méthode suivie pour réaliser cet objectif s'articule autour de deux axes indissociables :

[Page 214]

- la *lemmatisation* des sources – les textes patristiques et byzantins – selon des principes explicites émanant d'une conception de la

<sup>1</sup> Son intitulé complet est «Projet de recherche en lexicologie grecque: élaboration de concordances lemmatisées des auteurs grecs patristiques et byzantins» (direction: Prof. B. Coulie) ; cfr COULIE, *Lemmatisation* ; COULIE, *Thesaurus Patrum Graecorum* et KINDT, *Avancées* (cfr les *Abréviations bibliographiques*), ainsi que le site web à l'adresse <http://tpg.fltr.ucl.ac.be>. Ces travaux sont réalisés en collaboration avec le CENTAL, ex-CETEDOC (*Centre de Traitement Automatique du Langage* ; direction: Prof. C. Fairon ; cfr <http://cental.fltr.ucl.ac.be>).

<sup>2</sup> Pour reprendre des termes utilisés par D. Corbin dans CORBIN, *Morphologie*, p. 44.

langue (travaux qui se concrétisent par la publication de concordances lemmatisées)<sup>3</sup> ;

- la constitution progressive d'un *dictionnaire électronique*, le *Dictionnaire Automatique Grec (D.A.G.)*, rassemblant les données lexicales tirées des textes analysés.

La présente contribution illustre ces deux axes en redéfinissant la nature du *Projet* et en formulant les principes de lemmatisation propres au *D.A.G.*<sup>4</sup>. Elle donne donc suite à l'article du Prof. B. Coulie publié dans cette revue en 1996<sup>5</sup> et offre une synthèse et une mise à jour des principes formulés antérieurement mais dispersés dans les introductions des volumes du *T.P.G.* La *Bibliographie* et la *Liste des abréviations* précèdent les développements articulés en cinq chapitres : la *Définition du Projet* (I, §1-22) et les *Principes formels de lemmatisation* (II, §23-33) en fournissent les postulats de base et les règles générales ; les *Principes morpho-syntaxiques et sémantiques* (III, §34-44) montrent comment sont distingués les lemmes homographes ; les deux derniers chapitres abordent les cas particuliers liés à la classe morpho-syntaxique (IV, §45-89) et au statut lexical (V, §90-111) des mots.

## BIBLIOGRAPHIE

A = F.R. ADRADOS, *Diccionario Griego-Español*, Madrid, depuis 1980.

APOTHÉLOZ, *Construction* = D. APOTHÉLOZ, *La construction du lexique français. Principes de morphologie dérivationnelle (L'essentiel français)*, Paris, 2002.

B = A. BAILLY, *Dictionnaire Grec Français*, rédigé avec le concours de E. EGGER, éd. revue et corrigée par L. SECHAN et P. CHANTRAINE, 26<sup>e</sup> éd., Paris, 1963 (réimpr. 2000).

BODSON, *Index* = L. BODSON, *Aristote, De Partibus Animalium. Index Verborum. Listes de fréquence (Travaux publiés par le Centre Informatique de Philosophie et Lettres. Laboratoire d'Analyse Statistique des Langues Anciennes, 17)*, Liège, 1990.

<sup>3</sup> Ces concordances sont publiées dans le *Thesaurus Patrum Graecorum (T.P.G.)*, une sous-collection du *Corpus Christianorum* diffusée par Brepols Publishers (<http://www.brepols.net> ; <http://www.corpuschristianorum.org>).

<sup>4</sup> Cette contribution a fait l'objet d'un travail de D.E.A. présenté à l'Institut orientaliste de l'Université catholique de Louvain le 24 juin 2003. L'auteur tient à renouveler l'expression de sa gratitude envers tous les collègues et professeurs de la Faculté de Philosophie et Lettres, de l'Institut orientaliste et du CENTAL qui l'ont accompagné pendant la réalisation de ce travail.

<sup>5</sup> COULIE, *Lemmatisation*.

BROWNING, *Greek* = R. BROWNING, *Medieval and Modern Greek*, 2<sup>e</sup> ed., Cambridge, 1983 (réimpr. 1999).

[Page 215]

BRUNET, *Qui lemmatise*<sup>6</sup> = É. BRUNET, *Qui lemmatise dilemme attise*, dans *Lexicometrica*, 2 (2000), p. 1-19.

C = P. CHANTRAINE, *Dictionnaire étymologique de la langue grecque. Histoire des mots*, Paris, 1968-1984.

CHANTRAINE, *Formation* = P. CHANTRAINE, *La formation des noms en grec ancien* (*Société de linguistique de Paris. Collection linguistique de Paris*, 38), Paris, 1933 (réimpr. 1979).

CHANTRAINE, *Morphologie* = P. CHANTRAINE, *Morphologie historique du grec* (*Nouvelle collection à l'usage des classes*, 34), 2<sup>e</sup> éd., Paris, 1984.

*Concordantia Herodotea* = C. SCHRADER, *Concordantia Herodotea* (*Alpha-Omega. Reihe A. Lexika. Indizes. Konkordanzen zur klassischen Philologie*, 98, 1-5), Hildesheim, Zürich, New York, 1996.

*Concordantia in Appianum* = É. FAMERIE, *Concordantia in Appianum. Concordance d'Appien* (*Alpha-Omega. Reihe A. Lexika. Indizes. Konkordanzen zur klassischen Philologie*, 133, 1-3), Hildesheim, Zürich, New York, 1993.

*Concordantia Thucydidea* = C. SCHRADER, *Concordantia Thucydidea* (*Alpha-Omega. Reihe A. Lexika. Indizes. Konkordanzen zur klassischen Philologie*, 99, 1-4), Hildesheim, Zürich, New York, 1998.

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.

DENIS, *Concordance* = A.-M. DENIS, *Concordance des Pseudépigraphes d'Ancien Testament. Concordance. Corpus des textes. Indices*, Louvain-la-Neuve, 1987.

DUBOIS – DUBOIS, *Introduction* = J. DUBOIS, Cl. DUBOIS, *Introduction à la lexicographie : le dictionnaire (Langue et langage)*, Paris, 1971.

DUBOIS, *Dictionnaire* = J. DUBOIS, M. GIACOMO, L. GUESPIN, Ch. MARCELLESI, J.-B. MARCELLESI, J.-P. MÉVEL, *Dictionnaire de linguistique*, 2<sup>e</sup> éd., Paris, 2001.

E = H. ESTIENNE, *Thesaurus Graecae Linguae*, post editionem anglicam novis additamentis auctum, ordineque alphabetico digestum, tertio ediderunt C.B. HASE, G. DINDORF, L. DINDORF, Paris, 1831-1865 (réimpr. Graz, 1954).

[Page 216]

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*

<sup>6</sup> = <http://www.cavi.univ-paris3.fr/lexicometrica/article/numero2.htm>.



historiques et philologiques, III. Hautes Études du Monde Gréco-Romain, 24), Genève, 1998.

*Le grand Gaffiot. Dictionnaire Latin Français*, ed. F. GAFFIOT, nouvelle édition revue et augmentée sous la dir. de P. FLOBERT, Paris, 2000.

GAUDIN – GUESPIN, *Initiation* = F. GAUDIN, L. GUESPIN, *Initiation à la lexicologie française. De la néologie aux dictionnaires (Champs linguistiques. Manuels)*, Bruxelles, 2000.

GÉRARD – KINDT, *Du D.A.G. au D.D.G.* = R. GÉRARD, B. KINDT, [D'un dictionnaire de lemmatisation \(D.A.G.\) à un dictionnaire dérivationnel du grec ancien \(D.D.G.\)](#), dans 7<sup>èmes</sup> Journées internationales d'Analyse Statistique des Données Textuelles, 10-12 mars 2004, Louvain-la-Neuve [à paraître].

JANNARIS, *Grammar* = A.N. JANNARIS, *An Historical Greek Grammar Chiefly of the Attic Dialect* [...], Londres, 1897 (réimpr. Hildesheim, 1968).

JUSTI, *Namenbuch* = F. JUSTI, *Iranisches Namenbuch*, Marburg, 1895 (réimpr. Hildesheim, 1963).

KEVERS – KINDT, *Vers un concordanceur* = L. KEVERS, B. KINDT, [Vers un concordanceur-lemmatiser en ligne du grec ancien](#), dans *L'Antiquité Classique*, 73 (2004), p. 203-213.

KINDT, *Avancées* = B. KINDT, [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](#), dans *Lexicometrica*, numéro spécial «Autour de la lemmatisation» (dir. D. Labbé) (2003), p. 1-17<sup>7</sup>.

L = G.W.H. LAMPE, *A Patristic Greek Lexicon*, Oxford, 1961 (réimpr. 1968).

LEHMANN – MARTIN-BERTHET, *Introduction* = A. LEHMANN, F. MARTIN-BERTHET, *Introduction à la lexicologie. Sémantique et morphologie (Lettres Sup.)*, Paris, 1998.

LEJEUNE, *Phonétique* = M. LEJEUNE, *Phonétique historique du mycénien et du grec ancien (Tradition de l'humanisme, 9)*, Paris, 1972.

LS = LIDDELL – SCOTT = 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.

MANDILARAS = B.G. MANDILARAS, *The Verb in the Greek Non-Literary Papyri*, Athens, 1973.

[Page 217]

MARTIN, *Sémantique* = R. MARTIN, *Sémantique et automate. L'apport du dictionnaire informatisé (Écritures Électroniques)*, Paris, 2001.

MELLET – PURNELLE, *Atouts*<sup>8</sup> = 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, vol. 2, Rennes, 2002, p. 529-538.

MELLET, *Lemmatisation*<sup>9</sup> = S. MELLET, [Lemmatisation et encodage grammatical : un luxe inutile?](#), dans *Lexicometrica*, 3 (2001), p. 1-5.

<sup>7</sup> = <http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema1.htm>.

<sup>8</sup> = <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/tocJADT2002.htm>.

<sup>9</sup> = <http://www.cavi.univ-paris3.fr/lexicometrica/article/numero3.htm>.

MORTUREUX, *Lexicologie* = M.-F. MORTUREUX, *La lexicologie entre langue et discours (Campus. Linguistique)*, Paris, 1997.

MULLER, *Initiation* = Ch. MULLER, *Initiation à la statistique lexicale (Langue et langage)*, Paris, 1968.

PB = W. PAPE, G.E. BENSELER, *Wörterbuch der griechischen Eigennamen*, Brunswig, 1870 (4<sup>e</sup> tir., 1911, réimpr. Graz, 1959).

PIRENNE-DELFORGE – PURNELLE, *Index* = V. PIRENNE-DELFORGE, G. PURNELLE, *Pausanias, Periegesis. Index Verborum. Listes de fréquence. Index Nominum (Travaux publiés par le Centre Informatique de Philosophie et Lettres. Laboratoire d'Analyse Statistique des Langues Anciennes, 24 ; Suppléments à Kernos, 5)*, Liège, 1997.

PSALTES, *Grammatik* = S.B. PSALTES, *Grammatik der byzantinischen Chroniken (Forschungen zur griechischen und lateinischen Grammatik, 2. Heft)*, Göttingen, 1913 (réimpr. 1974).

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).

SAUSSURE, *Cours* = F. DE SAUSSURE, *Cours de linguistique générale*, ed. Ch. BALLY, A. SECHEHAYE, A. RIEDLINGER, T. DE MAURO (*Payothèque*), Paris, 1974.

SCHMIDT = *Basilii Minimi in Gregorii Nazianzeni Orationem XXXVIII Commentarii*, ed. Th.S. SCHMIDT (*Corpus Christianorum. Series Graeca, 46. Corpus Nazianzenum, 13*), Turnhout, Leuven, 2001.

SCHWYZER, *Grammatik* = E. SCHWYZER, *Griechische Grammatik*, vol. 1. *Allgemeiner Teil. Lautlehre. Wortbildung. Flexion (Handbuch der Altertumswissenschaft, II, 1, 1)*, Munich, 1968 ; E. SCHWYZER, A. DEBRUNNER, *Griechische Grammatik*, vol. 2. *Syntax und syntaktische Stilistik (Handbuch der Altertumswissenschaft, II, 1, 2)*, Munich, 1966 ; E. SCHWYZER, D.J. GEORGACAS, *Griechische Grammatik*, vol. 3. *Register (Handbuch der Altertumswissenschaft, II, 1, 3)*, Munich, 1968.

[Page 218]

ThAgMyr = B. COULIE et CENTAL, *Thesaurus Agathiae Myrinaei. Historia et Epigrammata (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout, 2006.

ThAmIc = B. COULIE et CETEDOC, *Thesaurus Amphilochei Iconiensis (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout, 1994.

ThAstFirm = B. COULIE, B. KINDT et CETEDOC, *Thesaurus Asterii Amaseni et Firmi Caesariensis, Opera Omnia (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout, 2001.

ThAthAlex = B. COULIE, P. VAN DEUN et CENTAL, *Thesaurus Athanasii Alexandrini, Opera Omnia (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout [en préparation].

ThBasCaes = B. COULIE, B. KINDT et CETEDOC, *Thesaurus Basilii Caesariensis, Opera Omnia. Pars I. Introductio. Enumeratio Lemmatum et Formarum A-I. Pars II. Enumeratio Lemmatum et Formarum K-Ω (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout, 2002.

ThConcOec = B. COULIE et CETEDOC, *Thesaurus Conciliorum Oecumenicorum (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Turnhout, 1998.

ThGrNaz = J. MOSSAY et CETEDOC, *Thesaurus Sancti Gregorii Nazianzeni*, vol. I. *Enumeratio Lemmatum, Orationes, Epistulae, Testamentum* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1990 ; J. MOSSAY, B. COULIE et CETEDOC, *Thesaurus Sancti Gregorii Nazianzeni*, vol. II. *Enumeratio Lemmatum, Carmina, Christus Patiens, Vita* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1991.

ThGrNys = B. COULIE, B. KINDT et CENTAL, *Thesaurus Gregorii Nysseni, Opera Omnia* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout [en préparation].

ThLeoCons = J. NORET et CETEDOC, *Thesaurus Leontii Presbyteri Constantinopolitani* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1992.

ThMenPro = B. COULIE et CENTAL, *Thesaurus Menandri Protectoris* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Louvain-la-Neuve, Turnhout, 2006.

ThPhotCons = J. SCHAMP, B. KINDT et CENTAL, *Thesaurus Photii Constantinopolitani, Bibliotheca* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 2005.

ThProcCaes = B. COULIE, B. KINDT et CETEDOC, *Thesaurus Procopii Caesariensis, De Bellis, Historia Arcana, De Aedificiis* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 2000.

ThPsNon, Com. = B. COULIE, J. NIMMO SMITH et CETEDOC, *Thesaurus Pseudo-Nonni, Commentarii in IV Orationes Gregorii Nazianzeni* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1999.

[Page 219]

ThPsNonPan, *Paraphr.* = B. COULIE, L.F. SHERRY et CETEDOC, *Thesaurus Pseudo-Nonni quondam Panopolitani, Paraphrasis Evangelii S. Ioannis* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1995.

ThTheoConf = B. COULIE, P. YANNOPOULOS et CETEDOC, *Thesaurus Theophanis Confessoris*, vol. I. *Chronographia* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 1998 ; B. COULIE, B. KINDT, P. YANNOPOULOS et CETEDOC, *Thesaurus Theophanis Confessoris*, vol. II. *Index Nominum* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 2000.

ThTheoSim = A. DE SIENA et CENTAL, *Thesaurus Theophylacti Simocattae* (*Corpus Christianorum. Thesaurus Patrum Graecorum*), Turnhout, 2007.

TOMBEUR, *Thesaurus = Thesaurus Linguae Scriptorum Operumque Latino-Belgicorum Medii Aevi*, Première partie. *Le vocabulaire des origines à l'an mil* ; vol. I. *Méthodologie et informatique : du texte aux analyses. Index général des lemmes* ; vol. II. *Index des lemmes selon les siècles et les genres littéraires. Index inverse des lemmes. Fréquences décroissantes* ; vol. III. *Index des formes et des lemmes*, ed. P. TOMBEUR, Bruxelles, 1996.

## LISTE DES ABRÉVIATIONS

act. = actif	impft = imparfait	pron. = pronom
adj. = adjectif	ind. = indicatif	sing. = singulier
adv. = adverbe	inf. = infinitif	<i>s.l.</i> = <i>sub lemmate</i>
anthr. = anthroponyme	lat. = latin	subj. = subjonctif
aor. = aoriste	masc. = masculin	subst. = substantif
comp. = comparatif	nom. = nominatif	sup. = superlatif
conj. = conjonction	par ex. = par exemple	<i>T.L.G.</i> = <i>Thesaurus Linguae</i>
CP = Constantinople	part. = participe	<i>Graecae</i>
<i>D.A.G.</i> = <i>Dictionnaire</i>	parf. = parfait	top. = toponyme
<i>Automatique Grec</i>	pas. = passif	<i>T.P.G.</i> = <i>Thesaurus Patrum</i>
dat. = datif	pers. = personne	<i>Graecorum</i>
ethn. = subst. ou adj.	pl. = pluriel	<i>s.v.</i> = <i>sub verbo</i>
ethnique	prép. = préposition	v. = verbe
excl. = exclamation	prés. = présent	<i>vs</i> = <i>versus</i>
fém. = féminin		
gén. = génitif		

## I. DÉFINITION DU PROJET (§1-22)

1. La lemmatisation en grec ancien (§1-7). 2. Le D.A.G. (§8-15). 3. L'enquête lexicographique (§16-19). 4. Les principes d'économie et de fidélité (§20-22).

### 1. La lemmatisation en grec ancien

**§1.** La *lemmatisation* est l'opération par laquelle les formes d'un texte, ou d'un *corpus* de textes, sont classées sous le mot qui pourrait leur servir d'«entrée [Page 220] lexicale» dans un dictionnaire. Cette entrée est appelée le *lemme*. Les méthodes de traitement automatique des langues permettent d'effectuer des dénombrements et des tris sur les mots d'un *corpus* afin d'y appliquer, ensuite, des traitements statistiques. Dans le cas des langues à flexion, le lemme permet de plus d'éviter de compter séparément les formes fléchies d'un même mot : les mots  $\phi\omega\nu\alpha\acute{\iota}\varsigma$ ,  $\phi\omega\nu\acute{\eta}$  et  $\phi\omega\nu\eta\varsigma$  seront considérés comme les formes fléchies du seul lemme  $\phi\omega\nu\acute{\eta}$ . Dans le cas du grec ancien, des *variations phonétiques*, *accentuelles* et *dialectales* accompagnent les variations flexionnelles et multiplient, parfois de manière sensible, le nombre des formes regroupées sous chaque entrée lexicale.

**§2.** Dans son état actuel, le D.A.G. rassemble 238 formes différentes du lemme  $\gamma\acute{\iota}\nu\omicron\mu\alpha\iota$ . Le paradigme de ce verbe n'est cependant pas encore complet car il ne comprend que des formes rencontrées dans le *corpus* actuel des textes analysés. Les variations phonétiques sont illustrées par des formes du type  $\acute{\epsilon}\gamma\acute{\iota}\nu\epsilon\tau\omicron/\acute{\epsilon}\gamma\acute{\iota}\nu\epsilon\tau\omicron$  ou  $\acute{\epsilon}\gamma\acute{\iota}\nu\epsilon\tau'/\acute{\epsilon}\gamma\acute{\iota}\nu\epsilon\tau'$ , les variations accentuelles par les couples  $\gamma\acute{\iota}\nu\epsilon\sigma\theta\alpha\iota/\gamma\acute{\iota}\nu\epsilon\sigma\theta\alpha\acute{\iota}$  ou  $\gamma\acute{\iota}\nu\epsilon\sigma\theta\alpha\acute{\iota}/\gamma\acute{\iota}\nu\epsilon\sigma\theta\alpha\iota$ , les variations dialectales par  $\gamma\acute{\iota}\nu\omicron\mu\acute{\epsilon}\nu\omicron\upsilon$  ou  $\gamma\acute{\iota}\nu\omicron\mu\acute{\epsilon}\nu\omicron\upsilon$  en face de  $\gamma\acute{\iota}\nu\omicron\mu\acute{\epsilon}\nu\omicron\iota\omicron$ . Le lemme  $\upsilon\acute{\iota}\omicron\varsigma$  rassemble à lui seul trente-six formes différentes, de  $\upsilon\acute{\iota}\omicron\varsigma$  à  $\upsilon\acute{\iota}\acute{\epsilon}\epsilon\sigma\iota$  ou  $\upsilon\acute{\iota}\eta\sigma\iota\nu$ . Ces variations affectent non seulement les mots soumis aux flexions nominales et verbales, mais aussi les invariables : l'adverbe  $\acute{\alpha}\epsilon\acute{\iota}$  regroupe les quatre formes  $\acute{\alpha}\epsilon\acute{\iota}$ ,  $\acute{\alpha}\iota\acute{\epsilon}\acute{\iota}$ ,  $\acute{\alpha}\acute{\iota}\acute{\epsilon}\nu$  et  $\acute{\alpha}\acute{\iota}\acute{\epsilon}\varsigma$  ; la préposition  $\pi\epsilon\rho\acute{\iota}$  rassemble les formes  $\pi\epsilon\rho\acute{\iota}$ ,  $\pi\acute{\epsilon}\rho\iota$  et  $\pi\epsilon\rho'$  ; la conjonction  $\eta\acute{\nu}\acute{\iota}\kappa\alpha$  apparaît sous les formes  $\eta\acute{\nu}\acute{\iota}\kappa\alpha$ ,  $\eta\acute{\nu}\acute{\iota}\kappa'$  ou  $\eta\acute{\nu}\acute{\iota}\chi'$ . Les quatre types de variations énoncées combinent de plus leur effets : la préposition  $\pi\rho\acute{\omicron}\sigma\omega$  rassemble les six formes  $\pi\rho\acute{\omicron}\sigma\omega$ ,  $\pi\acute{\omicron}\rho\rho\omega$ ,  $\pi\rho\omicron\sigma\omega\tau\acute{\epsilon}\rho\omega$ ,  $\pi\omicron\rho\rho\omega\tau\acute{\epsilon}\rho\omega$ ,  $\pi\omicron\rho\rho\omega\tau\acute{\epsilon}\rho\omega$  et  $\pi\omicron\rho\rho\omega\tau\acute{\alpha}\tau\omega$ .

**§3.** Le lemme est souvent défini de manière métaphorique : il est «l'adresse» des formes qu'il rassemble<sup>10</sup>, une «étiquette conventionnelle»<sup>11</sup>, une «unité théorique [et] une forme arbitraire»<sup>12</sup>, une «"entrée lexicale" du type "entrée de dictionnaire"»<sup>13</sup>, la «forme de citation» d'un mot<sup>14</sup>, une «forme canonique, [...] un artifice métalinguistique»<sup>15</sup>, etc. Trois notions complémentaires émergent de ces définitions et déterminent l'essence du *lemme* :

[Page 221]

- il est une «forme de regroupement»<sup>16</sup>, afin de rassembler sous un même élément dénombrable les différentes formes qui s'y rapportent, et de rendre possibles tris et calculs ;

- il est une «forme de départ»<sup>17</sup>, une forme censée être à la tête d'une série productive, que cette forme soit attestée ou non ;

- il est une «forme neutralisée du discours métalinguistique»<sup>18</sup>, afin d'être utile à la description sans être affecté de toutes les transformations – dont l'inventaire reste à faire – que lui imposent les règles qui président aux variations flexionnelles, phonétiques, accentuelles ou dialectales.

**§4.** La lemmatisation peut être automatisée, au moins partiellement. Les *corpus* analysés sont alors normalisés afin que les textes reçoivent un format compatible aux systèmes de traitement informatique. À chaque étape de l'analyse, les informations relatives à la référence du mot dans le texte restent attachées aux formes<sup>19</sup>.

<sup>10</sup> J. PRUVOST, "Introductions" et "Manuels" de lexicologie (1995-1998): un très bon cru à l'aube du XXI<sup>e</sup> s., dans *Cahiers de lexicologie*, 76 (2000), p. 194.

<sup>11</sup> REY-DEBOVE, *Étude*, p. 147 ; DUBOIS – DUBOIS, *Introduction*, p. 61 ; MELLET – PURNELLE, *Atouts*, p. 529 ; LEHMANN – MARTIN-BERTHET, *Introduction*, p. 1-2.

<sup>12</sup> DUBOIS – DUBOIS, *Introduction*, p. 61.

<sup>13</sup> TOMBEUR, *Thesaurus*, I, p. 28 ; COULIE, *Lemmatisation*, p. 37-38.

<sup>14</sup> APOTHÉLOZ, *Construction*, p. 11.

