

# Philologie et informatique : résultats et projets

Valeria LOMANTO et Nino MARINONE

**Abstract.** The following is a report on four computerized works about Latin language: a concordance to Symmachus, a concordance based on Keil's collection of the Latin grammarians, an index to Latin grammar texts (2nd–9th century), a project for automatic lemmatization of Latin. All of them have been carried out by Nino Marinone and Valeria Lomanto, both professors at Turin University, with the co-operation of the Istituto di Linguistica Computazionale in Pisa. The various problems are discussed and the accepted solutions are described.

**Keywords:** Latin language, ancient grammarians, Symmachus, concordances, classification methods, lexicography, lemmatization, writing and morphological spelling, word-listing, word-searching.

**Mots-clés :** Langue latine, grammairiens anciens, Symmaque, concordances, méthodes de classification, lexicographie, lemmatisation, altérations graphico-morphologiques, listing, recherche des mots.

## 1. Origine des initiatives du groupe de recherche turinois

En 1975, sollicités par un groupe de professeurs d'universités italiennes réunis pour un séminaire ayant pour thème «Les Grammairiens latins de l'époque impériale»<sup>1</sup>, nous avons mis en chantier un projet pour la rédaction des concordances des grammairiens latins en utilisant les procédés électroniques. Comme base, nous avons adopté les sept volumes publiés par H. Keil à Leipzig entre 1855 et 1880, en nous réservant de remplacer l'édition de Keil par les éditions disponibles plus récentes et les œuvres représentées partiellement ou sous forme de *specimen* dans le recueil de Keil par les éditions postérieures

---

<sup>1</sup> Les exposés ont été publiés en un volume de mélanges par l'Istituto di Filologia classica e medioevale de l'Université de Gênes en 1976.

---

✉ Università degli studi di Torino; Dipartimento di Filologia, Linguistica e Tradizione Classica;  
Via S. Ottavio, 20; I-10124 Torino (Italia).  
Fax : + 39 11 817 45 57

---

complètes. Malgré cette mise à jour, toujours en cours, étant donné que de nouvelles éditions de textes grammaticaux continuent de paraître<sup>2</sup>, la plus grande partie du matériel est constituée par le *corpus* établi par Keil selon des critères d'édition dépassés, et fondé sur un choix et sur une collation des manuscrits largement incomplets et pas toujours crédibles. Cette limite évidente n'a pas découragé notre initiative; elle en a au contraire souligné l'opportunité. En effet si l'on considère le caractère formel du langage grammatical, un instrument lexicographique, même construit sur des textes médiocres du point de vue philologique, pourrait fournir un support utile pour de nouvelles éditions, permettant d'orienter le choix entre les variantes ou d'intégrer les passages lacunaires d'une œuvre sur la base de la tradition manuscrite parallèle. Cette destination primordiale de notre travail a contribué à déterminer l'option pour la formule de la concordance sans l'élaboration de lemmes, certes moins raffinée que l'index lemmatisé ou le lexique, mais justement, pour cette raison, apte non seulement à ne pas compromettre l'interprétation des utilisateurs ou à mettre en évidence les passages parallèles, mais en quelque sorte imposée par la nature particulière des textes, où un grand nombre de formes est constitué par des graphèmes, des morphèmes, des segments de phrases correspondant à des unités métriques.

---

<sup>2</sup> On signale les éditions postérieures à celles qui sont indiquées à la p. 16 du fascicule «Concordanza dei Grammatici latini». Présentation de GRILLI (Alberto), MARINONE (Nino), LOMANTO (Valeria), ZAMPOLLI (Antonio), BROGNA (Domenico), FIOCCHI (Laura) : 1979, suppl. au vol. CXII des *Atti dell'Accademia delle Scienze di Torino. Classe di Scienze morali, storiche e filologiche* :

HOLTZ (Louis) : 1981, *Donat et la tradition de l'enseignement grammatical. Étude sur l'ars Donati et sa diffusion* (Paris : Editions du CNRS);