<sup>15</sup> GAUDIN – GUESPIN, *Initiation*, p. 112.

<sup>16</sup> R. ÉLUERD, *La lexicologie (Que sais-je?, 3548)*, Paris, 2000, p. 24-25 ; BRUNET, *Qui lemmatise*, p. 1.

<sup>17</sup> COULIE, *Lemmatisation*, p. 38.

<sup>18</sup> REY-DEBOVE, *Étude*, p. 174, qui reprend les propos d'É. Benveniste à propos de l'infinitif français, cfr É. BENVENISTE, *Problèmes de linguistique générale (Tel)*, vol. 1, Paris, 1966, p. 255.

<sup>19</sup> Sur la préparation des données, cfr COULIE, *Lemmatisation*, p. 39-40.

Ces dernières sont ensuite classées sous les lemmes correspondants. En pratique, une suite de programmes compare les données lexicales du *corpus* étudié à celles d'un dictionnaire de référence<sup>20</sup> :

- une forme du *corpus* connue du dictionnaire est «reconnue» par le logiciel et reçoit le lemme qui lui correspond : φωναῖς, φωνή et φωνῆς *s.l.* φωνή ;

- une forme du *corpus* connue du dictionnaire mais faisant l'objet de plusieurs analyses est «reconnue» par le logiciel qui lui attribue les différents lemmes possibles ; c'est à celui qui lemmatise, le lemmatiseur, que reviendra la tâche de définir quel lemme est pertinent pour telle forme *in textu* : φωνῆ, φωνῆς et φωνῶν *s.l.* φωνή et φωνέω<sup>21</sup> ;

[Page 222]

- une forme du *corpus* inconnue du dictionnaire n'est pas «reconnue» par le logiciel et ne reçoit aucune proposition de lemme ; c'est au lemmatiseur que reviendra la tâche de lui en attribuer un :

- si ce dernier existe déjà dans le dictionnaire, il l'attribue à la forme ;

- si la forme ne correspond à aucun lemme déjà enregistré, le lemmatiseur doit vérifier l'exactitude de la forme – car une forme présentant une coquille orthographique ne sera pas reconnue<sup>22</sup> – pour ensuite formuler un nouveau lemme adéquat.

La lemmatisation proposée par l'ordinateur demeure toujours partielle et provisoire, toujours sujette à d'éventuelles corrections, et, malgré le développement des technologies mises en oeuvre, une intervention humaine demeure indispensable<sup>23</sup>.

<sup>20</sup> Sur la lemmatisation des sources classiques, cfr, par ex., BRUNET, *Qui lemmatise* ; MELLET, *Lemmatization* ; MELLET – PURNELLE, *Atouts*. Sur les traitements applicables aux données lexicales qui en sont issues, cfr, par ex., ÉVRARD – MELLET, *Méthodes quantitatives* ; S. MELLET, *Quelques réflexions sur l'exploitation statistique des données informatisées*, dans *Les Études Classiques*, 58 (1990), p. 105-113.

<sup>21</sup> Le logiciel ne tranche pas en faveur de tel ou tel lemme. En d'autres termes, dans son état actuel, il ne fait pas de sémantique, et la lemmatisation est dite «formelle», aspect qui sera détaillé ci-dessous ; sur les formes ou lemmes homographes, cfr **§34-44**. Un programme d'assistance à la désambiguïsation est actuellement mis au point au CENTAL par Mr L. Kevers ; cfr KINDT, *Avancées*, p. 10-12.

<sup>22</sup> Sur les fautes d'édition, cfr **§109**.

<sup>23</sup> GAUDIN – GUESPIN, *Initiation*, p. 97.

§5. Quand une analyse porte sur des *corpus* différents, les formes identiques issues des diverses sources étudiées doivent être représentées par les mêmes lemmes. La formulation de ces derniers doit donc reposer sur des règles strictes et durables assurant la cohérence interne de la description au-delà des œuvres traitées ou des personnes qui lemmatisent ; la lemmatisation implique la définition d'une *norme* explicite, d'un *standard*<sup>24</sup>, comprenant, au minimum, la « désignation d'un dictionnaire » et « la liste des quelques cas où l'on entend s'écarter de la nomenclature (de ce dernier) »<sup>25</sup>.

Le choix d'une norme, c'est-à-dire le choix du dictionnaire de référence, constitue une gageure. De nombreux outils lexicographiques sont à la disposition des hellénistes. Mais deux remarques doivent être formulées :

- aucun dictionnaire n'est vraiment *exhaustif* : les *corpus* dépouillés ne sont pas les mêmes d'un dictionnaire à l'autre et, pour une même source, les dénombrements ne sont jamais complets ;

- aucun dictionnaire n'est parfaitement *cohérent* : les dictionnaires fournissent des entrées différentes pour des mots rares ou des mots présentant plusieurs [Page 223] graphies, dégroupent<sup>26</sup> différemment

<sup>24</sup> LABBÉ, *Analyse*, p. 7.

<sup>25</sup> MULLER, *Initiation*, p. 142-144. Cfr aussi, par ex., J. BAJARD, *De la Septante au Nouveau Testament. Approche quantitative de la Bible grecque*, dans *Le nombre et le texte. Hommage à Étienne Évrard*, = *Revue Informatique et Statistique dans les Sciences Humaines*, 24 (1988), p. 20 (qui cite Ch. Muller): « la qualité la plus assurée d'une norme lexicologique est de ne satisfaire personne, à commencer par celui qui l'a élaborée [...]. Ce qui importe, c'est de ne pas avoir à comparer entre eux des résultats obtenus avec des normes différentes, et surtout avec des normes que leurs auteurs n'ont pas jugé bon d'explicitier » ; BRUNET, *Qui lemmatise*, p. 3 ; ÉVRARD – MELLET, *Méthodes quantitatives*, p. 126-127 ; MELLET, *Lemmatisation*, p. 3.

<sup>26</sup> En lexicographie, le « dégrouplement » correspond à l'attitude du lexicographe qui décide d'enregistrer deux entrées différentes pour des mots homographes. Il opte ainsi pour une analyse homonymique (qui distingue deux termes différents dans le même mot), et non polysémique (qui enregistre différents sens pour un même mot), motivant son choix par des critères linguistiques qui devraient être, en théorie, explicites.



les entrées homographes, enregistrent comme entrées lexicales des formes fléchies, etc.<sup>27</sup>

Dans les dictionnaires traditionnels, les notions de «forme de départ» (qui permettrait de définir des règles pour formuler les lemmes de mots rares dont le nom. ou la 1<sup>ère</sup> pers. ne sont pas attestés dans les sources, et pour choisir le lemme des formes présentant plusieurs graphies), de «forme de regroupement» (qui permettrait de définir des règles pour distinguer les formes homographes, pour choisir entre un traitement homonymique ou polysémique d'un mot), et de «forme neutralisée du discours métalinguistique» (qui permettrait de distinguer ce qui doit être enregistré comme «entrée lexicale» et ce qui doit être classé sous une «entrée lexicale»), n'ont pas fait l'objet d'une étude linguistique préalable<sup>28</sup>.

Pourtant, d'une manière assez générale, les auteurs choisissent leur dictionnaire de référence parmi les outils traditionnels, le *Greek-English Lexicon* dans la majorité des cas<sup>29</sup>. Pour les lemmes particuliers (mots présentant plusieurs graphies, mots absents du dictionnaire choisi), ils fournissent un lemme déduit des formes attestées dans le corpus analysé. L'auteur de la concordance d'Appien écrit : «partant du principe que notre but était avant tout d'analyser la langue d'Appien et non de constituer un dictionnaire de référence, nous avons [Page 224] élaboré un système de lemmatisation qui tient

<sup>27</sup> Pour une critique des dictionnaires en général, cfr CORBIN, *Morphologie*, p. 21-44 ; REY-DEBOVE, *Étude*, p. 60-64. Les dictionnaires de grec ancien n'ont pas encore fait l'objet d'examen analogues ; cfr cependant F.R. ADRADOS, E. GANGUTIA, J. LOPEZ FACAL, C. SERRANO AYBAR, *Introducción a la lexicografía griega (Consejo Superior de Investigaciones Científicas. Manuales y Anejos de «Emerita», 33)*, Madrid, 1977, p. 107-133 ; O. MASSON, *Pape-Benseleriana VIII. Remarques sur le Wörterbuch de W. Pape et C.E. Benseler (1863-1870)*, dans *Z.P.E.*, 42 (1981), p. 193-204.

<sup>28</sup> Dans le domaine de l'étude du français, certains linguistes n'ont pas hésité à parler de dictionnaires «athéoriques» en parlant des dictionnaires traditionnels (M. TEMPLE, *Pour une sémantique des mots construits [Sens et Structure]*, Villeneuve d'Ascq, 1996, p. 25-27). Cette appellation est applicable aux dictionnaires du grec ancien.

<sup>29</sup> *Concordantia in Appianum*, I, p. XV ; DENIS, *Concordance*, p. XII ; cfr aussi BODSON, *Index*, p. V ; PIRENNE-DELFORGE – PURNELLE, *Index*, p. VI (pour des index lemmatisés) ; *a contrario*, l'introduction des concordances d'Hérodote et de Thucydide ne fait état d'aucun dictionnaire de référence et les principes de lemmatisation sont décrits en quelques lignes, cfr *Concordantia Herodotea*, I, p. XIV-XV ; *Concordantia Thucydidea*, I, p. VIII.

compte de l'usage propre à cet auteur»<sup>30</sup>. Dans ce cas, la lemmatisation ne dépasse pas le niveau du lexique propre au *corpus* analysé.

**§6.** Dans le cadre du *Projet*, les principes de lemmatisation se basent sur une conception différente de la langue et de son lexique. En aval de l'opposition saussurienne entre *langue* et *discours*, les lexicologues et lexicographes établissent la distinction entre *lexèmes* et *vocables*<sup>31</sup>. Les lexèmes sont les représentants des vocables au niveau de la langue. L'ensemble des lexèmes d'une langue ou d'un *corpus* en constitue le *lexique*. Les vocables sont les actualisations discursives des lexèmes. Ils constituent le *vocabulaire* d'un *corpus* et, en tant qu'unités du discours, ils sont les seules réalités tangibles et directement dénombrables. Les lexèmes demeurent donc des abstractions, car, pour former le tissu d'un texte, «aucun auteur ne s'exprime en lemmes»<sup>32</sup>.

Pour les concepteurs du *Projet*, la dichotomie entre langue et discours d'abord, entre lexèmes et vocables ensuite, est également assumée au niveau des *lemmes* et des *formes*. Chaque analyse d'un *corpus* génère des données lexicales qui, rassemblées, nourrissent le dictionnaire de référence, en l'occurrence le *D.A.G.* Toutes les données lexicales nouvelles, lemmes et formes, fruits d'une analyse antérieure, y sont incorporées. Enrichi par ces matériaux, le dictionnaire est utilisé pour des comparaisons ultérieures qui, elles aussi, amènent leur lot de nouveautés.

La formulation d'un nouveau lemme ne se base plus simplement sur les formes attestées dans la source analysée mais sur toutes les formes censées relever du même lemme et accessibles dans l'ensemble des sources, une approche holistique de la forme en *discours* permettant de définir un lemme pertinent en *langue*. Le lemme n'est plus ici une «étiquette des formes», mais une «étiquette du lexème». La comparaison des lemmes du *D.A.G.* à ceux des concordances lemmatisées de Thucydide et d'Hérodote permet d'illustrer ce fait : les mots ἀβουλίαν et ἀβουλίην sont classés sous le lemme ἀβουλία, ἡ dans celle de Thucydide et sous ἀβουλίη, ἡ dans

<sup>30</sup> *Concordantia in Appianum*, p. XVIII ; É. FAMERIE, *Appien. Histoire romaine. Index nominum* (Travaux publiés par le Centre Informatique de Philosophie et Lettres. Série du Laboratoire d'Analyse Statistique des Langues Anciennes, 25), Liège, 1998, p. VIII ; une déclaration analogue figure dans PIRENNE-DELFORGE – PURNELLE, *Index*, p. VI.

<sup>31</sup> Pour une description de cette «opposition saussurienne», cfr GAUDIN – GUESPIN, *Initiation*, p. 15-16 et 170-172 ; MORTUREUX, *Lexicologie*, p. 5-6 et 11-12.

<sup>32</sup> COULIE, *Lemmatisation*, p. 37.

celle d'Hérodote<sup>33</sup> ; le D.A.G. connaît les [Page 225] deux formes mais n'enregistre que le seul lemme, ἀβουλία. De la même manière, dans les concordances précitées, un lemme δέω rassemble les formes de δέω «lier» et δέω «manquer»<sup>34</sup>. Le D.A.G. distingue δέω (δήσω) pour δέω «lier» et δέω (δεήσω) pour δέω «manquer». La lemmatisation se situe alors au niveau de la langue<sup>35</sup>.

Ce *modus operandi* permet d'accomplir des dénombrements complets pour les sources abordées (ce qui justifie le nom de *Thesaurus* donné aux concordances lemmatisées publiées) et de garantir un traitement homogène, quasi au cas par cas, tant des lemmes déjà enregistrés que des lemmes nouveaux.

**§7.** La lemmatisation reste une opération qui réclame rigueur et patience<sup>36</sup>. Elle représente un coût certain régulièrement mis en évidence<sup>37</sup>. La méthode présentée dans ces lignes, qui tire profit des analyses précédentes pour améliorer peu à peu la rentabilité des analyses futures, réduit sensiblement ce coût. Elle suppose aussi remplies trois conditions :

- le D.A.G. doit être bâti sur une structure souple et ouverte : le D.A.G. est un *dictionnaire électronique* (§8-15) ;
- l'ensemble des sources disponibles doit faire l'objet d'un examen apportant les éléments nécessaires et suffisants pour la formulation pertinente des nouveaux lemmes : une *enquête lexicographique* – organisée en quatre volets – permet de prendre connaissance du comportement d'un mot dans cet «ensemble des sources disponibles» (§16-19) ;
- les principes de lemmatisation doivent être explicites : c'est l'objet principal de cet article.

<sup>33</sup> *Concordantia Thucydidea*, I, p. 1 ; *Concordantia Herodotea*, I, p. 1.

<sup>34</sup> *Concordantia Herodotea*, I, p. 473-476 ; *Concordantia Thucydidea*, I, p. 367-369.

<sup>35</sup> Sur la distinction des lemmes homographes dans le D.A.G., cfr §34-44.

<sup>36</sup> TOMBEUR, *Thesaurus*, I, p. 56.

<sup>37</sup> BRUNET, *Qui lemmatise*, p. 3 et 18-19 ; MELLET, *Lemmatization*, p. 4.

## 2. Le D.A.G.

§8. Comme le montre le *tableau 1*, les textes déjà analysés et les textes en cours de traitement, mis bout à bout, constitueraient un *corpus* de 4.284.343 mots-occurrences<sup>38</sup>.

[Page 226]

Les lemmatisations successives ont permis d'en extraire 174.758 formes différentes classées sous 33.874 lemmes. La majorité des sources traitées relève de la patristique (2.787.998 occurrences, dont 2.027.998 issues des textes des Pères Cappadociens, soit 47,33% de la totalité du *corpus*) et de l'historiographie byzantine (803.094 occurrences, soit 18,74% de la totalité du *corpus*). Ces lemmes et ces formes constituent respectivement la macrostructure et la microstructure<sup>39</sup> du dictionnaire de référence, base de la lemmatisation dans le cadre du *Projet*.

<sup>38</sup> La dénomination des chantiers renvoie aux *Thesauri*; 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, c'est-à-dire: le vocabulaire des pseudépigraphes de l'*Ancien Testament* (DENIS, *Concordance*); le vocabulaire d'Appien (II<sup>e</sup> s.; *Concordantia in Appianum*); le vocabulaire de la *Vita Silvestri Romani* (B.H.G. 1628-1629; = F. COMBÉFIS, *Illustrium Christi martyrum lecti triumpho*, Paris, 1660, p. 258-336); le vocabulaire des *Commentaires* de Basile le Minime (X<sup>e</sup> s.) sur le *Discours XXXVIII* de Grégoire de Nazianze (= SCHMIDT; pour le texte grec, cfr p. 2-119; la lemmatisation a fourni l'index exhaustif de l'édition, p. 121-194); le vocabulaire de la *Vie des saints Jason et Sosipatros* (X<sup>e</sup> s.; B.H.G. 776; = B. KINDT, *La version longue du récit légendaire de l'évangélisation de l'île de Corfou par les saints Jason et Sosipatros. Entre mythe et réalité*, dans A.B., 116 [1998], p. 253-295; pour le texte grec, cfr p. 260-295). 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.

<sup>39</sup> Sur ces deux notions, cfr REY-DEBOVE, *Étude*, p. 21, 114 et 148; GAUDIN – GUESPIN, *Initiation*, p. 110-112.

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

Chantiers	Occurrences	Lemmes	Formes
ThAgMyr (VI <sup>e</sup> s.)	66.659		
ThAmIc (IV <sup>e</sup> s.)	38.740	4.158	9.989
ThAstFirm (Astérius) (IV <sup>e</sup> -V <sup>e</sup> s.)	48.174	5.479	13.799
ThAstFirm (Firmus) (IV <sup>e</sup> -V <sup>e</sup> s.)	4.181	1.130	1.894
ThAthAlex* (IV <sup>e</sup> s.)	760.000		
ThBasCaes (IV <sup>e</sup> s.)	707.853	14.843	68.867
ThConcOec ( <i>varia</i> )	27.011	2.986	6.638
ThGrNaz, I (IV <sup>e</sup> s.)	268.327	10.506	41.617
ThGrNaz, II (IV <sup>e</sup> s.)	153.723	9.854	31.304
ThGrNys* (IV <sup>e</sup> s.)	807.000		
ThLeoCons (VI <sup>e</sup> s.)	30.819	3.679	8.329
ThMenPro (VI <sup>e</sup> s.)	29.990		
ThPhotCons (IX <sup>e</sup> s.)	347.145	17.158	54.184
ThProcCaes (VI <sup>e</sup> s.)	292.552	7.982	28.997
ThPsNon, <i>Comm.</i> (VI <sup>e</sup> s.)	20.248	2.818	5.579
ThPsNonPan, <i>Paraph.</i> (V <sup>e</sup> -VI <sup>e</sup> s.)	23.391	3.389	7.356
ThTheoConf (VIII <sup>e</sup> -IX <sup>e</sup> s.)	131.948	8.692	22.448
ThTheoSim (VI <sup>e</sup> -VII <sup>e</sup> s.)	78.874		
Autres	313.398		
Total général	4.284.343		
Total Cappadociens	2.027.998		
Total historiens**	803.094		

[Page 227]

§9. Le D.A.G. est un dictionnaire électronique structuré en une base de données relationnelle. Par rapport aux dictionnaires «papier» traditionnels, un dictionnaire électronique présente une structure souple et ouverte comprenant, au moins, trois avantages : possibilité d'extension ; possibilité de révision ; possibilité d'extraction<sup>40</sup>.

1. *Possibilité d'extension* : chaque lemmatisation amène son lot de formes nouvelles et de lemmes nouveaux. Le tableau 2 fournit la proportion de nouveaux lemmes et de nouvelles formes acquis au terme de chantiers récents et indique dans quelle mesure cela a permis d'accroître le D.A.G. Une procédure permet de mettre à jour le

<sup>40</sup> Sur ces trois notions, cfr G. GROSS, *Élaboration d'un dictionnaire électronique*, dans B.S.L., 94 (1999), p. 114 et, pour les termes employés, MARTIN, *Sémantique*, p. 61-66 ; R. Martin ajoute la *possibilité de synthèse avec d'autres ouvrages informatisés*, ce qui correspondrait, dans le cas actuel, à la compatibilité du dictionnaire avec d'autres outils, en l'occurrence le concordanceur-lemmatiseur et les modules de désambiguïsation ou de caractérisation morphologique des lemmes. Cfr aussi M. ZOCK, J. CARROLL, *Les dictionnaires électroniques*, dans *Traitement automatique des langues*, 44/2 (2003), p. 7-8.

dictionnaire qui, enrichi de données lexicales nouvelles, renouvellera ses comparaisons sur des bases plus larges.

Tableau 2 : proportion des nouveaux lemmes et des nouvelles formes dans quelques chantiers récents et taux d'accroissement du D.A.G.

Chantiers	Occurrences	Lemmes	Nouveaux lemmes	Accroissement du D.A.G.	Formes	Nouvelles formes	Accroissement du D.A.G.
ThTheoConf	131.948	8.692	2.927	9,58%	22.448	9.466	5,95%
ThConcOec	2.701	2.986	707	2,26%	6.638	1.583	0,99%
ThPsNon	20.248	2.818	341	1,08%	5.579	1.616	1,00%
ThProcCaes	292.552	7.982	2.281	6,73%	28.997	12.437	7,12%
ThBasCaes	707.853	14.843	3.014	9,89%	68.867	13.381	9,34%
ThPhotCons	347.145		4.521	13,32%	18.525		13,00%

Étant susceptible de s'enrichir sans limite, la nomenclature du D.A.G. peut accueillir les noms propres, les numéraux et leurs dérivés respectifs<sup>41</sup>, mots négligés dans les dictionnaires «papier» traditionnels car susceptibles de provoquer un accroissement exponentiel de leur macrostructure<sup>42</sup>. En grec ancien, les noms propres fournissent par dérivation ou composition :

- des noms communs tirés, originellement, de toponymes<sup>43</sup> : Ἀττικισμός, Λακωνισμός, Σκυθισμός, Ὠριγενιασμός ;

[Page 228]

- des noms communs tirés d'anthroponymes : Μανιχαΐσμός, Ὠριγενιασμός ;

- des adjectifs tirés de toponymes : φιλαθήναιος ;

- des adjectifs tirés d'anthroponymes : Αἰσώπειος, Δαυϊδικός, Διοήτηρ, Ἰσοκρατικός, Κυρίλλειος, Λυσιακός, Μιθριδάτειος, ψευδαντωνίνος, Ὠριγενιαστής ;

<sup>41</sup> Sur le traitement des numéraux dans le D.A.G., cfr §66-71.

<sup>42</sup> Pour une illustration des conséquences de la présence des noms propres, et de leurs dérivés, sur l'accroissement de la nomenclature des dictionnaires, cfr REY-DEBOVE, *Étude*, p. 69-70, 88-90 et 94 ; V. TOLÉDANO, D. CANDEL, *Dérivation suffixale de toponymes: étude d'un terrain propice à la création lexicale*, dans *Meta*, 47 (2002), p. 120-121 (pour des dérivés toponymiques) ; sur le statut des noms des «êtres de fiction», en général, cfr MARTIN, *Sémantique*, p. 27, et, dans le cas du grec (noms de divinités, noms de héros), PIRENNE-DELFORGE – PURNELLE, *Index*, p. X.

<sup>43</sup> Dans ces lignes, la définition des toponymes demeure, du moins pour l'instant, très large. Cette notion rassemble tant les noms de lieu que les appellations relevant de l'orographie et de l'hydrographie.

- des adverbes tirés de toponymes : Αἴγυπτισί, Λυκαο-  
νιστί, Ὀλυμπίαζε ;
- des adverbes tirés d'anthroponymes : Δημοσθενικῶς ;
- des verbes tirés de toponymes : ἀττικίζω, ἐξαττικίζω, ὑπεραττικίζω,  
διοनुσιάζω, ἰωνίζω ;
- des verbes tirés d'anthroponymes : πλατωνίζω ; φιλιππίζω,  
φιλωνίζω (< Φίλων), μακκοάω (< Μάκκω selon Suidas).

Quant aux numéraux, ils produisent par dérivation ou composition :

- des noms communs : πεντηκοντάρχος ;
- des adjectifs : δεκαδικός ; τεταρταῖος ; τριαδικός ;
- des adverbes : πεντάκις ; τριακοντάκις ;
- des verbes : τριάζω.

Une description globale du lexique doit prendre en compte ces créations lexicales.

2. *Possibilité de révision* : chaque lemmatisation permet de tester la pertinence de la formulation des lemmes. Certains usages d'une forme, négligés dans un premier temps, peuvent justifier la correction d'un lemme ou l'attribution d'une forme à un autre lemme : l'index des noms propres tiré de la concordance de la *Chronique* de Théophane le Confesseur corrige le lemme Μακεδονία en Μακεδονία (τοπ.), afin d'établir une distinction entre le lemme toponymique «la Macédoine» et le lemme anthroponymique Μακεδονία (προσ.) «Macédoine», mots attestés dans l'œuvre de Procope de Césarée traitée après celle du chroniqueur ; le lemme Ἄνται, ethnique<sup>44</sup> du peuple des Antes, est corrigé en Ἄντης, sur base d'attestations rencontrées au singulier, également chez [Page 229] Procope de Césarée<sup>45</sup> ; l'entrée Βεῖουδαές, plus proche de l'étymon hébreux *Bê Iûdâ*<sup>46</sup>, remplacera le lemme Βεῖουδές, forme attestée dans la *Chronique* de Théophane le Confesseur traitée antérieurement.