DE NONNO (Mario) : 1982, *La grammatica dell'Anonymus Bobiensis* : GL 1,533–565 (Roma : Edizioni di Storia e Letteratura);

MARIOTTI (Scevola) : 1984, «Il fragmentum Bobiense de nomine» (GL 7,540–544), *Il libro e il testo* (Urbino : Quattro Venti), pp. 37–68;

PASSALACQUA (Marina) : 1984, *Tre testi grammaticali bobbiesi* : GL 5,555–566.634–654; GL 4,207–216 (Roma : Edizioni di Storia e Letteratura);

PASSALACQUA (Marina) : 1987, *Prisciani Caesariensis opuscula : de figuris numerorum; de metris Terentii; praeexercitamina* (Roma : Edizioni di Storia e Letteratura);

DE NONNO (Mario) : 1988, «Traduzione e diffusione di Mario Vittorino grammatico, con edizione degli excerpta de orthographia», *Rivista di Filologia e di Istruzione classica* CXVI, pp. 5–59;

DE PAOLIS (Paolo) : 1990, *Macrobiani Theodosii de uerborum Graeci et Latini differentiis uel societatis excerpta* (Urbino : Quattro Venti);

PASSALACQUA (Marina) : 1992, *Prisciani Caesariensis institutio de nomine et pronomine et uerbo* (Urbino : Quattro Venti).

## 2. La concordance de Symmaque<sup>3</sup>

Afin de mettre au point le projet relatif aux grammairiens sur un texte plus court et surtout beaucoup moins complexe que les *artes* de la basse antiquité, le programme de concordance a été expérimenté sur Symmaque dans l'édition de O. Seeck, la seule aujourd'hui encore qui soit complète<sup>4</sup>. La fonction pour ainsi dire propédeutique attribuée au travail sur Symmaque a déterminé l'organisation des données sous forme de concordance sans lemmes, alors qu'il aurait été plus conforme à la nature de cette œuvre d'en établir un index lemmatisé ou un lexique; ladite fonction a de même suggéré la façon d'établir le contexte.

Deux possibilités se présentaient : ou bien découper le contexte sur la base de la ponctuation, de sorte que le mot-vedette se trouverait à chaque fois en un point différent de la ligne, mais aurait plus de probabilités d'être interprété correctement, ou bien découper le contexte de façon mécanique, de sorte que le mot-vedette serait immédiatement reconnaissable au centre de la ligne. La première solution non seulement est plutôt aventureuse sur le plan de l'interprétation<sup>5</sup>, mais elle contribue aussi à conférer à la concordance une sorte d'autonomie illusoire en faisant apparaître superflu le recours au texte; elle a donc été mise de côté en faveur du système KWIC (*keyword in context*). Les mots-vedettes y sont disposés en ordre alphabétique au centre d'un contexte qui peut atteindre 150–160 caractères, se conformant, du moins dans la plupart des cas, à l'analyse morphologique, et non pas syntaxique et sémantique du *keyword*. Les occurrences du même mot-vedette se suivent selon l'ordre alphabétique des paroles qui constituent la partie droite du contexte : cette disposition permet de mettre en évidence les *iuncturae* qui reviennent, ayant chez Symmaque une certaine fréquence<sup>6</sup> et une fréquence très élevée

<sup>3</sup> Cf. LOMANTO (Valeria) : 1983, *Concordantiae in Q. Aurelii Symmachi opera* (Hildesheim – Zürich – New York : Olms).

<sup>4</sup> Cf. SEECK (Otto) : 1883, *Q. Aurelii Symmachi quae supersunt* (Berlin : « Monumenta Germaniae Historica » VI.1). Entre-temps deux volumes des lettres de Symmaque ont été publiés, contenant les livres I–IV, par CALLU (J.P.) dans la collection « Les Belles Lettres », Paris 1972 et 1982.