L'évolution du *Projet* justifie parfois une révision des principes de lemmatisation. Ainsi, pour le traitement des verbes supplétifs, les

<sup>44</sup> L'appellation «ethnique» regroupe tous les noms ou adjectifs désignant un peuple ou les habitants d'une région. Les ethniques sont, dans la majorité des cas, dérivés d'un toponyme.

<sup>45</sup> Cfr ThTheoConf, II, p. IX-XI ; ThProcCaes, p. XIII-XIV.

<sup>46</sup> R.E., V, col. 196-197 (Stuttgart, 1897), s.v. *Beiudaes*.

règles formulées jadis suivaient partiellement celles du *Greek-English Lexicon*<sup>47</sup> : λέγω, ἔρω et εἶπον étaient dégroupés (de même que ὀράω et εἶδον, et ἐσθίω, ἔδω et φαγεῖν), mais ἦλθον, δραμοῦμαι et οἴσομαι étaient regroupés respectivement sous ἔρχομαι, τρέχω et φέρω. Les formes du type ὄπωπα ne faisaient cependant pas l'objet d'une entrée lexicale particulière et étaient classées sous ὀράω, contrairement au dictionnaire de référence explicitement proposé. Ces principes de lemmatisation, débouchant sur une représentation lexicographique peu satisfaisante dans le cadre d'un traitement automatique, ont donc été revus.

3. *Possibilité d'extraction* : les données lexicales du D.A.G. sont interrogeables soit à partir des lemmes, soit à partir des formes. Ce dernier mode d'interrogation permet d'atteindre, par exemple, les formes supplétives ne faisant pas directement l'objet d'une «entrée lexicale»<sup>48</sup>.

Un projet en cours de développement décrit la constitution morphologique de l'ensemble des lemmes. Il vise à dénombrer les morphèmes constitutifs des mots et à mettre en relation les lemmes présentant des morphèmes communs, tout en tenant compte des problèmes posés par les allomorphes (10.500 lemmes ont déjà été traités, soit 30,99% de l'ensemble)<sup>49</sup>. Le D.A.G. devient ainsi un dictionnaire dérivationnel interrogeable sur base des morphèmes (thèmes, suffixes, préverbes, préfixes, radicaux) constitutifs des lemmes. Dans la foulée de ce nouveau projet, les lemmes ont également été munis d'une caractérisation de leur classe morpho-syntaxique (30.826 lemmes ont été traités, soit 91% de l'ensemble)<sup>50</sup>. Enfin, certaines catégories de lemmes (les mots rares, les formes supplétives, les lemmes défectifs) ont été marquées d'indices spéciaux pouvant également intervenir dans les requêtes formulées par un utilisateur.

**§10.** Enrichi au fil des analyses successives, le D.A.G. devient un dictionnaire décrivant l'ensemble du lexique grec indépendamment des genres littéraires (épopée, poésie lyrique, tragédie, comédie, éloquence, etc.), de la nature des [Page 230] sources (sources

<sup>47</sup> ThGrNaz, I, p. XIV.

<sup>48</sup> C'est un des avantages du dictionnaire électronique sur le dictionnaire «papier» traditionnel: un apprenant cherchant le sens de la forme *yeux* dans un dictionnaire, ne trouvera pas ce mot, ni même un renvoi à «œil» (GAUDIN – GUESPIN, *Initiation*, p. 112).

<sup>49</sup> GÉRARD – KINDT, *Du D.A.G. au D.D.G.*

<sup>50</sup> L'analyse flexionnelle est encore, quant à elle, à l'état de projet.



littéraires transmises par la tradition manuscrite, sources documentaires, épigraphiques ou papyrologiques, etc.), ou de leur datation (d'Homère à la fin de l'empire byzantin). Comme le montre le *tableau 3*, le *D.A.G.* affecte une proposition de lemme à plus de 90% des occurrences rencontrées dans les chantiers récents.

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

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

**§11.** Cette estimation optimiste doit cependant être nuancée car le nombre de formes recevant plusieurs propositions de lemme atteint désormais les 10%, ce qui accroît considérablement la tâche du lemmatiseur. De plus, les logiciels utilisés ne reconnaissant que ce qu'ils connaissent, des formes très «communes» peuvent encore manquer au *D.A.G.*, même après de nombreuses analyses : la forme *γίγνομένη* n'y a été introduite qu'à la suite du traitement des textes de Basile de Césarée (*Thesaurus* paru en 2002), les formes *γεγεννημένοι* et *ἀγαγούσης* n'ont été rencontrées pour la première fois que lors de l'analyse des textes de la *Bibliothèque* de Photios (*Thesaurus* prévu en 2004), alors que les lemmes *γίγνομαι* et *ἄγω* regroupaient déjà, respectivement, 238 et 140 formes différentes.

**§12.** Le *D.A.G.* peut être dès à présent appliqué à des sources d'époque classique, car les textes patristiques déjà analysés se caractérisent par un grand nombre d'occurrences, d'un point de vue quantitatif, et, du point de vue qualitatif, par une langue fortement classicisante<sup>51</sup>, empreinte d'un purisme peu perméable aux

<sup>51</sup> À ce propos, cfr, par ex., E. NORDEN, *Die antike Kunstprosa vom VI. Jahrhundert v. Chr. bis in die Zeit der Renaissance*, Leipzig, Berlin, 1915-1918 (réimpr. Stuttgart, 1971), p. 550-572 ; C. FABRICIUS, *Der sprachliche Klassizismus des griechischen Kirchenväter: ein philologisches und geistesgeschichtliches Problem*, dans *J.A.C.*, 10 (1967), p. 187-199 ; *Handbook of Classical Rhetoric in the Hellenistic Period. 330 B.C.-A.D. 400*, ed. S.E. PORTER, Leyden, New York, Cologne, 1997, p. 633-670 (= chapitre 21. *The Greek Christian Writers*, par W. KINZIG).

[Page 231] innovations propres à la langue parlée connues dès l'époque hellénistique<sup>52</sup>. Quand les données lexicales du D.A.G. sont confrontées au vocabulaire du *Contre Ératosthène* de Lysias<sup>53</sup>, 78 formes ne reçoivent pas de proposition de lemme, mais seuls sept lemmes manquent : δυσκολαίνω, είσαρπάζω, Εύφίλητος, Θεσμοφόρια, Ὀῆθεν, ὑποπέμπω et ψιμυθιόω.

§13. Des sources plus récentes ont justifié ou justifieront la création d'entrées telles que Γραικός à côté de Ἑλλήν<sup>54</sup>, κοντάριον à côté de δόρυ, λουλούδι à côté de ἄνθος, ou τέντα à côté de σκηνή, ce dernier lemme produisant par dérivation τεντώω à côté de σκηνώω<sup>55</sup>. L'évolution de la langue, lorsqu'elle produit de nouveaux paradigmes flexionnels, sera perceptible par l'évolution du lexique. Le subst. ψάρι, historiquement issu de ὀψάριον<sup>56</sup>, le nominatif πατέρας (au lieu du classique πατήρ)<sup>57</sup>, ou la particule ἄς, dérivée par syncope de la forme ἄφες (du lemme ἀφήμι)<sup>58</sup>, font l'objet d'entrées distinctes. Ces trois mots «nouveaux» ne relèvent plus du même paradigme flexionnel et morpho-syntaxique que ceux dont ils sont issus. Mais étant formelle, la lemmatisation ne tire pas parti de la syntaxe : les participes είσελθών et συνλαβών, qui déterminent la forme γυνή dans un papyrus du IV<sup>e</sup> s. ap. J.-C., seront lemmatisés sous είσέρχομαι et συλλαμβάνω<sup>59</sup> ; dans l'expression ήμείς βλέποντα, le participe invariable βλέποντα conservera le lemme βλέπω ; la forme πάντων restera classée sous le lemme πᾶς même si, *in textu*, elle détermine un subst. féminin, comme dans l'expression ἐκ πάντων τῶν ἐν τῇ Ἰνδία γυναικῶν<sup>60</sup>.

<sup>52</sup> Sur cette notion, cfr BROWNING, *Greek*, p. 49-50.

<sup>53</sup> Cette analyse, réalisée à titre expérimental, est basée sur le texte de l'édition d'U. Albin (*Lisia, I discorsi*, ed. U. ALBINI, Florence, 1955, p. 6-16).

<sup>54</sup> ThConcOec, p. XIII.

<sup>55</sup> Pour les formes τέντας et τεντώσας, cfr par ex. les §16 et 17 de l'*Histoire byzantine* de Doukas (XV<sup>e</sup> s.) (= *Ducæ Michaelis Ducæ Nepotis Historia Byzantina*, ed. I. BEKKER [C.S.H.B.], Bonn, 1834, p. 17, l. 13 et p. 72, l. 20).

<sup>56</sup> C 846, s.v. ὄψον.

<sup>57</sup> Cfr ὁ πατέρας dans la Recensio F de l'*Historia Alexandri Magni*, 34 (5) (= *Ps.-Kallisthenes: Zwei mittellgriechische Prosa-Fassungen des Alexanderromans*. Teil I. ed. A. LOLOS [Beiträge zur Klassischen Philologie, 141], Königstein/Ts., 1983, p. 152, l. 21).

<sup>58</sup> MANDILARAS, *Verb*, p. 291, §680 (2).

<sup>59</sup> MANDILARAS, *Verb*, p. 335, §877.

<sup>60</sup> Cfr, au III<sup>e</sup> s. ap. J.-C., les *Actes de Thomas*, 116 (= *Acta Apostolorum Apocrypha*. Volumen I. Pars II. *Acta Philippi et Acta Thomae accedunt Acta Barnabae*, ed. M. BONNET, Leipzig, 1903 [réimpr. Hildesheim, 1972], p. 226, l. 7-8) ; ces illustrations s'inspirent de celles fournies dans BROWNING, *Greek*, p. 57, 59 et 64.

**§14.** Les principes de lemmatisation contribuent à normaliser la formulation des lemmes, ce qui revient, dans le cas du grec, à classiciser la graphie des «entrées lexicales» du dictionnaire. C'est la raison pour laquelle une lecture correcte [Page 232] des données lexicales doit aborder tant la macrostructure (l'ensemble des lemmes relevés en langue) que la microstructure (l'ensemble constitué des formes attestées en discours, et de toutes les informations qui y sont attachées).

**§15.** À côté des dictionnaires généraux (ceux de H. Estienne, d'A. Bailly, de H.G. Liddell et R. Scott, et de F.R. Adrados), des ouvrages spécialisés offrent aux chercheurs le vocabulaire grec des époques romaine et byzantine. Les outils édités jadis par E.A. Sophocles, plus récemment par E. Trapp<sup>61</sup>, restreignent leur nomenclature sur base de la date d'attestation des mots, et établissent une liste, plus ou moins complète selon les sources dépouillées, des vocables connus à ces époques. Ils reconstituent ainsi «une image du lexique byzantin», mais non celle «du lexique des Byzantins». Ce dernier est un ensemble composite, formé de vocables attestés depuis l'époque classique, de termes nouveaux (dérivés du grec ou empruntés) et, immanquablement, de mots classiques non attestés dans les sources conservées et qui n'apparaissent que dans les témoignages postérieurs réputés tardifs. Appliqué à un *corpus* très large, le *D.A.G.* contribuera, au fil des analyses, à décrire plus nettement ce «lexique des Byzantins». «Aucun dictionnaire existant n'est fondé sur une vue aussi large de la langue grecque, seule perspective permettant d'éviter, dans le traitement de textes d'époques et de genres différents, des incohérences qui empêcheraient leur comparaison»<sup>62</sup>. À terme, le *D.A.G.* permettra «d'évaluer l'ensemble des sources grecques à la même aune»<sup>63</sup>.

### 3. L'enquête lexicographique

**§16.** La formulation du lemme repose sur une *enquête lexicographique* qui suggère une méthode d'approche de l'«ensemble lexical de référence»<sup>64</sup>, c'est-à-dire l'ensemble des sources disponibles et

<sup>61</sup> *Lexikon zur byzantinischen Gräzität besonders des 9.-12. Jahrhunderts*, ed. E. TRAPP, Vienne, depuis 1994.

<sup>62</sup> ThGrNaz, I, p. XIII.

<sup>63</sup> B. KINDT, recension du ThTheoConf dans *Le Muséon*, 111 (1998), p. 245.

<sup>64</sup> Pour reprendre les termes utilisés par É. Famerie dans FAMERIE, *Appien*, p. 255-259.

accessibles, ce qui constitue la partie encore visible du *discours*. Dans le cadre du *Projet*, cette enquête est organisée en quatre volets :

- la consultation des dictionnaires «papier» traditionnels ;
- l'examen des sources ;
- l'utilisation des ressources électroniques ;
- l'application des *principes d'économie* et de *fidélité au texte de l'édition*.

[Page 233]

Ces investigations permettent de rassembler les informations utiles non seulement à la formulation du lemme, mais aussi à la rédaction de notices précisant :

- l'inventaire des graphies ;
- la motivation du choix du lemme ;
- la justification des corrections apportées aux éditions ;
- les remarques sur la fréquence ou l'usage des formes ;
- des *realia*, etc.<sup>65</sup>

**§17.** L'élaboration du dictionnaire électronique ne peut se passer de la consultation des dictionnaires «papier» traditionnels ; cette première étape de l'enquête lexicographique fournit des informations de premier ordre à propos du mot analysé, à savoir :

- la ou le(s) graphie(s) rencontrée(s) dans les sources ;
- le ou les sens enregistré(s) ;
- les dégroupements effectués pour distinguer les entrées lexicales homographes : pour la forme *διώνυμον*, par exemple, ce premier examen amène le lemmatiseur à poser deux entrées lexicales afin de distinguer *διώνυμος (διά)* «dont le nom se répand au loin», de *διώνυμος (δίς)* «avec deux noms» (cfr B 526, s.v. 2 *δι-ώνυμος* / 1 *δι-ώνυμος* ; LS 441, *διώνυμος I* et *διώνυμος II*)<sup>66</sup>. Aucun lemmatiseur, aussi compétent soit-il, ne maîtrise en permanence l'ensemble de ces connaissances pour les mots qu'il est amené à analyser<sup>67</sup>.

<sup>65</sup> Ces notices sont publiées dans les *Index des lemmes particuliers des Thesauri*, cfr ThTheoConf, I, p. XVIII-XXVI ; ThProcCaes, p. XV et XXIV-XXIX ; ThAstFirm, p. XVI-XIX ; ThBasCaes, I, p. XXXV-L.

<sup>66</sup> Les références aux dictionnaires sont réduites à la première lettre du nom ou des auteurs, cfr les *Abréviations bibliographiques* en début d'article.

<sup>67</sup> Cfr COULIE, *Lemmatization*, p. 36 et 46-47 ; cfr aussi ci-dessous, **§37**.

L'enquête doit impérativement s'étendre à plusieurs ouvrages car, d'un dictionnaire à l'autre, la représentation fournie des multiples sens attestés, des différentes graphies rencontrées, et des dégroupements présente des différences parfois sensibles, ainsi :

- pour le mot Κάμπανος, par ex., la situation est la suivante :
- B 1015, une seule entrée Καμπανοί, ὧν (οἱ), «les Campaniens» ;
- LS 873, une seule entrée καμπανός, ὄ, «weighing-machine» ;
- PB 613, une entrée Καμπανία sous laquelle apparaît Καμπανοί, et une entrée Καμπανός *Eigenn.* (*sic*).

Le *D.A.G.* enregistre trois entrées différentes : καμπανός, pour «weighing-machine» ; Καμπανός (ὄ), pour «Campanus » (anthr.) ; Καμπανός (Καμπανία), pour «Campanien, de Campanie» (ethn.), les sources attestant une «forme de départ» pour chacun de ces trois lemmes.

- le *Dictionnaire Grec Français* et le *Greek-English Lexicon* fournissent deux représentations lexicographiques sensiblement différentes de l'entrée ἀλέκτωρ :

[Page 234]

- B 74	1 ἀλέκτωρ [...] 2 ἀλέκτωρ [...]	1 (ἀ priv.) «vierge» 2 (ἀ cop.) «éroux» «coq» [...] ( <i>cf</i> ἀλέκω?)
- LS 62	ἀλέκτωρ (A) [...] ἀλέκτωρ (B) [...]	I. «cock» II. «husband» (ἀ priv., λέγω) = ἄλεκτρος

Aucun des deux dictionnaires ne justifie le regroupement du mot ἀλέκτωρ «éroux» avec ἀλέκτωρ «vierge» ou avec ἀλέκτωρ «coq». Le *D.A.G.*, sur base des principes formels, enregistre trois lemmes distincts : ἀλέκτωρ (χωρίς) pour ἀλέκτωρ «vierge», ἀλέκτωρ (ἄμα) pour ἀλέκτωρ «éroux», et ἀλέκτωρ (ἀλέξω) pour ἀλέκτωρ «coq»<sup>68</sup>.

**§18.** La lemmatisation insère le vocabulaire d'un auteur dans le lexique de la langue. La formulation des nouveaux lemmes doit donc se baser sur un large *examen des sources*, embrassant l'ensemble des manifestations encore disponibles du mot étudié. À l'issue d'un examen limité aux textes de Procope de Césarée, le lemme Βυζακηνός aurait été formulé au pluriel, Βυζακηνοί, comme dans PB 232. Le singulier attesté dans les autres sources justifie pourtant une entrée lexicale au singulier<sup>69</sup>. Cette approche permet aussi de

<sup>68</sup> Description qui rejoint celle fournie dans A I 141.

<sup>69</sup> ThProcCaes, p. X-XI.

réaliser l'inventaire des différentes graphies de certains mots et d'en définir le paradigme. L'enquête menée préalablement à la définition du lemme anthroponymique perse Ζαδέσπρας a permis de voir que ce nom présentait trois paradigmes flexionnels différents : Ζαδέσπρας (acc. Ζαδέσπραν ; gén. Ζαδέσπρα) ; Ζαδεσπράτης (acc. Ζαδεσπράτην ; gén. Ζαδεσπράτου) ; Ζαδεσπράμ (indéclinable)<sup>70</sup>.

Toutes les manifestations conservées de la langue peuvent être prises en compte. En plus des sources littéraires, l'analyse doit s'ouvrir aux sources hagiographiques<sup>71</sup> ou juridiques<sup>72</sup>, aux commentaires d'auteurs ou aux scholies<sup>73</sup>, aux documents épigraphiques (qu'ils relèvent de l'*instrumentum publicum*<sup>74</sup> ou de l'*instrumentum domesticum*<sup>75</sup>), papyrologiques<sup>76</sup>, voire numismatiques<sup>77</sup>.

**§19.** L'examen des sources doit tirer profit de l'utilisation des ressources électroniques (bases textuelles, analyseurs lexicaux et grammaticaux) désormais disponibles :

<sup>70</sup> JUSTI, *Namenbuch*, p. 384, s.v. Zātsparham.

<sup>71</sup> Le lemme Σίβερις (fleuve) est déduit des formes Sivberin de Procope de Césarée et des formes Σίβερην et Σιβέρεως issues d'une source hagiographique (ThProcCaes, p. X).

<sup>72</sup> Le lemme piskinhv est déduit des formes piskinaiv de Procope de Césarée et des formes πισκινή d'Hippolyte et πισκίνης des *Basiliques*, une source juridique (LVIII, 21, *Περὶ πηγῆς* = *Basilicorum Libri LX. Series A. Volumen VII. Textus Librorum LIII-LIX*, ed. H.J. SCHELTEMA, N. VAN DER WAL [*Scripta Universitatis Groninganae*], Groningen, 1974, p. 2701, l. 3).

<sup>73</sup> Le singulier Δάκης, pour l'ethnique des Daces, apparaît dans une scholie des *Gûêpes* d'Aristophane (ThProcCaes, p. X, note 29).

<sup>74</sup> La «forme de départ» du lemme ἰχθυόβρωτος apparaît dans une inscription de Smyrne (G. DITTENBERGER, *Sylloge Inscriptionum Graecarum*, vol. III, Leipzig, 1920 [réimpr. Hildesheim, 1960], p. 129, n° 997, l. 7.) ; celle du lemme ἐπιδημιουργός, enregistré au pluriel dans les dictionnaires car les sources littéraires ne l'attestent qu'au pluriel, est déduite de la forme ἐπιδαμιουργόν (*sic*) d'une inscription dorienne (ThProcCaes, p. X) ; le nominatif du lemme Ἀριστόβρωτος apparaît sur une inscription de Rhodes (V. KONTORINI, *Inscriptions inédites de Rhodes*, II, Athènes, 1989 [en grec avec titre et résumé en français], p. 48, G, l. 6) ; le lemme Σίλουινος est déduit des formes Σιλουῖνον et Σιλβίνου lues sur des inscriptions de Termessos et de Scythopolis (ThBasCaes, I, p. XLVII).

<sup>75</sup> Le nominatif du lemme anthroponymique Ὑψις apparaît sur l'inscription d'un vase grec du VI<sup>e</sup> s. av. J.-C. (ThBasCaes, I, p. XLIX).

<sup>76</sup> La forme de départ du lemme ἔλεφαντηγός, lue chez Photios, apparaît sur un papyrus de la collection Petri (J.P. MAHAFFY, *On the Flinders Petrie Papyri. Transcriptions, Commentaries, and Index*, Part. II (Royal Irish Academy. *Cunningham Memoirs*, 9), Dublin, 1893, p. 135, n°40, a, l. 26).

<sup>77</sup> Le nominatif du lemme toponymique Ἄνισα apparaît sur une légende monétaire (ThBasCaes, I, XXXVI).

- le CD ROM #E du *T.L.G.*<sup>78</sup> ;
- le CD ROM #7 du *P.H.I., Greek Documentary Texts*<sup>79</sup> ;
- les ressources du *Greek Morphological Analysis* du *Perseus Project*<sup>80</sup>.

Ces réalisations font désormais partie des outils de consultation, au même titre que les dictionnaires «papier» traditionnels. Elles partagent cependant les mêmes défauts : elles ne sont pas exhaustives et sont conçues sur des *corpus* différents, à partir de présupposés et avec des objectifs qui leur sont propres. Les lemmes Ὑψις, Σίβερις et ἐπιδημιουργός (cités dans les notes précédentes) ne figurent pas dans la base textuelle du *T.L.G.* Ces ressources n'étant pas lemmatisées, une interrogation formulée sous la forme <σερουι> visant à obtenir les occurrences du mot Σέρουιος (l'anthroponyme latin *Servius*), ne fournira pas les occurrences orthographiées avec -β- au lieu de -ου- (du type Σέρβιος), ni la forme Ἐρούιου rencontrée chez Photios dans l'expression « Ἐρούιου [Page 236] Τυλλίου». Quant à l'analyseur morphologique du *Perseus Project*, il a le *Greek-English Lexicon* pour base lexicale. Une interrogation sur Κιθαρίζων, un toponyme arménien attesté chez Procope de Césarée, fournit une analyse relative au v. καθαρίζω ; le *Greek-English Lexicon* prend en compte le vocabulaire de Procope de Césarée, mais n'enregistre pas systématiquement les noms propres<sup>81</sup>.

#### 4. Les principes d'économie et de fidélité

§20. Toutes les occurrences d'un *corpus* analysé sont censées recevoir un lemme. Les principes d'économie et de fidélité correspondent au besoin d'apporter une réponse pragmatique à certaines difficultés, tout en respectant les principes de lemmatisation ainsi que le choix des éditeurs, sans préjuger du caractère productif ou systématisable du lemme<sup>82</sup>. Le premier principe décrit évite de créer un lemme pour une forme exceptionnelle. Le second autorise la création d'un lemme

<sup>78</sup> M. PANTELIA, "Noûs into Chaos": the Creation of the Thesaurus of the Greek Language, dans *International Journal of Lexicography*, 13 (2001), p. 1-11 ; cfr aussi l'adresse <http://tlg.uci.edu>.

<sup>79</sup> Cfr <http://132.236.125.30/default.htm>.

<sup>80</sup> Cfr <http://www.perseus.tufts.edu>.

<sup>81</sup> ThProcCaes, p. LXI ; le texte de l'historien, ἐν φρουρίῳ, ὅπερ Κιθαρίζων καλοῦσιν, ne présente par ailleurs aucune ambiguïté.

<sup>82</sup> Dans ces lignes et celles qui suivent, l'adjectif «système» est utilisé au sens fort du terme indiquant la possibilité de faire entrer une unité dans le «système de la langue».

ou d'un «lemme-forme» sans remonter à une «forme de départ» *stricto sensu*. Dans les deux cas, les notices lexicologiques publiées dans les *Thesauri* motivent les options prises pour chaque lemme. Si des éléments nouveaux complètent une analyse antérieure, il est toujours possible de revoir la formulation d'un lemme lors d'une mise à jour du *D.A.G.*

**§21.** Le *principe d'économie* permet de classer sous un seul lemme des formes présentant des graphies ou des formes dérivationnelles ou flexionnelles exceptionnelles, souvent uniques, que le lemmatiseur ne considère pas comme productives :

- στομαχῶντες *s.l.* στομαχέω (et non \*στομαχάω, inconnu par ailleurs ; cfr *ThBasCaes*, I, p. XLVII.) ;
- δυσμενῆ *s.l.* δυσμενέων (et non δυσμενέω, inconnu par ailleurs ; cfr *ThAstFirm*, p. XVI) ;
- Μελιττέων *s.l.* Μελιτεύς (cfr *ThAstFirm*, p. XVII) ;
- Ἑλληνοποντίων *s.l.* Ἑλλησπόντιος (et non \*Ἑλληνοπόντιος, inconnu par ailleurs ; cfr *ThBasCaes*, p. XXXIX) ;
- Νουμίδιων *s.l.* Νουμίδης (et non \*Νουμίδιος, inconnu par ailleurs ; cfr *ThProcCaes*, p. XXVII), etc.

Ce *principe d'économie* sera aussi invoqué pour définir les principes de lemmatisation relatifs aux mots défectifs ou aux mots dont le paradigme fait appel à des formes supplétives (cfr **§98-107**).

[Page 237]

**§22.** Le *principe de fidélité à l'édition* permet de créer un lemme pour une forme exceptionnelle car «[les éditeurs] gardent en définitive la responsabilité des textes qu'ils ont fournis»<sup>83</sup> :

- chez Astérius (*ThAstFirm*, p. XVII), la forme ἐνευλογισθήσονται a reçu pour «forme de départ» un lemme ἐνευλογίζομαι, non attesté, mais défini sur base de l'existence d'une série productive de *v.* composés comprenant les éléments ἐν- ou εὐ- à l'initiale et λογίζομαι comme second élément (εὐλογίζεσθαι et ἐνευλογεῖσθαι étant attestés dans les sources) ;
- μοιχιάϊαν *s.l.* μοιχιαῖος ; la forme *in textu*, chez Basile de Césarée (*ThBasCaes*, p. XLIII-XLIV), pourrait justifier une correction en μοιχιδιάν. Pour être régulier, le lemme μοιχιαῖος devrait dériver

<sup>83</sup> TOMBEUR, *Thesaurus*, I, p. 19 ; *ThGrNaz*, I, p. IX.



d'un substantif \*μοιχία absent des dictionnaires (comme βία → βίαιος)<sup>84</sup> ;

- chez Photios, une forme anthroponymique Δόκιον, apparaissant dans l'expression «πρὸς Δόκιον», est lemmatisée Δόκιος, bien qu'il puisse s'agir soit d'une erreur ou d'un diminutif (forme hypocoristique) pour Προσδόκιον.

Les *lemmes-formes*, conservant le texte de l'édition, relèvent de ce principe (cfr §93, 3) :

- **ἀνακαινήσασαν** *s.l.* ἀνακαινήσασαν, lemme-forme, car aucun v. du type \*καινέω n'est attesté dans les sources. La correction ἀνακαινίσασαν aurait été possible (ThBasCaes, p. XXXV). Mais le lemmatiseur n'est pas l'éditeur, et le *principe de fidélité à l'édition* permet de conserver la forme controversée ;

- ὠρθρῶσαν *s.l.* ὠρθρῶσαν, lemme-forme, l'entrée proposée dans LS 1250 renvoyant à une forme unique et mal attestée (ThBasCaes, p. L).

<sup>84</sup> D'autres exemples de graphies en -ίδιος vs -ιαῖος existent. En face des formes ἐπινυκτίδιος et προμετωπίδιος enregistrées dans le D.A.G., une interrogation du T.L.G. fournit νυκτιαῖος et περιμετωπιαῖος. L'entrée κατανωτίδιος du *Greek-English Lexicon* (LS 903) est accompagnée de la remarque «v.l. -ιαῖος».

## II. PRINCIPES FORMELS DE LEMMATISATION (§23-33)

1. Variations formelles justifiant la création d'un lemme (§23-27). 2. Variations formelles ne justifiant pas la création d'un lemme (§28-33).

### 1. Variations formelles justifiant la création d'un lemme

**§23.** La lemmatisation se fonde d'abord sur une *analyse formelle* : toutes les formes issue d'un même mot sont classées sous un même lemme. Tel qu'il est actuellement conçu, le programme de reconnaissance des formes se base sur les mots du *corpus analysé*, ces mots étant définis comme autant de «séquences de [Page 238] caractères séparées par des blancs». Opérant mot à mot, il ne tient pas compte de l'environnement dont la forme est extraite, il ne résout pas les ambiguïtés (quand une forme fait l'objet de plusieurs propositions de lemme) et n'analyse pas comme un tout les locutions constituées de plusieurs mots.

**§24.** Le lemme est formulé en ionien-attique classique – dialecte servant de norme –, au nom. sing. pour les formes relevant de la flexion nominale, adjectivale ou pronominale, à l'act. ind. prés. 1<sup>ère</sup> pers. sing. pour les formes relevant de la flexion verbale. Les *variations phonétiques, accentuelles et dialectales*, neutralisées, ne justifient pas la création d'un lemme distinct :

- les voyelles consécutives ne sont pas contractées : ἀγαθουργίαν *s.l.* ἀγαθοεργία ; ἀνάπλων *s.l.* ἀνάπλοος (et non ἀνάπλους) ; ὕϊδοῦν *s.l.* ὕϊδέος (et non ὕϊδοῦς) ; ὄρᾱσθε *s.l.* ὄράω (et non ὄρῶ) ; les adjectifs en -κλῆς ont un lemme oxyton en -κλειής (δυσκλειής), les anthroponymes un lemme paroxyton en -κλέης (Περικλέης)<sup>85</sup> ;

- les voyelles en *hiatus* sont surmontées d'un tréma : κακοποιία *s.l.* κακοποιΐα ;

- les consonnes consécutives, en frontière de morphèmes<sup>86</sup>, sont assimilées : ἐνεχείρισα *s.l.* ἐγχειρίζω ; κατόψεται *s.l.* καθοράω ; συνεκινεῖτο *s.l.* συγκινέω ; συνελήφθη *s.l.* συλλαμβάνω ; κρύβδην ; etc.

<sup>85</sup> Ces substantifs et adjectifs contractes ont pour lemme une forme non contractée même s'il s'agit d'un graphie caractérisée comme «poétique», «érique» ou «ionienne» dans les dictionnaires (par ex. Ἴεροκλέης dans B 960).

<sup>86</sup> Les graphies du type χερρο- issues de χέρσος ou de ses composés, constituant un atticisme, ne sont donc pas assimilées au niveau des lemmes, cfr B 2135 ; C 1255, s.v. χέρσος.

**§25.** Des mots proches les uns des autres mais mettant en œuvre des procédés dérivationnels différents font l'objet d'entrées distinctes (de tels cas apparaissent dans toutes les classes morpho-syntaxiques) :

- *subst.* : ἀντέρεισμα vs ἀντέρεισις, ὀλολυγή vs ὀλόλυγμα vs ὀλολυγμός<sup>87</sup> ; παροδίτης vs πάροδος (ὀ) ;

- *anthr.* : Τατζάτιος vs Τατζάτης, Ἄρτασήρας vs Ἄρτασήριος, Ἑκατόμνωσ vs Ἑκάτομνος ;

- *surnom* : Τριφύλης vs Τριφύλιος ;

- *ethn.* : Πεπουζηνοί vs Πεπουζιανοί vs Πεπουζίται<sup>88</sup> ;

- *top.* : Γράσση vs Χρᾶσις ; Νισίβη vs Νίσιβις ; Σαρδῶ vs Σαρδανία vs Σάρδος<sup>89</sup> ; Τομέα vs Τόμις ; Τύανος vs Τύανα ;

- *adj.* : ἄσωμος vs ἀσώματος ; βλακώδης vs βλακικός ; ἴσαῖος vs ἴσος ; Εὐτυχηταί vs Εὐτυχιανός vs Εὐτυχιανιστής vs Εὐτυχιανικός ;

[Page 239]

- *adv.* : ἀστενάκτως vs ἀστενακτί ;

- *v.* : ἀνταιχμαλωτίζω vs ἀνταιχμαλωτεύω ; στερέω vs στερίσκω.

**§26.** Deux lemmes s'imposent aussi quand une forme déclinable s'oppose à une forme indéclinable<sup>90</sup> : ἀββᾶς vs ἀββᾶ, πάππας vs πάππα, ῥαββίς vs ῥαββί (pour des *subst.*) ; Δανίηλος vs Δανίηλ, Ἰάκωβος vs Ἰακώβ, Ἰσαμος vs Ἰσαμ, Καργαρίγας vs Καρδαρηγάν, Οὔμαρος vs Οὔμαρ<sup>91</sup>, Ζαδέσπρας vs Ζαδεσπράτης vs Ζαδεσπράμ (pour des *anthr.*) ; Ἀργάριζος vs Ἀργαρίζιν, Ἱεροσόλυμα vs Ἱερουσαλήμ (pour des *top.*).

**§27.** La création d'un nouveau lemme est également justifiée quand l'évolution sémantique d'un mot est accompagnée d'un changement du paradigme morpho-syntaxique : ἄνθος (ὀ) «bruant» (oiseau) vs ἄνθος (τό) «ce qui croît, ce qui pousse, une pousse [...], fleur»<sup>92</sup>.

<sup>87</sup> Les entrées ὀλολυγή et ὀλολυγμός sont regroupées dans E VI 1902.

<sup>88</sup> ThBasCaes, I, p. XLV.

<sup>89</sup> ThTheoConf, I, p. XV.

<sup>90</sup> ThTheoConf, I, p. XI.

<sup>91</sup> ThTheoConf, I, p. XV.

<sup>92</sup> Cfr B 162, s.v. 1 ἄνθος et <sup>2</sup> ἄνθος et LS 140, s.v. ἄνθος (A) et (B), mais C 89, s.v. ἄνθος.

## 2. Variations formelles ne justifiant pas la création d'un lemme

Variations dialectales (§29). Variations diachroniques (§30). Variations phonétiques (§31). Variations accentuelles (§32). Variations orthographiques (§33).

§28. En tant que «forme de départ» et «forme neutralisée du discours métalinguistique», le lemme est exempt des variations formelles propres aux mots en discours. Les marques casuelles, modales et temporelles n'apparaissent pas, les marques de genre et de nombre sont effacées. Un lemme unique regroupe les formes marquées d'une variation dialectale (propre à une région ou à un genre littéraire), diachronique, phonétique, accentuelle ou orthographique. Les lignes qui suivent en établissent un inventaire non exhaustif, certes (il ne s'agit pas d'écrire une grammaire), mais basé sur les formes et lemmes effectivement rencontrées dans les chantiers traités et enregistrés dans le D.A.G. Le classement adopté correspond à la nature des informations fournies par les dictionnaires et les grammaires pour justifier ces variations.

§29. Le lemme revêt une forme exempte de *variation dialectale*. Peuvent être cités et illustrés comme ne justifiant pas la création d'un nouveau lemme :

### 1. Des atticismes :

- la particule démonstrative -ί<sup>93</sup> : μονονουχί s.l. μονονού ; ούτοσί et τουτονί s.l. οὔτος ; οὐχί s.l. οὐ ; τοιονδί s.l. τοιόσδε ; ὠδί s.l. ὦδε

[Page 240]

- la graphie -ττω pour -ζω : ἀρμόττειν s.l. ἀρμόζω<sup>94</sup> ; σφαττόμενοι et ἐπικατασφάττει s.l. σφάζω<sup>95</sup> et ἐπικατασφάζω.

- la graphie -ττ- pour -σσ-<sup>96</sup> : θάλαττα s.l. θάλασσα, ἐφύλαττε s.l. φυλάσσω ;

<sup>93</sup> Sur cette appellation, cfr CHANTRAINE, *Morphologie*, p. 126, §135 et LEJEUNE, *Phonétique*, p. 317, §364 ; «déictique» dans ThTheoConf, I, p. XI ; «paragogicum» dans JANNARIS, *Grammar*, p. 161, §574.

<sup>94</sup> Forme présentée comme «attique c. ἀρμόζω» dans B 271.

<sup>95</sup> Forme présentée comme «néo-att.» dans B 1879 ; JANNARIS, *Grammar*, p. 223, §873. Le *Dictionnaire Grec Français* enregistre deux entrées pour ἀρμόζω et ἀρμόττω mais une seule pour σφάζω. Dans le *Greek-English Lexicon*, les deux entrées sont enregistrées pour les deux verbes, mais celles en -ττω (cfr respectivement LS 243 et 1737-1738) ne contiennent qu'un renvoi vers celles en -ζω (cfr respectivement LS 244 et 1739).

- l'assimilation de -ρσ- en -ρρ-<sup>97</sup> : ἄρρενα s.l. ἄρσην, θάρρει s.l. θάρσος ;  
μυρρίναις s.l. μυρσίνη ; χερρονήσου s.l. χερσόνησος ;

- les formes relevant de la déclinaison attique<sup>98</sup> : νεώς s.l. ναός ; λεώς  
s.l. λαός (et ses composés, par ex. Μενέλεως s.l. Μενέλαος) ; sur la  
métathèse de quantité, cfr §31, 1 ;

- autres cas rencontrés : ἄττα s.l. τις ; ὄστρειών s.l. ὄστρεον<sup>99</sup> ;  
πισύρων s.l. Ν-τέσσαρες ; σεύτλω s.l. τεῦτλον et σεύτλιον s.l.  
τεύτλιον<sup>100</sup> ; τήμερον s.l. σήμερον<sup>101</sup> ; σμύραινα s.l. μύραινα ; σμικρός  
s.l. μικρός<sup>102</sup> .

## 2. Des ionismes :

- la transformation d'un *alpha* long en η après certaines voyelles ou  
*rhô* : δυομέναν s.l. δύω (δύνω) ; σπείρημα s.l. σπείραμα ; πέτρα s.l.  
πέτρα ;

- la psilose<sup>103</sup> : ἀνδανέεις s.l. ἀνδανάω.

- l'allongement<sup>104</sup> : μονογενοῦς s.l. μονογενής ; μοῦνος s.l. μόνος ;  
οὔλον (sic, avec psilose) s.l. ὄλος.

- la graphie ξυν- pour συν-<sup>105</sup> : ξυγγελῶν s.l. συγγελάω ; ξυμμαχία s.l.  
συμμαχία ; ξυναγούσης s.l. συνάγω ;

[Page 241]

- la graphie ἐσ- pour εἰσ- : ἐσβολή s.l. εἰσβολή, ἐσβάλλειν s.l.  
εἰσβάλλω<sup>106</sup> ;

<sup>96</sup> Le lemme en -σσ- est créé si cette seule forme est attestée, cfr ThGrNaz, I, p. XV. LEJEUNE, *Phonétique*, p. 105, §94: divergences dans le traitement des groupes composés d'une dentale ou d'une gutturale et de la semi-voyelle \*-y-.

<sup>97</sup> LEJEUNE, *Phonétique*, p. 125, §119: divergences entre l'attique et l'ionien, hésitation dans la κοινή.

<sup>98</sup> CHANTRAINE, *Morphologie*, p. 43-45, §25 ; JANNARIS, *Grammar*, p. 118, §323-326

<sup>99</sup> E VI 2318 ; LS 1264.

<sup>100</sup> ThBasCaes, I, p. XLVIII.

<sup>101</sup> LEJEUNE, *Phonétique*, p. 109, §100.

<sup>102</sup> LEJEUNE, *Phonétique*, p. 120, §113: \*sm- ancien conservé à l'initiale de quelques termes. Mais l'anthroponyme reste Σμῖκρος.

<sup>103</sup> LEJEUNE, *Phonétique*, p. 92, §82.

<sup>104</sup> LEJEUNE, *Phonétique*, p. 158-159, §159: «effacement de w» et «allongement compensatoire d'une voyelle brève qui précède».

<sup>105</sup> ThGrNaz, I, p. XV ; LEJEUNE, *Phonétique*, p. 73, §61.

<sup>106</sup> ThTheoConf, I, p. XI (subst. et adj.) et XIV (v.) ; LEJEUNE, *Phonétique*, p. 131, §125.

- les formes d'impft et d'aor. thématique en -σκ-ον ne créant pas une nouvelle série productive, ne justifient pas la création d'un lemme<sup>107</sup> : ζώεσκες s.l. ζάω ; ποθέεσκες s.l. ποθέω ; ὠδίνεσκε s.l. ὠδίνω ;

- autres cas rencontrés : μαργαρίδαι s.l. μαργαρίτης<sup>108</sup>.

3. Des dorismes :

- Ζανός et Ζήνα s.l. Ζεύς<sup>109</sup> ;

- Ἄγεμωνία s.l. ἡγεμονία ; σάμερον s.l. σήμερον<sup>110</sup> ;

- κάρρων (= κρείσσων) s.l. ἀγαθός<sup>111</sup>.

4. Des formes «ériques» ou «ροέtiques» :

- gén. en -αο<sup>112</sup> : Ἐπτηνορίδαο s.l. Ἐπτηνορίδης ; Κρονίδαο s.l. Κρονίδης ;

- gén. en -οιο<sup>113</sup> : ὕμνοιο s.l. ὕμνος ; ἀμωμήτοιο s.l. ἀμώμητος ;

- des cas de métathèse de quantité, cfr §31, 1 ;

- des cas de métathèse consonantique : πόρσω s.l. πρόσω ;

- des cas de flexions avec alternance de degré du radical<sup>114</sup> : ἀνερας s.l. ἀνήρ ;

- nom. pl. en -εες : τοκέες s.l. τοκεύς ;

- autres cas rencontrés : ἀλεγεινήν s.l. ἀλγεινός<sup>115</sup> ; ἐνί s.l. ἐν<sup>116</sup> ; πτολεμίζων s.l. πολεμίζω ; πτολέμοιο s.l. πόλεμος.

5. Remarque : les mots propres à un seul dialecte auront cependant un lemme propre, même si leur forme est explicable par la linguistique historique : βανά (béotien) pour γυνή<sup>117</sup> ; δάνος (ὀ) (macédonien) pour θάνατος<sup>118</sup> ; στίλπων (mot sybarite)<sup>119</sup>.

<sup>107</sup> SCHWYZER, *Grammatik*, I, p. 710-712 ; CHANTRAINE, *Morphologie*, p. 226-227, §261-262.

<sup>108</sup> B 1226 ; LS 1080.

<sup>109</sup> Cfr B 882, s.v. \*Ζήν.

<sup>110</sup> Cfr B 1744 ; LEJEUNE, *Phonétique*, p. 109, §100.

<sup>111</sup> Sur le traitement des formes comparatives supplétives, cfr §98-107.

<sup>112</sup> CHANTRAINE, *Morphologie*, p. 55, §44.

<sup>113</sup> CHANTRAINE, *Morphologie*, p. 37-38, §15.

<sup>114</sup> CHANTRAINE, *Morphologie*, p. 78, §74.

<sup>115</sup> C 55, s.v. ἄλγος : ἀλγεινός doublet homérique pour ἀλγεινός.

<sup>116</sup> B 664.

<sup>117</sup> B 346 ; LS 305 ; C 242, s.v. γυνή.

<sup>118</sup> B 430 ; LS 369.

§30. Le lemme revêt une forme exempte de *variation diachronique*.

1. Des formes «récentes» ou «tardives» :

- les formes verbales tardives et qui n'évoluent pas vers un paradigme productif<sup>120</sup> : εἶπας s.l. λέγω<sup>121</sup> ; ἦλθα s.l. ἔρχομαι ; ἐξελαθήσεται s.l. ἐξελαύνω<sup>122</sup> ; ἐπελέγη s.l. ἐπιλέγω<sup>123</sup> ; συναγηχώς s.l. συνάγω<sup>124</sup> ;

- sur la labilité du *iota* souscrit, cfr §33, 1.

- autres cas rencontrés : γαννύμενος s.l. γάννυμαι<sup>125</sup> ; γνόφος s.l. δνόφος<sup>126</sup> ; ἐξουθενεῖν s.l. ἐξουδενέω<sup>127</sup> ; καθαριωτέρω s.l. καθάρειος<sup>128</sup> ; κλίβανον s.l. κρίβανον<sup>129</sup> ; μετάλημψις s.l. μετάληψις<sup>130</sup> ; σιέλου s.l. σιάλον<sup>131</sup> ; ὕελον s.l. ὕαλος<sup>132</sup> .

2. Des formes byzantines :

- les graphies en -αρέα pour -αρία, dérivées de mots en -(ά)ριος<sup>133</sup> : κουβικουλαρέαν s.l. κουβικουλάριος<sup>134</sup> ;

<sup>119</sup> B 1794 ; LS 1646.

<sup>120</sup> Cfr aussi ThTheoConf, I, p. XIV.

<sup>121</sup> Cfr MANDILARAS, *Verb*, p. 149, §317 (3).

<sup>122</sup> JANNARIS, p. 260, §996, n°75.

<sup>123</sup> JANNARIS, p. 267, §996, n°151, 2.

<sup>124</sup> B 1835.

<sup>125</sup> B 388 ; LS 338.

<sup>126</sup> ThGrNaz, I, p. XIII ; ThTheoConf, I, p. XI ; C 290, s.v. δνόφος ; LEJEUNE, *Phonétique*, p. 78, §67.

<sup>127</sup> LS 598 ; C 835, s.v. οὐ.

<sup>128</sup> Cfr B 991.

<sup>129</sup> ThTheoConf, I, p. XI, cfr cependant C 583, s.v. κρίβανος : «on ignore même si la forme originelle est κρίβανος ou κλίβανος».

<sup>130</sup> ThTheoConf, I, p. XI ; JANNARIS, *Grammar*, p. 95, §193.

<sup>131</sup> C 1000, s.v. σιάλον.

<sup>132</sup> C 1150, s.v. ὕαλος.

<sup>133</sup> PSALTES, *Grammatik*, p. 267-268, §390 ; D. GEORGACAS, *Grammatische und etymologische Miscellen zum Spät- und Neugriechischen*, dans *Glotta*, 31 (1951), p. 213-214 ; sur la productivité du suffixe -άριος (< lat. -arius), cfr JANNARIS, *Grammar*, p. 293-295, §1040-1044.

<sup>134</sup> Sur ce lemme, cfr ThTheoConf, I, p. XI ; G.N. HATZIDAKIS, *Zur Wortbildungslehre des Mittel- und Neugriechischen*, dans *BZ*, 2 (1893), p. 278 ; E. KRIARAS, *Λεξικό τῆς μεσαιωνικῆς Ἑλληνικῆς δημόδου γραμματείας. 1100-1669*, vol. 8, Thessalonique, 1985, p. 313, s.v. κουβικουλαρία ; Suppl. LS 184 ; H. HOFMANN, *Die lateinischen Wörter im Griechischen bis 600 n. Chr.*, Diss. Erlangen-Nürnberg, 1989, p. 205, s.v. κουβικουκάριον, -ία (sic).

- les subst. masc. en -ις<sup>135</sup> ou neutres en -ιν<sup>136</sup> dérivés de mots en -ιος ou -ιον :

[Page 243]

- en -ιος : ἀλεκτόριν *s.l.* ἀλεκτόριος ; μαγγανάριν *s.l.* μαγγανάριος (subst.) ; Ἄρτασῆριν *s.l.* Ἄρτασῆριος et Λοῦκις *s.l.* Λούκιος (anthr.) ;

- en -ιον : ὄψάριν *s.l.* ὄψάριον ; ταβλίν *s.l.* ταβλίον (subst.) ; Ἀμῶριν *s.l.* Ἀμόριον (top.) ;

- évolution de la quantité des voyelles : Ἰοβῖνος, Ἰουβῖνος et Ἰωβίνω *s.l.* Ἰοβῖνος<sup>137</sup> ; Ἰοβίω, Ἰούβιος, Ἰωβίου *s.l.* Ἰόβιος ; Παιονίω *s.l.* Παιώνιος<sup>138</sup> ;

- évolution du timbre des voyelles : Δελματίας *s.l.* Δαλματία<sup>139</sup> ; Κουλλούρια *s.l.* Κολλούριον<sup>140</sup> ; Σερδικήν *s.l.* Σαρδικήν<sup>141</sup> ; τοβερίων *s.l.* τουβερίων<sup>142</sup> ;

- évolution du timbre et de la graphies des consonnes : Νιτσίβιν et Νιτζίβιος *s.l.* Νίσιβις<sup>143</sup> ; Δαβίδ *s.l.* Δαυίδ ;

- disparition du *sigma* initial des mots empruntés au latin : ἔνατον *s.l.* σενᾶτον ; Ἐρούιου *s.l.* Σέρουιος<sup>144</sup> ;

- les formes syncopées ne justifient pas la création d'un nouveau lemme si la forme non syncopée est attestée (cfr aussi §31, 1)<sup>145</sup> : Κάτλος *s.l.* Κάτουλος (pour lat. *Catulus*, surnom des membres de la gens *Lutatia*, cfr G 278) ; βᾶκλων *s.l.* βᾶκυλον mais σταύλων *s.l.*

<sup>135</sup> JANNARIS, *Grammar*, p. 293, §1040.