<sup>5</sup> Sur le peu de crédibilité de la ponctuation, qui, dans les éditions de textes antiques, obéit à des critères logico-syntaxiques, plutôt qu'à des critères métriques et rythmiques, cf. HAVET (Louis) : 1892, *La prose métrique de Symmaque et les origines métriques du cursus* (Paris), pp. 21 et suiv. et en particulier la version de la *relatio* 3 fournie par ce chercheur, pp. 24 et suiv.

<sup>6</sup> Cf. HAVET 1892, p. 4, note 2 : « Dans un écrivain quelconque, l'établissement du texte n'a un caractère définitif que quand il est aisé de comparer rapidement et sûrement les passages parallèles; il en est ainsi particulièrement chez Symmaque, qui revient à certaines tournures presque comme à des formules ».

chez les grammairiens. À la fin de la concordance, la liste des mots latins est suivie des numéraux, présentés en lettres majuscules selon l'usage romain, des mots grecs et des formes irrémédiablement corrompues.

La possibilité de remonter de la concordance au texte est garantie par une série d'indications alphanumériques dans la marge gauche de la page. Cette référence s'articule en deux sections, l'une relative aux subdivisions internes de l'œuvre, l'autre relative à l'édition sur laquelle est basée la concordance. Dans la section structurale de la référence les deux premières données renvoient aux livres des lettres (de 1 à 10), aux *relationes* (11), aux *orationes* (12); l'indication A1 correspond à la lettre dont l'auteur est le père de Symmaque (*epist.* 1,2); B1 désigne la lettre adressée à Symmaque par Ausone (*epist.* 1,32). Les trois données suivantes indiquent le numéro des lettres à l'intérieur de chaque livre, le numéro des *relationes* et des *orationes*; viennent ensuite deux indications correspondant au paragraphe. Deux points séparent la référence structurale de la référence topographique, désignant la page et la ligne de l'édition Seeck.

### 3. La concordance des Grammairiens Latins

Le système de références adopté pour les grammairiens est semblable à celui qui a été élaboré pour Symmaque, mais il diffère naturellement sur les contenus<sup>7</sup>. La première partie, relative au texte, s'articule en deux séries de sigles numériques :

- 1) un code de trois chiffres désigne les œuvres à l'intérieur du *corpus*. On n'a pas attribué de code d'auteur parce que, sur 104 *artes*, 32 sont anonymes et, parmi les 72 restantes, plusieurs présentent des problèmes complexes d'attribution que seules peut-être des études ultérieures permettront de résoudre;
- 2) un code de deux chiffres correspond à l'argument. Les arguments ont été isolés sur la base des classifications caractéristiques des *artes*, en évitant une subdivision trop analytique, qui parfois aurait été opportune, dans le but d'éviter une fragmentation excessive des textes. On a distingué les rubriques suivantes :

---

<sup>7</sup> Pour la description analytique du système de codification adopté cf. MARINONE in GRILLI etc. : 1979, pp. 13-19.

- |  |  |
|--|--|
| <ol style="list-style-type: none"> <li>1. <i>praefatio (et epilogus)</i></li> <li>2. <i>generalia et uaria</i></li> <li>3. <i>de litteris</i></li> <li>4. <i>de syllabis (etiam metrica ratione)</i></li> <li>5. <i>de accentibus</i></li> <li>6. <i>de distinctionibus</i></li> <li>7. <i>de orthographia</i></li> <li>8. <i>de oratione eiusque partibus</i></li> <li>9. <i>de nomine</i></li> <li>10. <i>de pronomine</i></li> <li>11. <i>de uerbo</i></li> <li>12. <i>de aduerbio</i></li> <li>13. <i>de participio</i></li> </ol> | <ol style="list-style-type: none"> <li>14. <i>de coniunctione</i></li> <li>15. <i>de praepositione</i></li> <li>16. <i>de interiectione</i></li> <li>17. <i>de uitiis et uirtutibus orationis</i></li> <li>18. <i>de idiomatibus, elocutionibus, differentiis, synonymis</i></li> <li>19. <i>de constructione uel syntaxi</i></li> <li>20. <i>de arte metrica (cum poemate, compositione, structura, musica)</i></li> <li>21. <i>de rhetorica</i></li> <li>22. <i>interpretationes grammaticae</i><sup>8</sup></li> <li>23. <i>de figuris numerorum</i><sup>9</sup></li> <li>24. <i>de notis.</i></li> </ol> |
|--|--|