<sup>136</sup> ThTheoConf, I, p. XI ; PSALTES, *Grammatik*, p. 111-112, §222 ; P. YANNOPOULOS, *Les neutres en -ιν dans la Chronique de Théophane: un témoignage privilégié entre le grec classique et le grec moderne*, dans *Kainotomia. Die Erneuerung der griechischen Tradition. Le renouvellement de la tradition hellénique. Colloquium Pavlos Tzermias (4.IX.1995)*, ed. M. BILLERBECK, J. SCHAMP, Fribourg, 1996, p. 57-68.

<sup>137</sup> ThBasCaes, I, p. XLI ; PSALTES, *Grammatik*, p. 39, §87.

<sup>138</sup> ThBasCaes, I, p. XLIV.

<sup>139</sup> PSALTES, *Grammatik*, p. 3, §2.

<sup>140</sup> PSALTES, *Grammatik*, p. 40, §88.

<sup>141</sup> PSALTES, *Grammatik*, p. 4, §7.

<sup>142</sup> PSALTES, *Grammatik*, p. 50-51, §104.

<sup>143</sup> PSALTES, *Grammatik*, p. 134-136, §250 ; JANNARIS, *Grammar*, p. 294, §1040).

<sup>144</sup> ThTheoConf, I, p. XXV ; PSALTES, *Grammatik*, p. 85, §174. Cette règle aurait peut-être justifié la correction de la forme Ούλπικίου en Ούλπικίου plutôt qu'en Σουλπικίου (*s.l.* Σουλπίκιος), comme cela a été proposé, à la suite des Mauristes, dans la concordance de Basile de Césarée (ThBasCaes, I, p. XVIII et XLVII). Pour Σιλουία (lat. *Silvia*), une interrogation du *T.L.G.* fournit aussi des formes du type Ἰλουία.

<sup>145</sup> PSALTES, p. 58-59, §119.



στάβλον<sup>146</sup> ; la syncope affecte fréquemment les noms propres, par ex. par la chute du *iota*<sup>147</sup> : Δέκ(ι)μος, Δόμν(ι)νος ; Κλεπτ(ι)τιος ; Τίφλις *s.l.* Τίφιλις ;

[Page 244]

- pour les verbes contractes, transition d'un type flexionnel à un autre<sup>148</sup> :

- verbe en -άω conjugué comme un verbe en -έω<sup>149</sup> : ἀγωνιούμεθα *s.l.* ἀγωνιάω ; δαπανούμενος *s.l.* δαπανάω ; μωμείσθαι *s.l.* μωμάομαι ;

- verbe en -έω conjugué comme un verbe en -άω<sup>150</sup> : στομαχῶντες *s.l.* στομαχέω ; δυσθανατῶντες *s.l.* δυσθανατέω.

**§31.** Le lemme revêt une forme exempte de *variation phonétique*. Aux variations propres aux mots isolés s'ajoutent les variations propres aux mots dans la phrase.

1. Les variations phonétiques propres aux mots isolés :

- l'*aphérèse*, chute d'élément(s) phonique(s) à l'initiale d'un mot : μαῦρον *s.l.* ἀμαυρός ; θέλετε *s.l.* ἐθέλω<sup>151</sup> ; Σπανίας *s.l.* Ἴσπανία et Σπανικήν *s.l.* Ἴσπανικός<sup>152</sup> ;

- la *syncope*, chute d'élément(s) phonique(s) à l'intérieur d'un mot : Βερνίκης *s.l.* Βερενίκη (τοπ.)<sup>153</sup> ; ἐμπριζομένου *s.l.* ἐμπυρίζω<sup>154</sup> ; σκόρδων *s.l.* σκόροδον<sup>155</sup> ;

- la *métathèse de quantité*, échange de la quantité de voyelles successives : λεωφόρος *s.l.* λαοφόρος<sup>156</sup> ; υἱῆα et υἱέα *s.l.* υἱός ; ἀχιερῆος *s.l.* ἀρχιερέως ; πόλῆος *s.l.* πόλις (cfr aussi les formes relevant de la déclinaison attique, **§29, 1**) ;

- la *métathèse vocalique*, échange de la position de deux voyelles : πιτύαν *s.l.* πυετία<sup>157</sup> ;

<sup>146</sup> ThTheoConf, I, p. XX et XXVI.

<sup>147</sup> PSALTES, *Grammatik*, p. 33, §72.

<sup>148</sup> JANNARIS, *Grammar*, p. 219, §849-p. 221, §863.

<sup>149</sup> MANDILARAS, *Verb*, p. 62, §52(2) ; PSALTES, *Grammatik*, p. 233, §347.

<sup>150</sup> MANDILARAS, *Verb*, p. 62, §53.

<sup>151</sup> ThTheoConf, I, p. XIV ; LEJEUNE, *Phonétique*, p. 222, §230.

<sup>152</sup> ThTheoConf, I, p. XXII ; PSALTES, *Grammatik*, p. 31, §68.

<sup>153</sup> ThTheoConf, I, p. XI ; LEJEUNE, *Phonétique*, p. 223, §231 (hellénistique).

<sup>154</sup> ThTheoConf, I, p. XIV et XXII ; PSALTES, *Grammatik*, p. 32, §71.

<sup>155</sup> ThTheoConf, I, p. XI ; LEJEUNE, *Phonétique*, p. 223, §231 (hellénistique).

<sup>156</sup> ThTheoConf, I, p. XI.

<sup>157</sup> C 956, s.v. πυός ; cfr KINDT, *Avancées*, p. 8.

- la *métathèse consonantique*, échange de la position de deux consonnes : Καλχηδόνος s.l. Χαλκηδών<sup>158</sup> ;

- la *contraction*, amalgame de deux voyelles ou d'une voyelle et d'une diphtongue à l'intérieur d'un mot (cfr aussi §24) : ἄδονται s.l. αείδω ; βουστασίων s.l. βοοστάσιον ; γηπόνον s.l. γεωπόνον<sup>159</sup> ; δουλουργῶ s.l. δολοεργός ; νουμηνίας s.l. νεομηνία ; νοῦν s.l. νόος et ses composés ; φιλοῦντας s.l. φιλέω.

[Page 245]

2. Les variations phonétiques propres aux mots dans la phrase :

- les consonnes finales facultatives, *nu* éphelcystique et *sigma* adverbial, ainsi que l'élargissement en *kappa* de la négation οὐ, ne justifient pas la création d'un nouveau lemme<sup>160</sup> : τουτέστιν s.l. τουτέστι ; ἄχρις s.l. ἄχρι ; οὐκ s.l. οὐ ;

- l'*élision* : γράμματ' s.l. γράμμα ; ἀβρότητ' s.l. ἀβρότης ; αὐτικ' et αὐτίχ' s.l. αὐτίκα ; ἔπ' s.l. ἐπί ; μάθοιτ' s.l. μανθάνω ; ἔστ' , pour ἐστί(ν) s.l. εἰμί ;

- l'*élision inverse* : γαθέ s.l. ἀγαθός ; μῶ s.l. ἐμός ; 'στι, pour ἐστί(ν), s.l. εἰμί ;

- les *crases* (une unité graphique composée de deux unités lexicales) sont résolues et reçoivent deux lemmes : ἄνθρωπος s.l. ὁ et ἄνθρωπος<sup>161</sup> ; κἄν s.l. καί et ἄν, d'une part, ou ἐάν, d'autre part<sup>162</sup> ; certaines formes historiquement formées d'une crase mais lexicalisées<sup>163</sup> ont cependant un lemme propre : ταῦτουργόν s.l. ταῦτουργός ; ταῦτολόγει s.l. ταῦτολογέω.

**§32.** Le lemme revêt une forme exempte de variation accentuelle :

- la *barytonisation* et l'*accent d'enclise* sont traités par programme avant la comparaison du vocabulaire d'un texte aux données lexicales du D.A.G. ;

- l'*anastrophe* : les préverbes ou prépositions disyllabiques placés en anastrophe, et donc marqués d'un accent sur leur premier élément, ne justifient pas la création d'un nouveau lemme : πάρα s.l. παρά ; πέρι s.l. περί ; κατά s.l. κατά ;

<sup>158</sup> SCHWYZER, *Grammatik*, I, p. 269 (métathèse des consonnes).

<sup>159</sup> ThTheoConf, I, p. XI.

<sup>160</sup> JANNARIS, *Grammar*, p. 100, §220-224 ; LEJEUNE, *Phonétique*, p. 315, §360.

<sup>161</sup> ThGrNaz, I, p. XV.

<sup>162</sup> ThGrNaz, I, p. XV.

<sup>163</sup> Sur cette notion, cfr §108.

- la *variation de la place de l'accent* : par *principe d'économie*, la variation de l'accent d'un mot dans les sources ne justifie pas la création d'un nouveau lemme : μυρίοι et μύριοι *s.l.* μυρίος<sup>164</sup> ; Ἀδάουλφος et Ἐδαοῦλφος *s.l.* Ἀδάουλφος ; Ἀρμενιακῶν et Ἀρμενιάκων *s.l.* Ἀρμενιακός ; ἐγκλειστοῦ *s.l.* ἔγκλειστος<sup>165</sup> ; Μαυριτανούς et Μαυριτάνους *s.l.* Μαυριτανοί<sup>166</sup> ; Οὐανδάλων et Οὐάνδαλοι *s.l.* Οὐανδαλός.

- Remarque : il subsiste des cas de mots marqués de deux accents, suite à la suffixation de la particule enclitique de mouvement -δε (même si cet accent est labile dans les éditions) : Αἴγυπτόνδε, Μέγαράδε, Ἐλευσινάδε ; ces formes sont considérées comme lexicalisées (cfr §108).

[Page 246]

**§33.** La *variation orthographique* d'un mot dans les sources et dans les éditions :

1. L'orthographe d'un mot peut varier dans les sources et les éditions sans justifier la création d'un nouveau lemme : ἔννατον *s.l.* Ν-ἔνατος ; Κολυτός *s.l.* Κολλυτός ; κουράλλιον *s.l.* κοράλλιον ; φάλαινα et φάλλαινα *s.l.* φάλλαινα (κῆτος) ; Οὐαλλέριον *s.l.* Οὐαλέριος ; cfr aussi le caractère labile du *iota* souscrit : ἀθώσσει *s.l.* ἀθώσσις ; Ἡρώδην *s.l.* Ἡρώδης ; κωδίω *s.l.* κώδιον ; λάθρα, λάθρα, λάθρη *s.l.* λάθρα, ἐνιαχῆ et ἐνιαχῆ *s.l.* ἐνιαχῆ<sup>167</sup> ; de telles variations sont très fréquentes au niveau des emprunts : Βέβιος, le Vésuve, *s.l.* Οὐέσσοιος. Des critères sont proposés pour choisir, parmi les différentes graphies attestées, laquelle fournira le lemme<sup>168</sup>.

2. Les *variations* conditionnées par les *principes typographiques* propres à un milieu, une époque ou certaines éditions, par ex. celles des Mauristes et de la *Patrologie*<sup>169</sup> :

- un esprit doux puis un esprit rude surmontent le double *rhô* (-ῥῥ-) ;
- l'inf. des verbes contractes en -άω est noté -ᾶν au lieu de -ᾷν ;
- la voyelle en *hiatus* marquée par la *coronis* au lieu du tréma : προῦπέγραφεν *s.l.* προῦπογράφω ; προῦπαρξιν *s.l.* προῦπαρξις.

<sup>164</sup> ThTheoConf, I, p. XIII.

<sup>165</sup> Cfr E IV 83.

<sup>166</sup> ThTheoConf, I, p. XV.

<sup>167</sup> Pour ces deux derniers mots, cfr aussi §73.

<sup>168</sup> Cfr §95-97.

<sup>169</sup> ThGrNaz, I, p. XI.

- sur le traitement des voyelles en *hiatus*, cfr §24. L'absence d'un double point dans le texte d'une édition ne justifie pas la création d'un nouveau lemme ;
- sur le traitement des fautes d'édition, cfr §109-110.

### III. PRINCIPES MORPHO-SYNTAXIQUES ET SÉMANTIQUES : DISTINCTION DES LEMMES HOMOGRAPHES (§34-44)

**§34.** Les cas d'homographie s'observent soit au niveau des formes soit au niveau des formes et des lemmes :

- la forme ἀγαθῶν qui, hors contexte, peut avoir pour lemme ἀγαθός (adj. au gén. pl.) ou ἀγαθόω (v. à l'act. part. prés. nom. masc. sing.), la forme ἔχει qui peut avoir pour lemme ἔχισ (subst. au dat. sing.), ἔχω (v. à l'act. ind. prés., 3<sup>e</sup> pers. sing.) ou χέω (v. à l'act. ind. impft, 3<sup>e</sup> pers. sing.), relèvent du premier cas. Ces formes sont homographes, mais les lemmes qui les représentent sont graphiquement distincts les uns des autres. «C'est à celui qui lemmatise, le lemmatiseur, que reviendra la tâche de définir quel lemme est pertinent pour telle forme *in textu*» (cfr §4) ;

- la forme καρπὸν qui, hors contexte, peut avoir deux analyses selon qu'il s'agit du lexème καρπός «fruit» ou de καρπός «poignet», ou la forme δεῖν, qui peut être l'act. inf. prés. du lexème δέω «lier» ou de δέω «manquer», relèvent du second [Page 247] cas, car les différents lemmes possibles et proposés pour l'une et l'autre forme sont homographes. Une formulation pertinente des lemmes devra rendre compte de la distinction établie dans le lexique entre ces différents lexèmes. Les lignes qui suivent ne concernent donc que ces cas d'homographie au niveau des lemmes.

**§35.** Dans les dictionnaires traditionnels, les entrées homographes sont dégroupées et distinguées les unes des autres par des indices numériques ou alphabétiques : B 1022, καρπός I et καρπός II ; LS 879, καρπός A et καρπός B ; B 451-453, 1 δέω et 2 δέω ; LS 383, δέω (A) et δέω (B). Les mots καρπός «fruit» et καρπός «poignet» ou δέω «lier» et δέω «manquer», font l'objet d'un traitement homonymique, et non polysémique<sup>170</sup>, car ils sont étymologiquement, et donc formellement, issus d'étymons distincts<sup>171</sup>. Cette méthode de différenciation est peu appropriée à un processus de lemmatisation formelle qui reconnaît les formes et les lemmes sur base des séquences de caractères qui les constituent. Dans le D.A.G., les

<sup>170</sup> Sur ces méthodes, cfr REY-DEBOVE, *Étude*, p. 126-133 ; GAUDIN – GUESPIN, *Initiation*, p. 113-116 ; sur la polysémie, relevant de l'évolution sémantique d'un même lemme, cfr §41.

<sup>171</sup> Cfr respectivement, C 500-501, s.v. 1 καρπός («rapprochement avec le latin *carpo* "cueillir"») et 2 καρπός (\*κφαρπός) et C 269-270, s.v. 1 δέω (*deδ<sub>1</sub>-/dδ<sub>1</sub>-*) et 2 δέω («thème *δευ-* ou *δευσ-*»).

lemmes sont alors différenciés par l'ajout d'informations appelées *spécifications*, par ex. καρπός (μῆλον) pour καρπός «fruit» et καρπός (χείρ) pour καρπός «poignet», δέω (δήσω) pour δέω «lier» et δέω (δεήσω) pour δέω «manquer». Cette méthode convient aux exigences de la reconnaissance automatique tout en offrant au lemmatiseur une relative lisibilité de la distinction établie. Certaines spécifications anciennement établies ont été revues dans ce sens ; l'opposition σχοινίς, ἴδος «de jonc» (adj.) vs σχοινίς, ἴδος «corde de jonc» (subst.), peu éclairante, est désormais formulée sous la forme σχοινίς (σχοινικός) «de jonc» (adj.) vs σχοινίς (σχοῖνος) «corde de jonc» (subst.).

**§36.** Sur les 33.841 du D.A.G., 1.193 possèdent une spécification, soit 3,5% des lemmes, répartis dans toutes les classes morpho-syntaxiques. Sur les 707.853 occurrences du corpus de Basile de Césarée, 22.928 sont classées sous un lemme pourvu d'une spécification, soit 3,23% des formes (et, parmi ces dernières, 18.019 ont reçu plus d'une proposition de lemme). Les principes décrits dans cette partie portent donc sur une partie quantitativement réduite du lexique. Du point de vue qualitatif, ils demeurent pourtant indispensables pour rendre compte dans la langue des oppositions lexicales attestées *in textu*. Dans le cadre du *Projet*, l'exercice de la lemmatisation se situe toujours au niveau de la langue et non au seul niveau du discours. Les concordances lemmatisées publiées fournissent par conséquent les lemmes munis des spécifications qui les [Page 248] accompagnent dans le dictionnaire, même si un seul des différents lexèmes qu'elles permettent de distinguer s'actualise dans le corpus analysé : dans le *Thesaurus* consacré à Astérius d'Amasée, les formes du type πιστοί sont classées sous le lemme πιστός (πείθω), cité avec la spécification qui l'accompagne dans le D.A.G., même si le lemme πιστός (πιπίσκω), aussi attribuable à ces formes, ne s'actualise pas dans les homélies de cet auteur<sup>172</sup>. La *Bibliothèque* de Photios atteste le lemme Κίρρα (τοπ.), un toponyme dont le lemme s'oppose, dans le D.A.G., à un anthroponyme formulé Κίρρα (προσ.).

**§37.** La levée de l'ambiguïté nécessite une lecture attentive des formes dans leur contexte. La lemmatisation, loin d'être une opération mécanique limitée à un examen mot à mot du texte traité, constitue un exercice mettant en jeu la connaissance de la langue et

<sup>172</sup> ThAstFirm, p. 62.

la compréhension de la source analysée<sup>173</sup>. Le texte de Procope de Césarée présente la forme Κιθαρίζων classée sous le lemme verbal κιθαρίζω à l'issue du traitement automatique. Mais il s'agit en fait d'un toponyme d'origine arménienne. Dans la *Bibliothèque* de Photios, les formes κολωνόν et Κολωνῶ avaient toutes deux été rangées sous le lemme κολωνός. Mais la seconde se rapporte au lemme toponymique désignant le toponyme Colone.

**§38.** Lors d'un traitement informatique, deux mots différents seront homographes si les séquences de caractères qui les constituent sont identiques en tous points. Les spécifications ne sont utilisées que dans ces cas d'*homographie absolue* au niveau des lemmes. Ne sont donc pas considérés comme homographes et constituent des lemmes différents :

- les lemmes qui se distinguent par l'accentuation : ἀσφοδελός «plein d'asphodèles» (adj.) vs ἀσφόδελος «asphodèle» (subst.) ; Ἀθήναιος «Athénée» (anthr.) vs ἸΑθηναῖος «d'Athènes» (adj.) ; βατός «accessible» vs βάτος «ronce»<sup>174</sup> ; Ἑρμαῖος «Hermaios» (anthr.) vs Ἑρμαῖος «relatif à Hermès» (adj.) ; Χῖος «Chios» (top.) vs Χῖος «de Chios» (ethn.).

- les lemmes qui se distinguent par la casse de l'initiale : Ἐφιάλτης «Éphialte» (anthr.) vs ἐφιάλτης «cauchemar» ; Θυέστης «Thyeste» (anthr.) vs θυέστης «pilon» ; Ἰκτῖνος «Ictinos» (anthr.) vs ἰκτῖνος «milan» (oiseau) ; Χριστιανός «Chrétien» (anthr.) vs χριστιανός «qui professe la religion du Christ» (adj.) ; Ἡπειρώτης «Épirote» (ethn.) vs ἡπειρώτης «du continent» (adj.) ; Παιώνιος «Péonios» (anthr.) vs παιώνιος «relatif au péan, salutaire» (adj.)<sup>175</sup> ;

- les lemmes graphiquement proches mais pourvus de terminaisons différentes : πατέομαι «manger, absorber» vs πατέω «fouler aux pieds».

[Page 249]

**§39.** Les spécifications sont *morpho-syntaxiques* et *sémantiques*. Les premières ajoutent au lemme une information relative à sa classe morpho-syntaxique, c'est-à-dire :

<sup>173</sup> Sur cet aspect, cfr COULIE, *Lemmatization*, p. 35.

<sup>174</sup> Ce dernier terme étant lui-même homographe avec le nom d'un poisson et une unité de mesure, cfr *infra*, §44 et tableau 4, n°5.

<sup>175</sup> Il faut y joindre Παιόνιος «péonien, de Péonie (Παιονία)».

- à son *genre*, et donc souvent, aussi, à son paradigme flexionnel : ἔφοδος (ή) «chemin qui mène vers» vs ἔφοδος (ό) «qui fait une ronde de surveillance» ; ἄρμα (ή) «union» vs ἄρμα (τό) «attelage» ;

- à sa nature *nominale* ou *adjectivale* : ἄνοδος (ή) «chemin pour monter» vs ἄνοδος, ος, ον «sans route».

**§40.** Les spécifications sémantiques ajoutent au lemme une information relative :

- au sens du *lexème* : ἔαρ (αἷμα) «sang» vs ἔαρ (ἐαρινός) «matin» ; φάλλαινα (κῆτος) «baleine» vs φάλλαινα (ἔντομος) «phalène» ;

- au sens de *l'un de ses composants* : ἄλοχος (ἄμα) «compagne de couche» (avec *alpha* copulatif) vs ἄλοχος (χωρίς) «qui n'a pas enfanté» (avec *alpha* privatif).

**§41.** Ces critères sémantiques restent au service de la lemmatisation formelle ; «c'est le dénombrement des signifiants que l'on entreprend, non celui des signifiés»<sup>176</sup> :

- les spécifications n'opposent donc pas les différentes acceptions que peut revêtir un même terme : ἄγγελος «messager» n'est pas distingué de ἄγγελος «ange»<sup>177</sup> ; μέλος «membre, articulation» n'est pas distingué de μέλος «chant» (évolutions sémantiques) ; [φρούριον τῆς] Λήθης *s.l.* λήθη<sup>178</sup> (acception particulière du lemme) ; Θέμις «la Loi» *s.l.* θέμις<sup>179</sup>, Θεός «Dieu» *s.l.* θεός, Οὔτις «Personne» (le surnom que se donne Ulysse devant le Cyclope) *s.l.* οὔτις et Χριστός *s.l.* χριστός<sup>180</sup> (emplois personnifiés) ; γάλακτος [ὄρος] *s.l.* γάλα<sup>181</sup>, ὁ Παρθενών (le Parthénon) *s.l.* παρθενών, ἡ Χαλκῆ (la Chalcé, au palais impérial de Constantinople) *s.l.* χάλκεος, Ἑβδομον (l'Hebdomon, à Constantinople) *s.l.* Ν-ἑβδομος<sup>182</sup> (emplois toponymiques). L'emploi d'un subst. ou d'un adj. comme surnom ne justifie pas non plus la création d'un nouveau lemme<sup>183</sup> ; les lemmes ne sont pas non plus différenciés en fonction du domaine d'emploi

<sup>176</sup> MULLER, *Initiation*, p. 149.

<sup>177</sup> ThGrNaz, I, p. XV.

<sup>178</sup> ThTheoConf, I, p. XV.

<sup>179</sup> ThTheoConf, I, p. XV.

<sup>180</sup> ThGrNaz, I, p. XII.

<sup>181</sup> ThProcCaes, p. XLIV.

<sup>182</sup> ThProcCaes, p. XCIII.

<sup>183</sup> Cfr §47 et 51.



dans lequel s'actualisent les formes qu'ils représentent<sup>184</sup>. Les anthroponymes et toponymes [Page 250] attribués à plusieurs personnages ou différents sites géographiques ne sont pas non plus distingués au niveau des lemmes<sup>185</sup>.

- l'usage des spécifications relève d'une analyse synchronique de la langue et non diachronique ; les spécifications (-ω) qui accompagnent les verbes en -μαι attestés tardivement à l'actif, par ex. οἴομαι (-ω), seront par conséquent effacées.