La section topographique de la référence s'articule également en deux parties. La première fournit le numéro du volume de l'édition Keil ou l'initiale du nom de l'éditeur plus récent; là où c'est nécessaire, le chiffre ou la lettre sont suivis de minuscules indiquant que le mot-vedette appartient à un passage supprimé (« a ») ou emprunté à l'apparat (« b ») ou tiré de la préface (« c »), etc. La seconde partie renvoie à la page et à la ligne de l'édition utilisée<sup>10</sup>. L'adoption de la référence topographique au lieu de la référence structurale (par livre, chapitre, paragraphe) est imposée par la diversité d'extension et par la structure des œuvres : certaines d'entre elles sont des traités en plusieurs livres, d'autres sont des fragments sans subdivisions. Cette codification permet de sélectionner le matériel selon les exigences les plus diverses de la part des utilisateurs, par ex. par œuvres; par arguments; par œuvres isolées + arguments isolés; par groupes d'œuvres + groupes d'arguments; par œuvres isolées + groupes d'arguments; par groupes d'œuvres + arguments isolés, etc.

<sup>8</sup> Cette rubrique indique les *partitiones duodecim uersuum Aeneidos principalium* de Priscien (GL 3,459–515), œuvre sans terme de comparaison dans le *corpus*, plus similaire aux commentaires de Donat et de Servius à Térence et à Virgile qu'aux *artes*, bien que destinée non à l'exégèse du texte, mais à l'apprentissage de la structure de la langue à partir d'un texte.

<sup>9</sup> Cette rubrique aussi ne correspond qu'à l'opuscule homonyme de Priscien (GL 3,405–417 = pp. 3–18 PASSALACQUA : 1987).

<sup>10</sup> Pour les éditions d'Agroecius (PUGLIARELLO (Mariarosa) : 1978, Milano) et de Arusianus Messius (DELLA CASA (Adriana) : 1977, Milano), le numéro de la page est remplacé par le numéro du lemme; pour l'édition du *fragmentum Bobiense de nomine* (MARIOTTI (Scevola) : 1984, Urbino) le numéro du lemme remplace le numéro de la ligne.

#### 4. *L'Index Grammaticus*<sup>11</sup>

En considération du caractère de *work in progress* de la concordance, des dimensions qu'aurait atteintes sa publication et surtout de la possibilité de consultation vidéo des textes enregistrés même sur ordinateur personnel, d'ailleurs non prévue quand le travail a débuté, nous avons estimé opportun de fournir aux utilisateurs un index, ou plutôt une liste des occurrences dans les textes grammaticaux. Les formes latines et grecques y sont enregistrées par ordre alphabétique et accompagnées de l'indication des lieux si la fréquence est inférieure à 100 pour les formes latines, inférieure à 20 pour les formes grecques. Nous sommes conscients de l'arbitraire de ce procédé; il a toutefois été adopté pour éviter d'autres inconvénients. L'indication complète des lieux aurait conféré à l'index une ampleur proche de celle de la concordance, mais sans fournir le contexte, et en aurait rendu la consultation peu aisée; la sélection des termes techniques en serait résultée inévitablement subjective et aurait sacrifié de façon systématique l'exemplification, caractéristique des textes grammaticaux; et surtout l'œuvre est destinée moins à des chercheurs se proposant d'analyser la terminologie grammaticale qu'à ceux qui auraient pour but d'examiner les formes rares ou reconstruites par les grammairiens, très nombreuses dans les *artes* : pour cette catégorie d'utilisateurs l'index peut représenter un point de départ efficace. L'index des formes en ordre alphabétique direct est suivi d'un index des formes en ordre inverse, comprenant aussi une section latine et une section grecque et muni des fréquences seulement.

Les modifications au texte ont été limitées à l'essentiel.

- La majuscule a été réservée aux noms propres pour en permettre la sélection<sup>12</sup>.
- En latin :
  - 1) pour *i* et *u* semi-consonnes on a adopté les signes vocaliques ⟨i⟩ et ⟨u⟩ conformément à l'usage antérieur à la réforme de Ramus;
  - 2) les abréviations ont été explicitées entre parenthèses;
  - 3) les enclitiques *-ne*, *-que*, *-ue* sont introduits séparément;
  - 4) les segments de paroles qui correspondent à des unités métriques ont été recomposés en entrées lexicales.

<sup>11</sup> Cf. LOMANTO (Valeria) et MARINONE (Nino) : 1990, *Index Grammaticus*, 3 vol. (Hildesheim - Zürich - New York : Olms).

<sup>12</sup> Pour les noms qui peuvent être ou communs ou propres, la lettre initiale de la forme avec majuscule est indiquée à côté de la forme en minuscules.

-- Pour le grec :

- 1) les différentes graphies d'une même forme ont été unifiées [par ex. ἐξ et ἐκ];
- 2) les voyelles élidées ont été réintroduites entre parenthèses et, dans le cas d'altération phonico-graphique, on a inséré un renvoi [par ex. ἔσθ' → ἐστ(ι)(ν)];
- 3) les cas de crase avec καὶ et τὸ ont été explicités et on a introduit un renvoi [par ex. καὶθηνις → ἸΑθηνις], tandis qu'on a conservé les crases en ταὐτ- et en προὐ-;
- 4) les enclitiques et les proclitiques ont toujours été enregistrés comme des paroles atones, tandis que les formes orthotoniques des prépositions ont été insérées à part quand elles ont une fonction différente des proclitiques correspondants [par ex. ἐπι est distingué de ἐπὶ, ἐπ', ἐφ'];
- 5) on n'a pas fait de distinction pour les formes monosyllabiques τίς/τις et τὶ/τι du fait que les formes atones peuvent être accentuées en cas d'enclises.

## 5. Pré-édition des textes<sup>13</sup>

Avant d'enregistrer les textes on a procédé à une révision formelle soit pour en éliminer les dissemblances soit pour les conformer aux critères de la philologie textuelle moderne. Nous avons déjà cité l'élimination des lettres de Ramus, l'explicitation des écritures en abrégé ou par acronymes<sup>14</sup>, la séparation des enclitiques les plus fréquents *-ne*, *-que*, *-ue*, la limitation des majuscules aux théonymes, anthroponymes, ethniques et leurs dérivés, aux noms de mois et de fêtes. En outre

<sup>13</sup> Pour la description analytique du procédé utilisé, accompagnée systématiquement d'exemples, cf. pour Symmaque les pp. VII–IX de la préface à la concordance; pour les grammairiens, LOMANTO in Grilli etc. : 1979, pp. 21–31 et 1990 : « A Concordance to Keil's Latin Grammarians », *Computers and the Humanities*, XXIV, pp. 427–435.

<sup>14</sup> Tandis que dans Symmaque les abréviations concernent presque exclusivement les dates et les titres officiels, dans les grammairiens elles sont utilisées soit pour les *praenomina*, soit pour les paroles qui figurent en acronymes dans une citation [par ex. Diomède GL 1,392,15 = Virgile, *Aen.* 1,687 : « cum dabit amplexus a(t)que o(sc)ula d(ul)cia f(iget) »] ou en abrégé [par ex. *fragmentum Lauantinum* GL 5,326,14 : ... ut est octo et rel(iqua)].