**§42.** Quand il est nécessaire de distinguer des lemmes homographes, les critères formels, d'une part, et morpho-syntaxiques et sémantiques, d'autre part, sont pris en compte. Les lignes qui suivent définissent les principes de formulation des spécifications puis décrivent les oppositions qu'elles permettent d'établir. La formulation des spécifications répond donc aux règles énoncées ci-dessous :

1. Une spécification est une (et une seule) forme grecque jointe au lemme. Ce principe permet de représenter les oppositions lexicales d'une façon cohérente et facilement compréhensible (car les informations distinctives apparaissent en même temps que les lemmes), tout en permettant d'effectuer des dénombrements, des classements et des tris alphabétiques sur la totalité de l'adresse lexicale.

2. Une spécification est une forme grecque : διώνυμος (διά) «dont le nom se répand au loin» vs διώνυμος (δίς) «avec deux noms» ; οὔλος (εἰλέω) «roulé fortement sur soi-même» vs οὔλος (ὄλλυμι) «funeste».

<sup>184</sup> Cfr par ex. l'article ἐχίνος de B 874, qui enregistre sept sens différents, de «hérisson» à «gâteau» ; l'article καθαρμόζω de B 991 enregistre les sens «arranger, adapter, ajuster», auxquels s'ajoute «marry, épouse» dans L 684 ; l'article καρκίνος «crabe» de B 1021, enregistre six sens différents, jusqu'à «sorte de chaussure» ; ἐμβόλισμος, adjectif désignant les jours ou les mois intercalaires du calendrier, mais signifiant aussi «supposé» en parlant d'un enfant (B 652) ; ἐναυλίζω «passer la nuit dans» et, «en parl. d'un mal» (B 671), «se fixer» ; δόλιχος «le long stade» et «sorte de haricot» ; ἐκρέω «couler de», puis, «en parl. de cheveux» (B 630), «tomber» ; etc. Un cas à signaler cependant: Ἀσινάρια ( Ἀσίνιος) (pour la porta Asinaria à Rome, en référence à la famille des Asinii) vs Ἀσινάρια ( Ἀσίναρος) (pour la fête des Asinaria, en Sicile, en référence au fleuve Asinaros), cfr ThProcCaes, p. XXV.

<sup>185</sup> Dans les concordances de Théophraste le Confesseur et de Procope de Césarée, les différents personnages portant un même nom sont cependant distingués, au niveau des formes, par un indice numérique renvoyant à un index, cfr par ex. ThTheoConf, II, p. VIII ; ThProcCaes, p. XVI.

Dans le cas de spécifications morpho-syntaxiques, cette forme grecque peut être :

- un *article* : ἄνθος (ὁ) «bruant» (oiseau) vs ἄνθος (τό) «ce qui croît» ;
- les *terminaisons de l'adj.* : ἄνοδος, ος, ον «sans route» (adj.) vs ἄνοδος (ἡ) «chemin pour monter» (subst.) ; θεῖος, α, ον «divin» vs θεῖος (ὁ) «oncle» ; Ἰουδαῖος, α, ον «juif» (adj.) vs Ἰουδαῖος (ὁ) «Jude» (anthr.) ; λαιμός, ἡ, ὄν «creux» vs λαιμός (ὁ) «gorge» ; νεουργός, ὅς, ὄν «fait récemment» vs νεουργός (ὁ) «constructeur de navires» ;

[Page 251]

- un cas particulier : ὅς, ἡ, ὅ «qui» (pron. relatif) vs ὅς, ἡ, ὅν «mon, ton, son» (pron. possessif).

Dans le cas de spécifications sémantiques, cette forme grecque peut être :

- une *forme synonymique* :

- du *lemme* dans son ensemble : ἴος (βέλος) «trait» vs ἴος (φάρμακον) «venin» ; μήν (μείς) «mois» (subst.), pour opposer ce dernier mot à μήν (μέν) «certes» (particule affirmative) ; λύγξ (λυκόλυγξ) «lynx» vs λύγξ (λύγμος) «hoquet» ;
- de la *base* d'un lemme relevant d'un mot construit : περίειμι (εἰμί) «être autour» vs περίειμι (εἶμι) «aller autour» (et tous les composés existant pour les deux v.) ;
- d'un *morphème* constitutif du lemme. Relèvent de cette catégorie tous les lemmes présentant un *alpha* privatif et qui s'opposent à un autre lemme issu de la même base dérivationnelle mais commençant par un *alpha* copulatif : ἀγάστωρ (χωρίς) «qui est à jeun» vs ἀγάστωρ (ἄμα) «né du même sein» ;
- une *forme hyperonymique* : βάτος (μέτρον) «mesure de 50 setiers», pour opposer ce terme à βάτος (βάτον) «ronce» ; μάψ (ὄρνις) (un oiseau), pour opposer ce dernier mot à μάψ (μαψιδίως) «vainement» ; φάλλαινα (κῆτος) «baleine» (famille des cétacés) vs φάλλαινα (ἔντομος) «phalène» (famille des insectes) ;
- une *forme holonymique*<sup>186</sup> : καρπός (χείρ) «poignet», pour opposer ce dernier mot à καρπός (μῆλον) «fruit» ;

<sup>186</sup> Sur ces notions d'*hyperonymie* et d'*holonymie* relevant de la sémantique et des relations lexicales, d'ordre hiérarchique, entre les mots, cfr GAUDIN – GUESPIN, *Initiation*, p. 187-195 ; MORTUREUX, *Lexicologie*, p. 82-85.

- une *forme relevant de la même classe morpho-syntaxique* : ἤ (αἶ) «hé!» (excl.) vs ἤ (καί) «ou» (conj.); μάψ (μαψιδίως) «vainement», pour opposer ce dernier mot à μάψ (ὄρνις) (un oiseau); μήν (μέν) «certes» (particule affirmative), pour opposer ce dernier mot à μήν (μείς) «mois» (subst.); κρηνίς (κρηναῖος) (adj.) vs κρηνίς (κρήνη) (subst.); σχοινίς (σχοινικός) «de jonc» (adj.) vs σχοινίς (σχοῖνος) «corde de jonc» (subst.); φυλλάς (φύλλον) «feuille» vs φυλλάς (φυλλικός) «couvert de feuilles»; ὡς (εἰς) «vers» (prép.) vs ὡς (ὅς) «comme, que» (adv. et conj.);

- une *forme relevant de la même famille étymologique* : ἀδείς (δέησις) «qui n'a pas besoin de», pour opposer ce dernier mot à ἀδείς (δέος) «qui ne craint pas»; οὔλος (εἰλέω) «roulé fortement sur soi-même» vs οὔλος (ὄλλυμι) «funeste»;

- une *forme dérivée du lemme* : ἔαρ (ἔαρινός) «matin», pour opposer ce dernier mot à ἔαρ (αἶμα) «sang»; ἔως (ἔωλος) «aurore», pour opposer ce dernier mot à ἔως (μέχρι) «jusqu'à ce que»;

[Page 252]

- une *forme servant de base dérivationnelle du lemme* : ἀδείς (δέος) «qui ne craint pas», pour opposer ce dernier mot à ἀδείς (δέησις) «qui n'a pas besoin de»; ἄθεος (θεός) «sans dieu», pour opposer ce dernier mot ἄθεος (θεάομαι) «sans vision»; ἀνάξιος (ἄναξ) «kingly» vs ἀνάξιος (ἄξιος) «unworthy»; εὐνόμος (νόμος) «régé par de bonnes lois» vs εὐνόμος (-νομός) «riche en pâturages»; tous les cas présentant un couple de lemmes dérivé de εἶμι «être» ou de εἶμι «aller»; ce principe vaut aussi pour les noms propres : Μηλιεύς (Μῆλος) «de Mélos (île de l'Égée)» vs Μηλιεύς (Μηλίσ) «de Mélis (ville en Thessalie)» et Μηλιάς (Μῆλος) «de Mélos (île de l'Égée)» vs Μηλιάς (Μηλίσ) «de Mélis (ville en Thessalie)»;

- deux *éléments relevant de la métalangue* : (προσ.) pour les anthroponymes vs (τοπ.) pour les toponymes; Μακεδονία (προσ.) «Macédoine» (anthr.) vs Μακεδονία (τοπ.) «Macédoine» (top.); Λυδία (προσ.) «Lydie» (anthr.) vs Λυδία (τοπ.) «la Lydie» (top.).

3. En dehors des cas d'homographie, aucun lemme n'est pourvu de spécification; la majorité des lemmes adjectivaux enregistrés dans le D.A.G. est donc formulée sans les terminaisons -ος, η, ον, la majorité des toponymes sans la spécification (τοπ.): par ex. ἐλεφάντινος «d'ivoire»; Ῥώμη «Rome»; πύρινος (πῦρ) «de feu» vs πύρινος (πυρός) «de blé», mais πύριος «de feu», sans spécification car les sources connues n'attestent pas de forme relevant d'un lexème \*πύριος (πυρός) «de blé».

§43. Au-delà de la précision du sens de deux homographes de même nature, l'utilisation des spécifications permet de distinguer le genre ou le paradigme flexionnel des lemmes ainsi que leur classe morpho-syntaxique.

1. Précision du *genre* et/ou du *paradigme flexionnel* :

- masc. vs fém. : ἄλς (ὅ) «sel» vs ἄλς (ἡ) «mer» ;
- fém. vs neutre : κράνος (ἡ) «cornouiller» vs κράνος (τό) «casque» ;
- neutre vs masc. : γάνος (τό) «éclat d'un liquide limpide et brillant» vs γάνος (ὅ) «hyène» ; τάφος (τό) «stupeur» vs τάφος (ὅ) «funérailles, sépulture».

2. Précision de la classe morpho-syntaxique :

- subst. vs adj. : πλάνος (ὅ) «course errante» vs πλάνος, η, ον «errant» ; νεουργός (ὅ) «shipbuilder» vs νεουργός, ὅς, ὄν «new made» ; φυλλάς (φύλλον) «feuillée» vs φυλλάς (φυλλικός), «couvert de feuilles» ;
  - subst. vs adv. : μάψ (ὄρνις) (un oiseau) vs μάψ (μαψιδίως) «vainement» ;
  - subst. vs conj. : ἔως (ἔωλος) «matin» vs ἔως (μέχρι) «jusqu'à ce que» ;
  - anthr. vs adj. : Ἄγαπητός «Agapet» vs ἀγαπητός «aimé, chéri» ; Πύθιος (ὅ) «Pythios» vs Πύθιος, α, ον «Pythique» ; les noms de mois ont un lemme adjectival pour les distinguer des anthroponymes : Ἰούλιος, α, ον «du mois de juillet, juillet» vs Ἰούλιος (ὅ) «Jules» ;
  - anthr. vs ethn. :
    - en priorité par des critères purement formels, quand c'est possible : Ἀττικός (ὅ) «Atticus» vs Ἀττικός, ἡ, ὄν «d'Attique» ;
- [Page 253]**
- dans le cas contraire, par la combinaison de critères formels et sémantiques, ces derniers faisant état du toponyme équivalent : Νικαιεύς (ὅ) vs Νικαιεύς (Νίκαια) ; Πατρεύς (ὅ) vs Πατρεύς (Πατραί) ; Σαρμάτης (ὅ) vs Σαρμάτης (Σαρματία) ; Ἴων (ὅ) vs Ἴων (Ἴωνία) ; Λυκάων (ὅ) vs Λυκάων (Λυκαονία) ;
  - anthr. vs top. : Μακεδονία (προσ.) vs Μακεδονία (τοπ.) ; Ἰορδάνης (προσ.) vs Ἰορδάνης (τοπ.) ; Καλώνυμος (προσ.) vs Καλώνυμος (τοπ.) ;
  - prép. vs v. : ἄνω (ἀνά) vs ἄνω (ἀνύω).

**§44.** Dans la majorité des cas, les oppositions concernent deux termes. Elles peuvent cependant s'étendre à des séries de trois, quatre ou cinq termes. Le lemmatiseur doit être attentif à utiliser des spécifications homogènes, rendant par les mêmes critères les mêmes types d'oppositions. Le tableau 4 en fournit des exemples.

Pour une même forme homographe, la série des lemmes qui lui correspondent peut présenter des lemmes avec spécifications aux côtés de lemmes sans spécification : la forme ἄγει relève ainsi des lemmes ἄγος (ἄγνυμι) «fragment»<sup>187</sup>, ἄγος (ἄζομαι) «crime contre les dieux» et de ἄγω «mener» ; la forme πιστοῖς relève des lemmes πιστός (πείθω) «qu'on peut croire», πιστός (πιπίσκω) «qu'on peut boire» et de πιστόω «lier par une promesse». Une forme grecque peut parfois faire l'objet de cinq analyses différentes et donc se voir attribuer cinq lemmes différents ; la forme νέων est enregistrée dans le D.A.G. sous νέος «nouveau» (adj.), νέω (νήχω) «nager» (v.), νέω (νήμα) «filer» (v.), νέω (νήεω) «entasser» (v.), et Νέων «Néon» (anthr.). Un delta seul (δ) peut, selon les signes diacritiques qui l'accompagnent, recevoir six analyses différentes : le cardinal «4» (N-4), l'ordinal «4<sup>e</sup>» (N-4-ος), le cardinal «4000» (N-4000), l'ordinal «4000<sup>e</sup>» (N-4000-ος), la lettre grecque «delta», et l'abréviation de la particule δέ.

[Page 254]

<sup>187</sup> Cfr LS 14 ; C 12, s.v. ἄγνυμι.

Tableau 4 : séries de spécifications.

1. ἀγάλακτος (χωρίς) «qui ne fournit pas de lait»	vs	ἀγάλακτος (ἄμα) «όμογάλακτος» <sup>188</sup>
ἀγάστωρ (χωρίς) «qui est à jeun»	vs	ἀγάστωρ (ἄμα) «né du même sein»
ἀλέκτωρ (χωρίς) «vierge»	vs	ἀλέκτωρ (ἄμα) «έρουκ» <sup>189</sup>
et donc aussi :		
ἄλοχος (χωρίς) «qui n'a pas enfanté»	vs	ἄλοχος ἄμα) «compagne de couche»
et non : *ἄλοχος, ος, ον	vs	*ἄλοχος (ή)
2. ἀνάξιος (ἄναξ) «à la façon d'un roi»	vs	ἀνάξιος (ἄξιος) «indigne»
ἀναξία (ἄναξ) «pouvoir»	vs	ἀναξία (ἄξιος) «indignité»
3. Ναυατιανός, ή, όν «novatien»	vs	Ναυατιανός (ό) «Novatianus»
Νεστοριανός, ή, όν «nestorien»	vs	Νεστοριανός (ό) «Nestorianus»
4. Πέρσης (ό) «Persès»	vs	Πέρσης (Πέρσις) «perse»
Σύρος (ό) «Syros»		Σύρος (Συρία) «syrien»
Περσικός (ό) «Persicos»	vs	Περσικός (Πέρσις) «perse»
5. βάτος (βάτον) «ronce»	vs	βάτος (μέτρον) «mesure de 50 setiers» <sup>190</sup>
κόρος (κοῦρος) «jeune garçon»	vs	κόρος (μέτρον) «mesure de 6 médimnes attiques» <sup>191</sup>

<sup>188</sup> Cité en grec dans LS 5.

<sup>189</sup> Il faut y joindre ἀλέκτωρ (ἀλέξω) «coq». B 74 et LS 62 dégroupent autrement ces trois termes, cfr §17.

<sup>190</sup> Il faut y joindre βάτος (ιχθύς) «sorte de raie».

<sup>191</sup> Il faut y joindre κόρος (κορέννυμι) «satiété» vs κόρος (κορέω) «ordure».

#### IV. PRINCIPES DE LEMMATISATION LIÈS À LA CLASSE MORPHO-SYNTAXIQUE DES MOTS (§45-89)

1. Les substantifs (§45-50). 2. Les adjectifs (§51-65). 3. Les numéraux (§66-71). 4. Les adverbes (§72-76). 5. Les verbes (§77-85). 6. Les pronoms (§86). 7. Les prépositions (§87-88). 8. Les conjonctions (§89).

##### 1. Les substantifs

**§45.** Les noms propres sont enregistrés dans le *D.A.G.* (cfr **§9, 1**); l'initiale des dérivés de noms propres est une majuscule (βακχασμόν *s.l.* Βακχασμός [de Βάκχος «Bacchus»]; δηλιακός *s.l.* Δηλιακός «déliaque»; Ἑλληνισμός) mais une minuscule quand le dérivé est un verbe (ἀττικίζω; ἑλληνίζω) ou un mot dérivé ou composé dont le nom propre constitue le second élément (ὑποφρύγιος; ψευδαντωνῖνος).

**§46.** Un subst. désignant un titre ou une fonction et utilisé comme anthr. ne justifie pas la création d'un nouveau lemme anthroponymique: Πιτυάξην et Χαναράγγου *s.l.* πιτυάξης et χαναράγγης<sup>192</sup>.

[Page 255]

**§47.** Un subst. utilisé comme surnom ne justifie pas la création d'un nouveau lemme anthroponymique, sauf si le mot demeure inconnu par ailleurs<sup>193</sup>: [Πέτρον τὸν] Κναφέα *s.l.* κναφεύς, «cardeur, foulon», mais [Ἰωάννην τὸν πατρίκιον, τὸ ἐπίκλην] Πιτζιαῦδιν *s.l.* Πιτζιαῦδης (non attesté).

**§48.** Un subst. utilisé comme top. ne justifie pas la création d'un nouveau lemme toponymique, sauf si l'étymologie rend le rapprochement improbable: Ἄλγος et Ὀμόνοια (noms de forts) *s.l.* ἄλγος et ὀμόνοια; mais Κρέας *s.l.* Κρέας (et non κρέας «chair»<sup>194</sup>); [τοῦ] Πελαργοῦ (quartier de la Cigogne à Constantinople) *s.l.* πελαργός.

**§49.** Un nom de région dérivé d'un adj. conserve cependant un lemme nominal<sup>195</sup>: Ἀττική, Ἰουδαία, Βυζακηνή, Δαυλία; Περαία, Σουσιανή, Τρωγλοδυτική.

<sup>192</sup> Cfr *ThProcCaes*, p. XXVIII-XXIX.

<sup>193</sup> *ThTheoConf*, p. XII.

<sup>194</sup> *ThProcCaes*, p. XII.

<sup>195</sup> *ThGrNaz*, I, p. XIV; *ThTheoConf*, I, p. X.

**§50.** Un anthr. utilisé au pluriel ne justifie pas la création d'un nouveau lemme : Ἄρταβάνους *s.l.* Ἄρταβάνης; Γρηγόριοι *s.l.* Γρηγόριος; Ἐλέναι *s.l.* Ἐλένη; Θησεῖς *s.l.* Θησεύς; Κορνηλιανούς *s.l.* Κορνηλιανός; Λαζάρους *s.l.* Λάζαρος; Μαρκιῶνες *s.l.* Μαρκίων; Οὐαλεντῖνοι *s.l.* Οὐαλεντῖνος (ὁ); [τῶν δύο] Τραχινίων *s.l.* Τραχίνιος.

## 2. Les adjectifs

**§51.** Un adjectif substantivé conserve un lemme adjectival.

**§52.** Un adj. utilisé comme surnom ne justifie pas la création d'un nouveau lemme anthroponymique, sauf si le mot demeure inconnu par ailleurs : [Τιμόθεος ὁ] Λευκός *s.l.* λευκός, [Ἰωάννης ὁ] Φιλόπονος *s.l.* φιλόπονος, «qui aime le travail»; mais [Στεφάνω πατρικίῳ, τὸ ἐπίκλην] Ἀσμίκτῳ *s.l.* Ἄσμικτος «non nettoyé» (non attesté)<sup>196</sup>.

**§53.** Un adj. utilisé comme top. ne justifie pas la création d'un nouveau lemme toponymique : [ἐν τῇ] Μέση (la Mésé à CP) *s.l.* μέσος; [ἐπὶ τὴν καλουμένην] Χαλκῆν (la Chalcé à CP) *s.l.* χάλκεος. Pour les noms de régions dérivés d'adj., du type Ἀττική, cfr **§49**.

**§54.** Un adj. de formation régulière désignant un ethnique a, par *principe d'économie*, un lemme singulier, même si les sources n'attestent le mot qu'au pluriel : Εὐαισηνοῖς *s.l.* Εὐαισηνός<sup>197</sup>.

**§55.** Un adj. de formation régulière désignant le partisan d'une hérésie a, par *principe d'économie*, un lemme au singulier, même si les sources n'attestent le mot qu'au pluriel : αὐτοματιστῶν *s.l.* αὐτοματίστης<sup>198</sup>.

[Page 256]

**§56.** Les adj. verbaux en -τός et leurs degrés de comparaison ont un lemme adjectival : λεπτότεροι *s.l.* λεπτός<sup>199</sup>; φευκότερον *s.l.* φευκτός, αἰρετώτατην *s.l.* αἰρετός.

**§57.** Pour les adj. verbaux en -τέος, cfr **§83**.

**§58.** Les adj. défectifs ont un lemme correspondant aux formes attestées, cfr **§104**.

**§59.** Pour les formes supplétives, cfr **§102-103**.

<sup>196</sup> ThTheoConf, I, p. XII.

<sup>197</sup> ThBasCaes, I, p. XL.

<sup>198</sup> ThTheoConf, I, p. XI.

<sup>199</sup> ThGrNaz, I, p. XIV.



§60. Sur un adj. dérivant un adv., cfr §72-73.

§61. Sur un adj. dérivant une prép., cfr §88.

§62. Les comp. et sup. dérivés d'un adj. ont pour lemme cet adjectif au positif nom. masc. sing.

§63. Les comp. et sup. dérivés d'un adv. ou d'une prép. (ἄνω [ἀνά], ἐγγύς, ἔνδον, ἔξω, πρόσω, ὑπέρ) ont un lemme adjectival de comp. : ἀνώτεροι s.l. ἀνώτερος; ἑξωτέρας s.l. ἑξώτερος (sur les formes adverbiales, cfr §76).

§64. Les comp. et sup. dérivés d'un subst. ont un lemme adjectival de comp., cfr §101.

§65. Les comp. et sup. dérivés d'un verbe au participe ont un lemme adjectival de comp., cfr §101.

### 3. Les numéraux

§66. Les numéraux sont enregistrés dans le D.A.G. car ils sont considérés comme faisant partie du lexique, cfr §9, 1.

§67. La formulation du lemme dépend de la forme sous laquelle apparaît le numéral *in textu*<sup>200</sup> :

- les numéraux cardinaux écrits en toutes lettres reçoivent un lemme en toutes lettres précédé de l'indication N- : πέντε s.l. N-πέντε ;

- les numéraux cardinaux écrits dans le système alpha-numérique – stigma = 6 (ς), koppa = 90 (ϩ) et sampi = 900 (ϳ) compris – reçoivent un lemme en chiffre arabe précédé de l'indication N- : ε' s.l. N-5 ;

- les numéraux ordinaux écrits en toutes lettres reçoivent un lemme en toutes lettres précédé de l'indication N- : πέμπτην s.l. N-πέμπτος ;

- les numéraux ordinaux écrits dans le système alpha-numérique reçoivent un lemme en chiffre arabe précédé de N- et suivi de -ος : ε' s.l. N-5-ος. Un numéral fonctionnant comme déterminant d'un subst. est classé sous le lemme ordinal ; dans l'expression «κεφάλαιον α'», la forme α' reçoit le lemme N-1-ος.

§68. Les formes de χίλιοι reçoivent le lemme numéral N-χίλιοι. Mais les formes du type μύριοι sont classées sous μυρίος qui fonctionne à la fois comme [Page 257] adj. («très nombreux, innombrable») et comme numéral («10.000»). La distinction μυρίος vs μύριοι basée sur

<sup>200</sup> Cfr ThTheoConf, I, p. XIII ; B. KINDT, recension de ThTheoConf, I, dans *Le Muséon*, 111 (1998), p. 247.

l'accent n'est pas retenue<sup>201</sup>. Les composés de ces deux termes reçoivent cependant des lemmes numéraux précédés de l'indication N- (par ex. N-δισχίλιοι ; N-τρισχίλιοι ; N-δισμύριοι ; N-τρισμύριοι, etc.).

**§69.** Un numéral utilisé comme toponyme ne justifie pas la création d'un nouveau lemme : Ἑβδομον (l'Hebdomon, à Constantinople) s.l. N- ἑβδομος.

**§70.** Dans des expressions du type «πρὶν ἂν τὸ λ' τῆς Ἰλιάδος ἀναγνώσειε» ou «ἐν τῇ ω τῆς Ὀδυσσεΐας» les formes λ' et ω signifient respectivement «onze», et non «trente», et «vingt-quatre», et non «800». Les lemmes sont donc λ, la lettre *lambda*, et ω, la lettre *oméga* (sur le traitement des lettres, cfr **§111**).

**§71.** Remarque : il résulte de ce mode de traitement que les nombres écrits en plusieurs mots sont distingués en plusieurs lemmes, alors que ceux écrits dans le système alpha-numérique ne reçoivent qu'un seul lemme : εἴκοσι πέντε reçoit deux lemmes, N-εἴκοσι et N-πέντε, mais κέ' est lemmatisé N-25.

#### 4. Les adverbes

**§72.** Les adv. dérivés d'une forme casuelle, nominale ou adjectivale, qui n'est plus attestée à date historique ont un lemme adverbial :

- les adv. en -ει et -οι, «anciennes formes de locatif»<sup>202</sup> : οἴκοι s.l. οἴκοι (et non \*οἴκος) ;

- les adv. en -η, c'est-à-dire les «adverbes en \*-ē, probablement d'anciens instrumentaux»<sup>203</sup>, qui revêtent la forme au datif avec *iota* souscrit, même si les manuscrits et les éditions présentent les graphies concurrentes en -α, -α/-αι, -η ou -η. Le choix de la forme au datif avec *iota* souscrit repose sur le *principe d'économie*, cfr p. ex. : εἰκῆ et εἰκῆ s.l. εἰκῆ ; λάθρα, λάθρα, λάθρη s.l. λάθρα<sup>204</sup> ; πῆ, πῆ s.l. πῆ ;

- les adv. en -ως<sup>205</sup>, quelle que soit la classe morpho-syntaxique de leur base dérivationnelle :

<sup>201</sup> B 1306, s.v. μυρίος (*in fine*).

<sup>202</sup> CHANTRAINE, *Morphologie*, p. 121, §132.

<sup>203</sup> CHANTRAINE, *Morphologie*, p. 120-121, §132.

<sup>204</sup> ThBasCaes, I, p. XLII.

<sup>205</sup> Également issu d'une ancienne forme d'instrumental, cfr CHANTRAINE, *Morphologie*, p. 121, §132.

- adjectivale : ἀγαθῶς (et non \*ἀγαθός); διατατικῶς (et non \*διατατικός); κεφαλαιωδῶς (et non \*κεφαλαιώδης);
- adjectivo-pronominale : οὐδαμῶς, οὕτως, πάντως;
- verbale, que les adv. soient formés à partir d'un part. prés. (δεόντως, λανθανόντως) ou parf. (διηρημένως<sup>206</sup>, περικεκαλυμμένως).

[Page 258]

**§73.** *A contrario*, les adv. dérivés d'une forme casuelle, nominale ou adjectivale, toujours attestée à date historique ne justifient pas la création d'un lemme adverbial; il s'agit, par exemple :

- des adv. issus d'acc. sing. : ἐναντίον *s.l.* ἐνάντιος; ἐνώπιον *s.l.* ἐνώπιος; μάτην *s.l.* μάτη<sup>207</sup>; ὀλίγον *s.l.* ὀλίγος<sup>208</sup>; πολύ *s.l.* πολύς (mais ἐπιπολύ, καταπολύ, παραπολύ); χάριν *s.l.* χάρις;
- des adv. issus d'acc. pl. : μέγала *s.l.* μέγας; πολλά *s.l.* πολύς.

**§74.** Les adv. formés par dérivation suffixale ont un lemme adverbial : -δε/-ζε : Μέγαράδε, Ἀθήναζε; -δην : κρύβδην; -δον : ἀγεληδόν, στοιχηδόν; -θα : ἐνταῦθα; -θε : Ἀσσυρήθε, ὑπόθε; -θεν : Ἀθήνηθεν, Μεγαρόθεν; -θι : ὑπόθι; -σε : ὑπόσε; -φι : νόσφι.

**§75.** La particule démonstrative -ι en finale des adv. dérivés de démonstratifs ne justifie pas la création d'un nouveau lemme (cfr **§29, 1**) : οὕτως *s.l.* οὕτως; οὐχί *s.l.* οὐ.

**§76.** Les adv. au comp. et sup. dérivés d'un adv. ou d'une prép. (ἄνω [ἀνά], ἐγγύς, ἔνδον, ἔξω, πρόσω, ὑπέρ) ont pour lemme cet adv. ou cette prép.<sup>209</sup> : ἀνωτέρω *s.l.* ἄνω; ἐγγυτάτω, ἐγγυτέρω *s.l.* ἐγγύς; ἐνδοτάτω *s.l.* ἔνδον; ἐξωτέρω *s.l.* ἔξω; προσωτέρω, πορροτάτω *s.l.* πρόσω; πρωΐτερος *s.l.* πρωΐ (sur les formes adjectivales, cfr **§63**).

## 5. Les verbes

**§77.** Les lemmes verbaux ont une terminaison active ou médio-passive ;

<sup>206</sup> ThGrNaz, I, p. XV.

<sup>207</sup> ThGrNaz, I, p. XV.

<sup>208</sup> ThGrNaz, I, p. XV.

<sup>209</sup> ThTheoConf, I, p. XIII.

**§78.** Les v. défectifs justifient la création d'un nouveau lemme correspondant aux formes attestées, cfr **§104**.

**§79.** Pour les cas de supplétisme, cfr **§107**.

**§80.** Les formes récentes ou tardives constituant la «forme de départ» d'une nouvelle série productive justifient la création d'un nouveau lemme : περιέκρυβεν *s.l.* περικρύβω (et non περικρύπτω<sup>210</sup>) ; πανθάνουσι *s.l.* πανθάνω (et non πάσχω) ; συστήκειν *s.l.* συστήκω (et non συνίστημι) ; ἀποδωσείοντες *s.l.* ἀποδωσείω<sup>211</sup> ; πέπτεται *s.l.* πέπτω ; ἐγεγώνησε *s.l.* γεγωνέω.

**§81.** Les v. thématiques dérivant tardivement d'un v. athématique et constituant une nouvelle série productive représentable par une «forme de départ» ont un lemme verbal thématique : σκεδαννύεις vs σκεδαννύω, mais καθειργνύει *s.l.* καθείργνυμι ; καταπληγνύναι *s.l.* καταπλήσσω.

**§82.** Sur les formes d'impft et d'aor. thématiques en -σκ-ον, cfr **§29, 2**.

**§83.** Un adj. verbal en -τέος a un lemme verbal, cfr **§57**.

[Page 259]

**§84.** Un part. substantivé a un lemme verbal.

**§85.** Sur un part. dérivant un adv., cfr **§72**.

## 6. Les pronoms

**§86.** Les pron. réfléchis ont pour lemme la forme au gén. : ἑμαυτοῦ ; σεαυτοῦ ; ἑαυτοῦ.

## 7. Les prépositions

**§87.** Une prép. dérivée d'un subst. a un lemme nominal : χάριν *s.l.* χάρις<sup>212</sup>.

**§88.** Une prép. dérivée d'un adj. a un lemme adjectival : ἔναντιον *s.l.* ἔναντιος<sup>213</sup>, cfr **§61**.

<sup>210</sup> ThTheoConf, I, p. XIV.

<sup>211</sup> ThProcCaes, p. XXIV.

<sup>212</sup> ThGrNaz, I, p. XV.

<sup>213</sup> ThGrNaz, I, p. XV.

## 8. Les conjonctions

§89. Une conjonction dérivée d'un adj. a un lemme adjectival : οἷον  
s.l. οἷος<sup>214</sup>.

<sup>214</sup> ThGrNaz, I, p. XV.

## V. PRINCIPES LIÉS AU STATUT LEXICAL DES LEMMES (§90-111)

1. Les formes rares (§90-94). 2. Les formes présentant plusieurs graphies et les emprunts (§95-97). 3. Les lemmes défectifs et les formes supplétives (§98-107). 4. Les formes et locutions lexicalisées (§108). Les fautes d'édition (§109). 6. Les lettres, les phonèmes et les morphèmes (§111).

### 1. Les formes rares

**§90.** Le lemme est une «forme de départ» censée être à la tête d'une série productive (cfr **§3**). Plus un mot est rare, plus la probabilité que cette «forme de départ» soit attestée *in textu* diminue. Deux options s'offrent alors aux lexicologues et lexicographes : une approche hypercritique ou une approche plus pragmatique.

La première consisterait à ne prendre comme lemmes que des formes attestées. Cette position est intenable car elle ferait du dictionnaire une «collection de monstres» encombrant de manière exponentielle la nomenclature et laissant croire à l'utilisateur qu'une grande part du lexique présente des flexions défectives. Cette première approche s'oppose également à l'hypothèse, déjà présente chez F. de Saussure, de l'existence d'un lexique potentiel constitué de mots non encore (ou pas encore) attestés, certes, mais déjà présents en langue (puisque relevant de [Page 260] règles de construction morphologiquement régulières, systématisables et prédictibles), leur non-attestation dans le discours étant qualifiée de «fait insignifiant»<sup>215</sup>.

**§91.** Dans le cas du grec, l'état fragmentaire des sources conservées incite de plus à la prudence<sup>216</sup> : l'absence d'une forme dans les sources, c'est-à-dire au niveau du discours, ne suffit jamais pour déterminer son absence au niveau du lexique, et donc en langue. Les textes offrent par ailleurs de nombreux cas de «formes de départ» attestées une seule fois<sup>217</sup>. Citées telles quelles dans le dictionnaire, sans mention de leur nature d'*hapax*, elles fourniraient du lexique

<sup>215</sup> SAUSSURE, *Cours*, p. 227 ; LEHMANN – BERTHET, *Introduction*, p. 5-6 ; CORBIN, *Morphologie*, p. 40-43.

<sup>216</sup> Pour des estimations chiffrées de l'importance des sources perdues, cfr FAMERIE, *Appien*, p. 252-253 (avec références bibliographiques).

<sup>217</sup> Pour des exemples, cfr *ThProcCaes*, p. X et *ThBasCaes*, I, p. XIV.

une image biaisée, encore une fois, puisque ce qui est rare y serait présenté comme ce qui est commun<sup>218</sup>.

**§92.** Le caractère fortement prédictible des flexifs, d'une part, et la productivité des éléments morphématiques constitutifs des mots analysés (productivité mise en évidence par la commutation paradigmatique), d'autre part, sont susceptibles d'offrir les indices suffisants pour définir, par déduction à partir des formes existantes, des lemmes cohérents, c'est-à-dire revêtant une forme possible et régulière. Ces lemmes fourniront une «forme de départ» non attestée certes, mais utile au classement des réalités formelles observées. C'est l'approche pragmatique habituellement adoptée par les lexicographes. En grec, elle s'impose pour la formulation des entrées lexicales verbales, connues sous un très grand nombre de formes différentes, mais elle est tout aussi valide pour les mots issus des autres classes morpho-syntaxiques ; c'est la méthode suivie pour le *D.A.G.* La formulation des lemmes des mots rares repose dès lors sur un examen des formes concernées. Dans ce contexte, la fréquence du mot compte moins que la variété des formes connues<sup>219</sup>. Deux formes fléchies différentes suffiront pour [Page 261] déduire une «forme de départ» ; une seule forme suffira aussi, moyennant un examen des éléments morphologiques qui la constituent. Ce *modus operandi* s'apparente partiellement à la démarche suivie par le linguiste qui examine la régularité d'une unité du lexique potentiel d'une langue moderne. Selon ses observations, il pourra conclure qu'il s'attend à rencontrer le terme étudié, et il le qualifiera de néologisme dès qu'il remarquera son actualisation dans le discours, oral ou écrit. Le lemmatiseur, pour sa part, conclura qu'il ne s'étonne pas de le rencontrer, mais l'état fragmentaire et provisoire des sources disponibles l'engagera à

<sup>218</sup> Qualifier un mot d'*hapax* signifie qu'il n'est attesté qu'une seule fois dans les sources disponibles, c'est-à-dire en discours. Cette notion n'affecte en rien le statut du mot au niveau de la langue, cfr J.-F. SABLAYROLLES, *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes* (*Lexica. Mots et dictionnaires*, 4), Paris, 2000, p. 165 et 198-203 ; CORBIN, *Morphologie*, p. 25-28 ; B. MOREUX, *L'utilisation des méthodes quantitatives en linguistique grecque et latine*, dans *L'Antiquité Classique*, 51 (1982), p. 295-296.

<sup>219</sup> L'interrogation du *T.L.G.* fournit quatre occurrences du lemme ἐπιτινάσσω, mais il s'agit à chaque fois de la même forme, cfr ThBasCaes, I, p. XL. Les lemmes αἰτιολόγημα, δυσπαρατηρήτος, ἐναποκεύω et ὑπογύναιος, toujours chez Basile de Césarée, se rencontrent aussi dans d'autres sources, mais toujours sous la forme qu'ils revêtent chez cet auteur, cfr, respectivement, ThBasCaes, I, p. XXV, XXXVIII-XXXIX et XLIX.

manipuler avec une extrême prudence les notions d'*hapax* ou de néologisme.

**§93.** Trois cas de figures se présentent :

1. Quand la «*forme de départ*» est attestée, que les sources fassent ou non état d'autres formes fléchies, la «*forme de départ*», même rare, voire unique, sert de lemme : dans le lexique de Basile de Césarée, il en est ainsi du subst. συνάθλησις, des anthroponymes Ἰωσάκης et Σύμπιος, de l'adj. φιλαireτικός, du v. συμπεριπτύσσομαι<sup>220</sup>.

2. Quand la «*forme de départ*» n'est pas attestée, mais que les sources font état d'au moins deux formes fléchies différentes, le caractère fortement prédictible des flexifs permet de déduire le lemme : le subst. ἀρχικουνίτης est déduit des formes ἀρχικουνίτη et ἀρχικουνιτῶν<sup>221</sup> ; l'adj. δυσамάρτητος est déduit des formes δυσамαρτήτω, rencontrée chez l'auteur analysé, et δυσамάρτητοι, attestée chez Vettius Valens<sup>222</sup> ; le v. εἰσλαμβάνω, est déduit des formes εἰσλαμβάνων et εἰσλαμβάνουσιν<sup>223</sup> ; l'anthr. Μάκης est déduit des formes Μακήτα et Μακήτι connue dans l'épigraphie<sup>224</sup>.

3. Quand la «*forme de départ*» n'est pas attestée, et que les sources ne font état que de la seule forme fléchie en cours de lemmatisation, la prédictibilité des flexifs et la productivité des morphèmes constitutifs de la forme permettent de confirmer ou d'infirmer la régularité du mot et, ensuite, de déduire un lemme pertinent :

- si la *forme est analysable*, si les éléments morphématiques qui la constituent sont lisibles, le mot présentant une certaine «*transparence formelle*»<sup>225</sup>, un [Page 262] «*étayage paradigmatique*»<sup>226</sup> permet de déduire le lemme sur des bases rationnelles : le subst. 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<sup>227</sup> ; l'anthroponyme Μαγνινιανός est déduit des formes

<sup>220</sup> ThBasCaes, I, p. XIV ; cfr aussi ThProcCaes, p. X et note 27. Un exemple tiré de Photios: καρβωνάριος.

<sup>221</sup> ThBasCaes, I, p. XIV.

<sup>222</sup> ThBasCaes, I, p. XV.

<sup>223</sup> ThAstFirm, p. XVII.

<sup>224</sup> Cfr par ex. C.I.G., n° 1740.

<sup>225</sup> Sur cette notion, cfr APOTHÉLOZ, *Construction*, p. 52 ; J. LEROT, *Précis de linguistique générale*, Paris, 1993, p. 351.

<sup>226</sup> Expression reprise à APOTHÉLOZ, *Construction*, p. 25.

<sup>227</sup> ThBasCaes, I, p. XV ; CHANTRAINE, *Morphologie*, p. 321, §258.



Μαγνηνανῶ et Μαγνινανῶ sur base de la productivité des noms en -ινανός, d'origine latine<sup>228</sup> ; le v. συγκαταδέχομαι<sup>229</sup> 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<sup>230</sup> et celle de la double préfixation en συ-κατα-<sup>231</sup> ;

- si la forme n'est pas analysable, si les éléments qui la constituent ne sont pas lisibles, elle sert de lemme à elle-même, le lemmatiseur ne prenant pas la responsabilité de définir une «forme de départ» systématisable : Πουτεδιν<sup>232</sup>, πολύπλαδον<sup>233</sup>. Ces entrées lexicales sont appelées «lemmes-formes»<sup>234</sup>. Il s'agit non seulement de mots non-analysables (Λακίζοις, un toponyme<sup>235</sup>), mais aussi, souvent, de formes assurément fautives (Ὀβούσιριν<sup>236</sup>), et de formes rencontrées dans des extraits présentant des problèmes d'établissement de texte (μοχλιάση, κονοπιασθῆ et κονοπισθῆ<sup>237</sup>). Les «lemmes-formes» relèvent donc des principes d'économie et de fidélité.

**§94.** Le corpus basilien rassemble 14.943 lemmes. Parmi ceux-ci, 148 appartiennent à la catégorie des lemmes pour lesquels une «forme de départ» non attestée est déduite selon les critères établis ci-dessus, soit 0,9%<sup>238</sup>. C'est peu proportionnellement à l'ensemble du vocabulaire de Basile et par rapport au lexique grec en général. C'est énorme en termes de temps de recherches, car chacun de ces 148 lemmes a fait l'objet d'une «enquête lexicographique». Mais en définitive, le statut lexical de ces mots, précisé par un indice spécial dans le [Page 263] D.A.G., y est mieux défini que dans les dictionnaires traditionnels car les unités lexicales rares et formulées par déduction y sont présentées comme telles. Le *Thesaurus Graecae Linguae*, le

<sup>228</sup> ThBasCaes, I, p. XV et XLIII.

<sup>229</sup> ThAstFirm, p. XVIII.

<sup>230</sup> C 268, s.v. δέχομαι.

<sup>231</sup> Une interrogation du T.L.G. fournit un exemple dès le VII<sup>e</sup> s. av. J.-C., dans la deuxième lettre de Périandre (= *Epistolographi Graeci*, ed. R. HERCHER, Paris, 1873 [réimpr. Amsterdam, 1965], p. 408).

<sup>232</sup> ThProcCaes, p. XI.

<sup>233</sup> ThAstFirm, p. XVIII.

<sup>234</sup> À ce propos, cfr aussi COULIE, *Lemmatisation*, p. 38 ; ThTheoConf, p. XVI ; ThProcCaes, p. XI.

<sup>235</sup> ThBasCaes, I, p. XV: la forme de départ devrait-elle être Λάζικος, Λάκιζον ou Λάκιζα?

<sup>236</sup> ThTheoConf, I, p. XXIV.

<sup>237</sup> ThBasCaes, I, p. XLII.

<sup>238</sup> ThBasCaes, I, p. XV.

*Dictionnaire Grec Français* et le *Greek-English Lexicon* enregistrent par exemple les entrées συναγραυλέω et συναλύω. Leurs microstructures font état d'occurrences connues chez Denys d'Halicarnasse et Basile de Césarée, pour la première, chez Plutarque de Chéronée et Basile de Césarée, pour la seconde. Mais rien n'indique cependant que ces deux mots sont extrêmement rares, et que les occurrences signalées sont les seules connues dans les sources actuellement disponibles<sup>239</sup>.

## 2. Les formes présentant plusieurs graphies et les emprunts

§95. Les mots grecs en général, et les mots empruntés à d'autres langues (le latin dans la majorité des cas, mais aussi l'hébreu, les langues germaniques, le vieux-perse, les langues anatoliennes, etc.), en particulier, sont susceptibles de présenter des variations orthographiques. Ces variations ne justifient pas la création d'un nouveau lemme. Une série de critères permet de déterminer quelle graphie est susceptible de fournir la «forme de départ» la plus pertinente :

- l'ancienneté des occurrences connues ;
- la fréquence des occurrences connues ;
- la proximité formelle des occurrences connues avec leur étymon ;
- la productivité des occurrences connues.

Les deux premiers arguments sont de type fréquentiel. Ils restent fortement conditionnés par la nature lacunaire des sources et le hasard des attestations. Leur poids est plus faible que les deux suivants qui proposent une argumentation davantage formelle. Selon les cas, un seul aspect peut être mis en évidence, ou plusieurs, qui se combinent pour corroborer la formulation la plus pertinente du lemme : μασήσει, μασσήσει et μασσήσεως s.l. μάσησις, dérivé de μασάομαι, l'argument de la proximité formelle avec la base dérivationnelle est privilégié. Dans certains cas, aucun critère ne peut être mis en évidence. Le lemmatiseur se retranche alors derrière l'autorité d'un dictionnaire ou d'un linguiste : la graphie de φάλλαινα (κῆτος), avec un ou deux *lambda*, suit le *Dictionnaire étymologique de la langue grecque*<sup>240</sup>, tout comme celle du lemme

<sup>239</sup> Cfr, respectivement, E VIII, 1173, B 1835, L 1295, LS 1295 et S 1034 ; E VIII, 1202, B 1838, L 1295, LS 1694 et S 1034.

<sup>240</sup> C 1175, s.v. φάλλαινα.

σιβύνη connu sous les formes du type ζιβύνη et σιβίνη dans d'autres sources<sup>241</sup>.

[Page 264]

#### §96. Application des critères. Exemples.

- L'ethnique des Aquitains présente en grec les graphies du type Ἄκοιταν-, Ἄκουιταν-, Ἄκυτ- et Ἄκουιταν-. La première est rare et tardive<sup>242</sup>. La deuxième est plus fréquente et attestée antérieurement<sup>243</sup>. Elle présente des occurrences toponymiques en -ία et un dérivé adjectival en -ός. La troisième, quoique plus tardive, est aussi fréquente que la deuxième et présente des dérivés plus variés en -ία, -ός, et -ικός<sup>244</sup>. La dernière, Ἄκουιταν-, est moins fréquente que les deux précédentes<sup>245</sup>, mais elle présente des formes en -ία et -ός, des dérivés du type Ἄκουιτάνιος et Ἄκουιτανικός, et s'apparente à la graphie latine *Aquitāniā*<sup>246</sup>. Plus productives, les formes du type Ἄκουιταν- fourniront le lemme. Les critères de productivité et de proximité avec l'étymon ont prévalu.

- Le nom grec du Vésuve présente des graphies du type Βέβιος, Βέσβιος, Ούέσσοιος et Ούεσσοούιος. Le lemme est Ούέσσοιος, qui correspond aux graphies les plus anciennes et les plus proches du latin *Vesuvius*<sup>247</sup>. Les critères d'ancienneté et de proximité avec le mot dans sa langue d'origine ont été suivis.

- Le toponyme Ἐρμωνθις est aussi orthographié Ἐρμουθις. Les deux graphies se côtoient dans les sources, et sont productives (elles dérivent l'une et l'autre un ethnique, Ἐρμωνθίτης et Ἐρμουθίτης). La forme la plus ancienne, Ἐρμωνθις, fournit le lemme.

<sup>241</sup> C 1001, s.v. σιβύνη. En ce domaine, le sentiment des locuteurs demeure parfois impossible à établir. La forme toponymique Κιθαρίζων (cfr §19 et 37) était-elle comprise par Procope de Césarée comme un nom étranger hellénisé ou comme un toponyme grec sémantiquement relié à la cithare? Cfr aussi l'exemple du mot ἀνασπάστους connu chez Hérodote, forme fléchie d'ἀνασπάω ou emprunt à un étymon perse *anašpašta-*? (cfr G. TRAINA, C.A. CIANCAGLINI, *La forteresse de l'Oubli*, dans *Le Muséon*, 115 [2002], p. 411-412).

<sup>242</sup> L'interrogation du T.L.G. en fournit 2 occurrences, au V<sup>e</sup> s. ap. J.-C.

<sup>243</sup> L'interrogation du T.L.G. en fournit 28 occurrences, dès le II<sup>e</sup> s. av. J.-C.

<sup>244</sup> L'interrogation du T.L.G. en fournit 28 occurrences, dès le II<sup>e</sup> s. ap. J.-C.

<sup>245</sup> L'interrogation du T.L.G. en fournit 17 occurrences, à partir du II<sup>e</sup> s. ap. J.-C.

<sup>246</sup> *Le grand Gaffiot. Dictionnaire Latin Français*, ed. F. GAFFIOT, nouvelle édition revue et augmentée sous la dir. de P. FLOBERT, Paris, 2000, p. 153.

<sup>247</sup> ThProcCaes, p. XXVIII.

- La forme ἀννώνα, proche du latin *annōna* et productive (elle dérive, par ex., ἀννωνάριος, ἀννωνεύω), sert de lemme à ἀνόνας<sup>248</sup>. Les critères relatifs à l'étymologie et à la productivité ont été déterminants.