- 5.1. dans Symmaque<sup>15</sup> on a supprimé les en-têtes, souvent manquants et parfois douteux, de lettres, *relationes* et *orationes*; dans les grammairiens<sup>16</sup> on a éliminé toutes les indications étrangères au texte, et notamment :
- 5.1.1. la numérotation des livres, des chapitres et des paragraphes;
  - 5.1.2. la barre simple ou double qui indique la fin de page d'éditions précédentes;
  - 5.1.3. les renvois internes ou à une autre œuvre;
  - 5.1.4. les indications accompagnant les citations d'auteur;
  - 5.1.5. les tirets isolés<sup>17</sup> qui opposent une forme erronée à une forme correcte<sup>18</sup> ou qui indiquent le passage à un autre sujet<sup>19</sup> ou qui remplacent un lemme<sup>20</sup>;
  - 5.1.6. les variantes signalées entre parenthèses<sup>21</sup>.
- 5.2. quant à la ponctuation, dans GL on a rectifié quelques aberrations dans les parties grecques<sup>22</sup>; en outre dans S comme dans GL les tirets, dont la fonction d'ouverture ou de fermeture n'est pas claire, ont été remplacés par des parenthèses quand ils désignent des incises, par des virgules quand ils désignent des mots isolés.
- 5.3. dans S comme dans GL, les titres des œuvres citées dans les textes sont en italique; dans GL les titres des livres ou des chapitres ou des sections sont mis entre barres doubles<sup>23</sup>.
- 5.4. dans S comme dans GL, les citations d'auteur sont indiquées par des guillemets et, dans le cas de citations poétiques, la fin de chaque vers est marquée par une barre<sup>24</sup>.

<sup>15</sup> Désormais indiqué dans le texte par S.

<sup>16</sup> Désormais indiqués dans le texte par GL.

<sup>17</sup> Cf. par ex. Cassiodore *orth.* J 45,954 : *quoties, toties, septies — sine n.*

<sup>18</sup> Cf. par ex. *dub. nom.* G 774,278 : ... *eburnea — non eborea — dicendum.*

<sup>19</sup> Cf. par ex. Clédonius GL 5,61,16s. : *accuso te et inuoco te : et in simplici et in composito ipsum casum tenet. — soleo caret participio praesenti...*

<sup>20</sup> Dans l'édition Glorie du *de dubiis nominibus* le tiret introduit systématiquement les paragraphes dans lesquels continue la discussion d'un lemme énoncé dans un paragraphe précédent.

<sup>21</sup> Cf. par ex. *explan.* GL 4b,549,26 : ... *cum osi erimus eritis erunt* (corr. *erint*).

<sup>22</sup> Par ex. dans Priscien *inst.* GL 3,295,4 nous avons remplacé les deux points par le point en position haussée : Ἴσοκράτης παραίνεσεν· οὕτω γὰρ...

<sup>23</sup> Les titres des différentes œuvres de S et de GL sont remplacés par leur code numérique respectif.

<sup>24</sup> Dans S les citations d'auteurs sont signalées en général par l'espacement; dans GL en plus de l'espacement on a utilisé l'italique, le renvoi à la ligne, les virgules hautes. Dans GL les virgules



- 5.5. dans GL on a unifié la scansion métrique moyennant l'emploi de points qui remplacent la division des mots ou les apex placés en dessous<sup>25</sup>.
- 5.6. dans S comme dans GL, on a adopté les signes diacritiques courants : [ ] pour les suppressions; ( ) pour les intégrations, au lieu de l'italique utilisé dans les éditions plus anciennes; \* pour les lacunes, au lieu des points de suspension ou, dans GL, au lieu de formules telles que *aliquid desideratur*, *deest folium unum* etc.; † pour les *loci desperati*.
- 5.7. dans GL les omissions dues à l'éditeur sont indiquées moyennant [...] qui remplace les tirets, précédés parfois de etc. ou de κτλ. ou de quelque autre signal ayant un sens analogue. Si une œuvre n'est publiée que sous forme d'extraits, chacun d'eux est précédé de [...] et suivi de [...] ||.
- 5.8. dans S les parties des *orationes*, considérées comme des rédactions alternatives et qui dans l'édition Seeck se présentent en corps plus petit entre barres doubles, sont mises entre accolades. En outre toutes les formes pour lesquelles sont signalées des variantes ou des omissions dans la tradition manuscrite ou au sujet desquelles l'éditeur ou d'autres philologues ont opéré ou proposé des interventions d'une nature quelconque sont indiquées par 1 en position plus haute.

## 6. Les variantes graphiques

La concordance de Symmaque réalisée par l'ordinateur selon notre programme et après la révision formelle exposée dans les paragraphes précédents a été soumise à une élaboration ultérieure. Les interventions effectuées ont eu pour objet exclusivement la disposition des données et comme but de rendre la consultation plus aisée.

La graphie du texte présente une relative homogénéité, soit à cause de l'harmonie des manuscrits, soit grâce à l'intervention unificatrice de l'éditeur<sup>26</sup>; toutefois elle n'est pas sans discordances qu'il aurait été arbitraire d'éliminer,

---

hautes ont été supprimées, quand elles signalent des formes isolées qui illustrent une norme grammaticale; elles ont été remplacées par des traits d'union quand elles indiquent des segments de paroles.

<sup>25</sup> Pour le recours à la division des mots, cf. Priscien *part.* GL 3,461,15 : *armaui rumqueca noTro iaequi primusab oris* (= *arma ui.rumque ca.no Tro.iae qui. primus ab. oris*); pour le recours aux signes graphiques au-dessous des lettres, cf. Diomède GL 1,519,2 *Maece,nas atauis, edite re,gibus* (= *Maece.nas atauis. edite re.gibus*).

<sup>26</sup> On peut encore partager l'opinion de HAVET qui remarque (1892 : p. 46, note 2) : « M. Seeck sacrifie trop facilement à l'orthographe moderne » et qui fournit quelques exemples.

mais qui auraient comporté une localisation des variantes dans des zones de la concordance parfois très éloignées. Devant chaque forme l'utilisateur aurait dû se demander s'il en existe des allographes et n'aurait pas pu effectuer une comparaison immédiate. Pour remédier à cet inconvénient, les variantes, scrupuleusement respectées, ont été rassemblées selon une des graphies adoptées dans le texte et des renvois opportuns ont été introduits pour faciliter leur recherche. Les cas intéressés concernent les phénomènes suivants<sup>27</sup>.

- 6.1. Les formes avec préverbe, tantôt assimilées tantôt dissimilées dans l'édition, sont disposées dans la concordance selon la graphie assimilée quand le préverbe est suivi d'une occlusive, selon la graphie dissimilée quand le préverbe est suivi d'une fricative, d'une nasale ou d'une liquide.
- 6.2. Les formes pronominales redoublées ou suffixées : *quamdam* (p. 339,1) est rangé selon la graphie assimilée par analogie avec les cas où la nasale s'assimile à la dentale du suffixe; les formes assimilées *quicquam* (pp. 55,10; 137,31; 181,13) et *quicquid* (p. 189,28) sont rangées parmi les occurrences beaucoup plus nombreuses des formes non assimilées *quidquam* et *quidquid*.
- 6.3. Les oscillations entre
  - 6.3.1. sourde et sonore : *aliut, aput, haut, set, uelud, adque* sont rangés parmi les occurrences de *aliud, apud, haud, sed, uelut, atque*;
  - 6.3.2. sourde et aspirée : *trophaeis* (p. 266,31) est rangé parmi les formes non aspirées plus fréquentes;
  - 6.3.3. aspirée et fricative : *triumfales* (p. 322,10) *triumfalia* (p. 324,9) *triumfum* (pp. 322,19 et 329,33) sont rangés parmi les formes beaucoup plus nombreuses avec la graphie *triumph-*;
  - 6.3.4. vélaire et labiovélaire : les forme de *urguere* sont rangées parmi celles de *urgere*;
  - 6.3.5. formes aspirées et non aspirées : *arenae* (p. 178,28) est rangé parmi les formes avec aspiration, tandis que *hauet* (p. 175,10) est rangé parmi les formes sans aspiration. En outre *nil* est rangé parmi les occurrences de *nihil, corte* (p. 244,11) parmi celles de *cohors, cohercet* (p. 326,33) parmi celles de *coercere*;
  - 6.3.6. formes simples et géminées : les occurrences de *cotidie* et de *cottidie* sont regroupées; *litis* (p. 23,26) est rangé parmi les formes avec géminée; *oblitterant* (p. 133,24) et *obliterari* (p. 336,18) sont rangés l'un à la suite de l'autre; *opportuna* (pp. 13,8 et 333,31) est rangé parmi les formes avec la graphie simple; *rennuerunt* (p. 339,16)