§97. Éventail de l'origine des emprunts :

- emprunts au *latin* : les sources patristiques et historiques byzantines, qui constituent la majorité des sources actuellement traitées (66,07% du total des occurrences, cfr §8), fourmillent d'emprunts au latin. Les lemmes grecs revêtent une forme proche de l'orthographe du mot latin : lemme ἀσηκρηῆτις pour le latin [Page 265] *a sēcrētis*<sup>249</sup> ; κίτατόριον s.l. κίτατόριον pour *citatōrium*<sup>250</sup> ; κωμέρκιον s.l. κομμέρκιον pour *commercium*<sup>251</sup> ;

- emprunts aux *langues germaniques* : les formes du type Τουτίλας attestées chez Procope de Césarée sont lemmatisées Τοτίλας, graphie rencontrée chez d'autres auteurs, dont Théophane le Confesseur, et conforme à l'étymologie germanique de l'anthroponyme<sup>252</sup> ;

- emprunts à l'*arabe* : les formes Βατμιζομανεῖς et Βαθημανεῖς (ethnique d'un peuple arabe) s.l. Βανιζομενεῖς, graphie attestée chez Diodore de Sicile et rendue relativement plus motivée<sup>253</sup> que les autres, par rapport à la langue d'origine, par la présence de l'élément *Banî-* «fils de» qui en arabe sert de premier terme au nom d'une tribu<sup>254</sup> ;

- emprunts au *vieux-perse* : de nombreux anthroponymes perses présentent plusieurs graphies correspondant parfois à des paradigmes flexionnels différents (cfr Ζαδέσπρας, cité §18). Les informations fournies par l'*Iranisches Namenbuch*<sup>255</sup> et par l'interrogation du *T.L.G.* montrent que l'anthroponyme *Julābzīn* est connu en grec sous les formes Ζαλαβζάν (indéclinable), Δαλαυζάν, Δολαύζαν, Δολαβζάν (pour ces trois formes, s'agit-il d'un accusatif

<sup>248</sup> ThTheoConf, I, p. XIX.

<sup>249</sup> ThTheoConf, I, p. XX. Pour les nombreuses graphies concurrentes en B- ou Oυ-, cfr ThTheoConf, I, p. XII.

<sup>250</sup> ThTheoConf, I, p. XXIII.

<sup>251</sup> ThTheoConf, I, p. XXIII.

<sup>252</sup> ThProcCaes, p. XXIX ; cfr aussi le lemme Ἄδηγις dans ThProcCaes, p. XXIV.

<sup>253</sup> Sur les notions de *motivation* et de *motivation relative*, d'origine saussurienne, cfr, par ex., SAUSSURE, *Cours*, p. 180-184 ; GAUDIN – GUESPIN, *Initiation*, p. 164 ; LEHMANN – MARTIN-BERTHET, *Introduction*, p. 105-107 ; MORTUREUX, *Lexicologie*, p. 20-21.

<sup>254</sup> Cfr *R.E.*, II, 2, IV (Stuttgart, 1896), col. 2848, s.v. *Banizomeneis*.

<sup>255</sup> JUSTI, *Namenbuch*, p. 8-9, s.v. *Ahura-Mazdāh* n° 24.

ou d'un indéclinable?) et Δόλβζα (gén. d'une forme Δόλβζας non attestée). Le lemme Ζαλαβζάν, plus proche du mot d'origine, regroupera toutes les formes ;

- emprunts aux langues anatoliennes : le lexique de Basile de Césarée présente de nombreux toponymes d'Asie Mineure qui ont fait l'objet de rapprochements avec d'éventuels étymons hittites, louvites et lyciens<sup>256</sup>.

### 3. Les lemmes défectifs et les formes supplétives

**§98.** Les lexicologues qualifient de *défectifs* les mots dont le paradigme flexionnel n'est pas complet<sup>257</sup>, et de *supplétifs* ceux qui complètent le paradigme [Page 266] flexionnel d'un autre mot<sup>258</sup>. En grec, les formes défectives et les cas de supplétion concernent les degrés des adjectifs et les radicaux verbaux. Dans les dictionnaires «papier» traditionnels, les premières ont une entrée lexicale correspondant aux formes attestées. Quant aux *supplétifs*, ils apparaissent soit sous l'entrée du paradigme dans lequel ils interviennent, soit sous celle dont ils relèvent eux-mêmes, soit, enfin, sous l'une et sous l'autre.

**§99.** Ces traitements traditionnels, dont l'utilisateur s'accommode, présentent des caractéristiques peu compatibles avec les traitements automatiques propres au *Projet* :

- des formes sont classées sous plusieurs entrées lexicales : ἐλάσσων, cité sous une entrée qui lui est propre (B 639 ; LS 528), apparaît aussi sous ἐλαχύς (B 639 ; LS 530) et sous μικρός (B 1283 ; LS 1133) ; ἐρεῖν apparaît sous λέγω (B 1175 ; LS 1033), εἶρω (B 491 ; LS 695) et ἐρῶ (B 813 ; LS 695) ;

- des formes sont classées sous des entrées hypothétiques : les entrées \*λῶις (B 1213) et \*χέρης (B 2134) permettent de classer respectivement les formes du type λῶων (citée comme entrée en B 1214) et χείρων (citée en B 2133) ; l'entrée \*εἶδω (B 584 ; LS 483) permet de classer les formes du type εἶδεται ;

- les renvois multiples ne sont pas toujours entièrement honorés : l'entrée ἥσσων est mentionnée comme comp. de κακός et de μικρός (LS 779), mais si κακός renvoie à ἥσσων (LS 863), aucun renvoi vers

<sup>256</sup> ThBasCaes, I, p. XXIII.

<sup>257</sup> DUBOIS, *Dictionnaire*, p. 131.

<sup>258</sup> DUBOIS, *Dictionnaire*, p. 458 ; LEHMANN – MARTIN-BERTHET, *Introduction*, p. 139-140 ; GAUDIN – GUESPIN, *Initiation*, p. 268.

ce dernier n'apparaît sous μικρός (LS 1133) ; les formes du type φέρτερος, φέρτατος et φέριστος sont mentionnées comme comp. et sup. d'ἀγαθός (B 5 ; LS 4) ; mais quand elles constituent elles-mêmes une entrée lexicale, le renvoi vers ἀγαθός est omis (B 2060 ; LS 1922)<sup>259</sup>.

**§100.** Les dictionnaires n'offrent donc pas une image cohérente de ces éléments du lexique. Dans le cadre du *Projet*, ces notions ont donc été réexaminées sur base de la définition stricte du lemme – considéré comme une «forme de départ» – et du *principe d'économie*. Traités selon la même grille d'analyse, les *défectifs* et les *supplétifs* sont décrits ici conjointement. La représentation proposée, quoiqu'encore imparfaite<sup>260</sup>, fournit une réponse pragmatique mieux appropriée aux objectifs de la lemmatisation.

Cette approche pragmatique s'articule en trois points :

1. Les formes possédant une «forme de départ» attestée et productive – un *positif* au nom. sing. pour les adj., un ind. prés. pour les verbes – sont lemmatisées [Page 267] selon les principes habituels, même si elles complètent par ailleurs le paradigme flexionnel d'un autre mot : ἐλάττων *s.l.* ἐλαχύς, même quand il fonctionne *in textu* comme le comparatif de μικρός ; ἔδεται *s.l.* ἔδω, ce lemme étant productif (et non sous ἐσθίω) ; mais εἰπών ne sera pas lemmatisé sous ἔπω, mot enregistré comme entrée dans certains dictionnaires mais qualifié de «récent et épique au présent»<sup>261</sup> ; εἶμι n'apparaîtra jamais sous ἔρχομαι.
2. Les formes ne possédant pas de «forme de départ», et qui, de plus, ne complètent pas un autre paradigme flexionnel, ont un lemme qui correspond aux formes attestées, ce lemme étant dit *défectif* : μεθελικέστερος, μέμονα.
3. Les formes ne possédant pas de «forme de départ», et qui, en outre, complètent le paradigme flexionnel d'un autre mot, sont dites *supplétives*, leur lemme étant celui du paradigme complété ; ἀμείνων

<sup>259</sup> Dans ce dernier cas, les renvois sont honorés dans le *Dictionnaire étymologique de la langue grecque*, cfr C 6, s.v. ἀγαθός et C 1188, s.v. φέρτερος.

<sup>260</sup> Par exemple, dans le cadre de l'analyse dérivationnelle, mentionnée **§9, 3**, le traitement appliqué aux formes supplétives aboutit à présenter ces dernières sous un lemme auquel elles ne sont en aucun cas apparentées morphologiquement.

<sup>261</sup> B 794 ; LS 678.

et φέρτεροι *s.l.* ἀγαθός<sup>262</sup>; δραμοῦμαι *s.l.* τρέχω; εἶπον *s.l.* λέγω; οἴσετε *s.l.* φέρω, etc.

Dans une telle description, la *supplétion* relève donc des formes, la *défectivité* se situe au niveau des lemmes. Dans le *D.A.G.*, un indice spécial est attaché aux *lemmes défectifs* et *formes supplétives*.

**§101.** Les comp. et sup. qui n'ont pas un positif pour «forme de départ», ainsi que les comp. et sup. dérivés d'un lemme appartenant à une autre classe morpho-syntaxique, sont considérés comme défectifs. Sont comptés dans cette catégorie :

- des comp. et sup. dont le positif n'est pas attesté : μεθελικέστεροι *s.l.* μεθελικέστερος (et non \*μεθῆλιξ)<sup>263</sup> ;
- des comp. et sup. dérivés d'un subst. : βασιλεύτατε *s.l.* βασιλεύτερος (et non \*βασιλεύς), ὀπλοτέρω et ὀπλοτάτης *s.l.* ὀπλότερος (et non \*ὄπλον) ;
- des comp. et sup. dérivés d'un part. : ἐρρωμενέστερον *s.l.* ἐρρωμενέστερος (et non \*ῥώννυμι) ;
- des comp. et sup. dérivés d'adv. ou de prép. (ἄνω [ἀνα], ἐγγύς, ἔνδον, ἔξω, πρόσω, ὑπέρ) : ἀνωτέρους *s.l.* ἀνώτερος, ἐγγυτάτην *s.l.* ἐγγύτερος ; ὑπερτάτους *s.l.* ὑπέρτερος.

**§102.** L'approche pragmatique proposée pour les cas de supplétion adjectivale repose sur deux observations et un principe :

- les formes adjectivales supplétives constituent une série close et systématisée en synchronie ; à l'époque classique, ἄριστος fonctionne comme le superlatif de ἀγαθός<sup>264</sup> ;

[Page 268]

- en diachronie, les formes supplétives entrent progressivement en concurrence avec des formations régulières, cfr les formes du type ἄριστος vs ἀγαθωτάτην enregistrées dans le *D.A.G.*<sup>265</sup> ;

<sup>262</sup> Même si, historiquement, ἀμείνων pourrait être «un "positif" entré dans le système du comparatif» ; C 74, s.v. ἀμείνων.

<sup>263</sup> *Contra* L 838, cfr ThAstFirm, p. XVII.

<sup>264</sup> CHANTRAINE, *Morphologie*, p. 110, §116.

<sup>265</sup> En grec moderne, les formations dérivationnelles en -τερος, -τατος et -ίων, -ιστος, etc., sont progressivement délaissées, les degrés de comparaison sont alors rendus par des tournures périphrastiques ; JANNARIS, *Grammar*, p. 144, §483-491 ; KALITSUNAKIS, *Grammatik der neugriechischen Volkssprache* (Götschen, 756 et 756a), Berlin, 1963, p. 78, §57-58.

- la constitution de la nomenclature du *D.A.G.* reste soumise à un *principe d'économie* qui évite de multiplier les entrées.

**§103.** Les *formes supplétives des comp. et sup.* sont donc classées sous le lemme adjectival dont elles complètent le paradigme<sup>266</sup> : les formes fléchies de ἀμείνων, ἄριστος, ἀρείων, βελτίων, κρείσσων<sup>267</sup> et λῶων *s.l.* ἀγαθός, tout comme les formes du type φέρτερος, φέριστος et φέρτατος qui en complètent également le paradigme<sup>268</sup> (et non sous φερτός, *ni* φέρτερος, *ni* φέρω)<sup>269</sup> ; ἥσων *s.l.* μικρός (et non sous l'adv. ἦκα) ; μήκιστον *s.l.* μακρός (et non sous le subst. μήκος) ; etc.

**§104.** Les *lemmes defectifs verbaux* ont un lemme correspondant aux formes attestées :

- un inf. aor.<sup>270</sup> : ἐντραγεῖν, καταπεφνεῖν, πορεῖν, τετμεῖν (et ses composés) ; ἀποσκληῖναι, τληῖναι (et ses composés) ;

- un ind. parf. : ἔοικα et ses dérivés, μέμονα, οἶδα et ses dérivés, πέπρωται (*v. impers.* «il est marqué par le destin»<sup>271</sup>), τέθηπα ;

- un participe : δυσμενέοντες *s.l.* δυσμενέων (et non \*δυσμενέω ; B 552, et LS 548)<sup>272</sup> ; κεκαφηότι *s.l.* κεκαφηώς ; ὑπερμενέοντα *s.l.* ὑπερμενέων (et non \*ὑπερμενέω, *cfr* LS 1866).

**§105.** L'approche pragmatique des verbes supplétifs se fonde sur les mêmes présupposés que les cas de supplétion adjectivale :

- les verbes faisant appel à la supplétion constituent une série close et systématisée en synchronie<sup>273</sup> ;

[Page 269]

<sup>266</sup> Pour un traitement identique, *cfr* aussi PIRENNE-DELFORGE – PURNELLE, *Index*, p. VII ; BODSON, *Index*, p. VI.

<sup>267</sup> Sur la forme κάρρων, *cfr* §29,3.

<sup>268</sup> C 6, *s.v.* ἀγαθός.

<sup>269</sup> CHANTRAINE, *Morphologie*, p. 116, §126 ; l'hypothèse posant fertov" comme positif, qui implique l'haplogie de φερ(τό)-τερος, est écartée dans C 1188.

<sup>270</sup> L'inf. est privilégié à l'ind. pour éviter l'éloignement, dû à l'augment, de formes apparentées présentées dans un index alphabétique.

<sup>271</sup> *Cfr* B 1515 ; *s.v.* πορεῖν dans C 928.

<sup>272</sup> ThAstFirm, p. XVI.

<sup>273</sup> SCHWYZER, *Grammatik*, II, p. 257-258 ; CHANTRAINE, *Morphologie*, p. 153-154, §171 et p. 157-158, §176 ; Y. Duhoux, *Le verbe grec ancien. Éléments de morphologie et de syntaxe historiques (Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain, 104)*, 2<sup>e</sup> éd., Louvain-la-Neuve, 2000, p. 67-69.



- en diachronie, les formes supplétives d'un paradigme entrent progressivement en concurrence avec des formations régulières, cfr les formes du type εἶπες vs ἔλεξας<sup>274</sup>.

- la constitution de la nomenclature du D.A.G. reste soumise à un principe d'économie qui évite de multiplier les entrées.

**§106.** Les formes verbales supplétives sont donc lemmatisées :

- sous un lemme propre quand elles possèdent une «forme de départ» productive : λέγω et ἐρώ, ἐσθίω et ἔδω, εἶμι et ἔρχομαι sont donc dégroupés ;

- dans le cas contraire sous le lemme dont elles complètent le paradigme : εἶπον s.l. λέγω, φαγεῖν s.l. ἐσθίω, ἦλθον s.l. ἔρχομαι, ὄπωπα s.l. ὀράω, etc.

**§107.** Les formes conjuguées sur le radical de εἶπεῖν (servant d'aor. tant à λέγω qu'à φημί et ἀγορεύω<sup>275</sup>) sont arbitrairement classées sous λέγω, en vertu toujours du principe d'économie. Pour les composés en -ειπεῖν, l'usage traditionnel des dictionnaires, qui distinguent des dérivés de λέγω, des dérivés de ἀγορεύω et des lemmes propres en -ειπεῖν, a été suivi, même s'il est conditionné par des facteurs sémantiques (cfr par ex. διαλέγω «mettre à part, choisir [...], parler, expliquer» vs διεπιπεῖν «dire en détail» dans B 476 et 503). Le D.A.G. retient donc les lemmes suivants :

- les formes du type ἀντειπεῖν et ἐπειπεῖν sont classées respectivement sous ἀντιλέγω (B 173 ; LS 150) et ἐπιλέγω (B 729 ; LS 614) ;

- les formes composées en ἀν-, ἀπ-, ἐξ-, κατα-, προ-, προσ-, συν- et ὑπ- sous le composé correspondant de ἀγορεύω ;

- le v. διεπιπεῖν possède un lemme qui lui est propre (B 503 ; LS 423)<sup>276</sup>.

<sup>274</sup> Le grec moderne ne conserve que quatre paradigmes faisant appel à la supplétion: βρέπω/εἶδα, ἔρχομαι/ἦρθα, λέ(γ)ω/εἶπα et τρώ(γ)ω/ἔφαγα, cfr CHANTRAINE, *Morphologie*, p. 154, §171 ; A. MIRAMBEL, *La langue grecque moderne* (Collection linguistique publiée par la Société de linguistique de Paris, 59), Paris, 1959, p. 173-174.

<sup>275</sup> H. FOURNIER, *Les verbes «dire» en grec ancien (exemple de conjugaison supplétive)*(Collection linguistique publiée par la Société de linguistique de Paris, 51), Paris, 1946, p. 1.

<sup>276</sup> Ce mode de traitement rejoint, dans ses grandes lignes, celui présenté aussi dans PIRENNE-DELFORGE – PURNELLE, *Index*, p. VII-VIII.

#### 4. Les formes et locutions lexicalisées

**§108.** La lexicalisation est le processus par lequel une innovation observée en discours confirme la création d'une unité lexicale au niveau de la langue et, par la suite, justifie la création d'une entrée dans les dictionnaires<sup>277</sup>. L'évolution [Page 270] progressive d'une locution comme ἐξ οὐρίας πνοῆς en ἐξουρίας, attestée dans les sources et enregistrée dans les dictionnaires anciens et modernes<sup>278</sup>, en constitue un exemple. La forme ἐταυτολόγησα est enregistrée sous ταῦτολογέω, et non \*τὸ αὐτὸ λογέω, la place de l'augment montrant que la forme est comprise comme une seule unité lexicale même si le texte des éditions conserve la *coronis*. Les formes du lemme καλοκάγαθία sont traitées de la même manière. Les formes et locutions lexicalisées justifient donc la création d'un nouveau lemme<sup>279</sup> : κατάκρας et κατάκραν ont un lemme propre et non \*κατ' ἄκρας ou κατ' ἄκραν, ni \*κατα et \*ἄκρα ; cfr aussi διόλου, ἐπένεικα, ἐπιπολύ, ἐπίσης, καθόλου, καταμόνας, καταπολύ, κατευθύ, μεθημέραν, παραπολύ, προτοῦ, προὔργου, etc.

#### 5. Les fautes d'édition

**§109.** Les formes non reconnues à l'issue de la lemmatisation automatique sont :

- soit des formes correctes jamais rencontrées dans les analyses antérieures, et pour lesquelles le lemme existe, ou non<sup>280</sup> ;
- soit des formes fautives (les erreurs relevant de l'orthographe d'usage, de l'accentuation, ou de coquilles typographiques) corrigées avant ou pendant l'élaboration des concordances lemmatisées : μήποτε pour μήποτε, διαίως pour δικαίως, exemples tirés des 193 corrections proposées dans le *Thesaurus* de Basile de Césarée<sup>281</sup>. Dans ce sens, la lemmatisation contribue aussi à améliorer les textes reçus<sup>282</sup>.

<sup>277</sup> CORBIN, *Morphologie*, p. 36-38 ; MORTUREUX, *Lexicologie*, p. 123-125.

<sup>278</sup> Cfr ThBasCaes, I, p. XXXIX.

<sup>279</sup> ThTheoConf, I, p. XIII ; ThAstFirm, p. XVI ; ThBasCaes, I, p. XXXIX.

<sup>280</sup> Sur des formes communes absentes du *D.A.G.*, cfr §11.

<sup>281</sup> ThBasCaes, I, p. XVII-XXII. La lemmatisation des homélies d'Astérius d'Amasée a permis de suggérer 134 corrections ; celle des lettres de Firmus de Césarée, 17 (ThAstFirm, p. XI-XIII). La concordance lemmatisée de la *Bibliothèque* de Photios présentera la liste, encore provisoire, des 227 corrections proposées.

<sup>282</sup> ThGrNaz, I, p. XI ; COULIE, *Lemmatisation*, p. 41-42 ; ThAstFirm, p. IX ; ThBasCaes, I, p. XVI.

**§110.** Contrairement à ce qui se rencontre dans les dictionnaires traditionnels, le *D.A.G.* ne retient pas comme entrée lexicale un mot caractérisé comme fautif et devant être corrigé ou rayé du lexique :

- E V 1060 : καταδιδάξας est corrigé en καταδείξας ;
- LS 1033, s.v. \*λέγω (A) : «erroneously inferred from λέξομαι, ἔλεκτο, etc.» ;
- S 1005, s.v. σπουδικόν : «σπουδικόν corrupt for δεσποτικόν».

[Page 271]

## 6. Les lettres, les phonèmes et les morphèmes

**§111.** Certains auteurs (les grammairiens mais aussi certains Pères et historiens) évoquent les lettres grecques, des syllabes ou des ensembles de caractères correspondant à ce que les linguistes modernes appellent phonèmes et morphèmes. Ces éléments du lexique sont traités dans le *D.A.G.*

- une lettre grecque citée en toutes lettres reçoit son nom comme lemme : ἄλφα s.l. ἄλφα ; βῆτα s.l. βῆτα ; γάμμα s.l. γάμμα ; μῦ s.l. μῦ (μ)<sup>283</sup> ; νῦ s.l. νῦ ; φεῖ, χεῖ, ψεῖ s.l. φεῖ, χεῖ, ψεῖ ;
- une lettre grecque seule sert de lemme à elle-même : α s.l. α ; γ s.l. γ ; η s.l. η ; ο s.l. ο ; σ s.l. σ ; χ s.l. χ ; ω s.l. ω ;
- une syllabe sert de lemme à elle-même : το et τε s.l. το et τε ; υξ s.l. υξ ;
- sur les lettres notant les numéraux, cfr **§66-71**.
- un *phonème* sert de lemme à lui-même : αυ s.l. αυ ; ει et εἰ s.l. εἰ ;
- un *morphème* sert de lemme à lui-même : δυς s.l. δυς<sup>284</sup> ; νη et νε s.l. νη (suffixe privatif)<sup>285</sup>.

<sup>283</sup> La spécification distingue la lettre de l'onomatopée μῦ (μύ), cfr B 1301.

<sup>284</sup> Basile le Minime, *Commentaire*, 74, 7 (= SCHMIDT, p. 58, l. 5).

<sup>285</sup> C 732, s.v. ν-, νε- ; la graphie du lemme (nh) correspond à la forme la plus productive.

## CONCLUSIONS

Les principes formulés devront désormais servir de référentiel à l'usage des concepteurs du *Projet*, de leurs collaborateurs et des utilisateurs des concordances lemmatisées – ces derniers l'appelaient d'ailleurs de leur vœu<sup>286</sup>. L'exercice de la lemmatisation ne se limite pas à établir des listes de mots. Telle que décrite ici, elle permet d'inscrire le vocabulaire d'un auteur dans le lexique de la langue ; il s'agit d'un travail de réflexion et d'analyse sur celle-ci. Aucune description, aussi précise soit-elle, n'épuisera cependant la richesse des ressources lexicales d'une langue. Ces principes demeurent donc provisoires, susceptibles d'être révisés, complétés ou corrigés selon les faits de langue rencontrés dans les sources et selon la technologie utilisée pour le traitement informatique appliqué à celles-ci. Les outils d'analyse actuellement en développement au CENTAL sont par exemple conçus sur un *D.A.G.* renouvelé qui ajoute aux lemmes des données relatives à leur classe morpho-syntaxique. Ces développements récents annoncent une révision de l'usage des spécifications, les éléments (προσ.), (τοπ.) et N- qui accompagnent les lemmes devenant inutiles. L'analyse de textes nouveaux, le perfectionnement de l'outil et l'enrichissement des principes d'analyse qui président à leur fonctionnement doivent donc se poursuivre.

Bastien KINDT  
Université catholique de Louvain  
Institut orientaliste  
Louvain-la-Neuve (Belgique)

<sup>286</sup> Cfr la recension du ThAstFirm par le Prof. Th. Schmidt dans *Le Muséon*, 115 (2002), p. 225.