---

<sup>27</sup> Pour un examen complet, cf. les pp. X-XVII de la préface à la concordance de Symmaque.

précède *renuis* (p. 227,3); *repperiet* (p. 298,35) est rangé parmi les formes d'inflectum avec graphie simple.

- 6.4. Les cas de métathèse : les formes de *accersere* sont rangées avec celles de *arcessere*.
- 6.5. Les changements de timbre :
  - 6.5.1. entre [a] et [e] : les formes de *aequiparare* sont rangées avec celles de *aequiperare*;
  - 6.5.2. entre [e] et [i] : *beniuolentiam* (pp. 123,27 et 194,19) est inséré parmi les formes avec *bene-*; *delinimenta* (p. 73,26) est rangé parmi les occurrences de *delenimenta*; *dimensione* (p. 43,12) *dimensionibus* (p. 161,7) *diuersor* (p. 219,18) *diuersorii* (p. 258,30) sont placés parmi les formes avec préverbe *de-*;
  - 6.5.3. entre [e] et [o] : les formes en (-)uort- sont placées parmi celles en (-)uert-;
  - 6.5.4. entre [o] et [u] : les formes de *adulescens* sont rangées parmi celles de *adolescentia* et *adolescere*; *epistolis* (p. 4,13) est placé parmi les formes de *epistula*; *Volcanus* (p. 7,5) suit immédiatement *Vulcani* (p. 328,38);
  - 6.5.5. entre [i], [u], [ü] : *bybliothecis* (p. 108,10) est rangé parmi les occurrences de *bibliotheca*; les formes *inclyt-* sont rangées avec les formes *inclut-*.
- 6.6. Les formes diphtonguées et monophthonguées : la mise en ordre est commune pour *glæba* (p. 328,14) *glebae* (pp. 119,25 et 315,18) et *glæbis* (p. 326,27); les occurrences de *caudex* sont rangées avec celles de *codex*; *promptario* (pp. 38,28 et 93,9) est placé à côté de *promptuario* (pp. 256,28 et 260,29); la forme *beneficientiae* (pp. 112,19 et 137,20), avec un *i* qui peut être considéré comme un indice de graphie étymologique ou de palatalisation, est insérée parmi les formes de *beneficent-*.
- 6.7. Les formes avec syncope de *domna domnus periculum uinclum* sont toujours insérées parmi les formes intactes.
- 6.8. Les formes avec apocope et les cas de phonsyntaxe : les variantes *atque* et *ac* sont mises ensemble, ainsi que *neque* et *nec*, *siue* et *seu*, *sicuti* et *sicut*, *ueluti* et *uelut*. Les occurrences de *a* et *abs* sont rangées avec *ab*; avec *ex* celles de *e*.
- 6.9. Les formes qui admettent une graphie unie ou divisée sont rangées selon la graphie divisée : le premier élément sert de *keyword*, le deuxième est considéré comme le premier mot du contexte. Les occurrences des formes avec double graphie sont donc regroupées de cette façon : par ex. *cumprimum* et *cum primum* sont situés dans la zone de concordance où le *keyword cum* est suivi de *primum* : la succession des passages suit comme

toujours l'ordre alphabétique du contexte de droite. Cette solution a été appliquée par analogie aux mots composés et juxtaposés qui, tout en ayant dans le texte la graphie unie seulement, sont typologiquement identiques aux formations qui admettent la graphie double : par ex. les occurrences de *eiusmodi* sont rangées parmi celles de *eius* suivi de *modi*. Par contre *quare*, à défaut de formations analogues avec graphie divisée, est considéré comme un *keyword*. Si la différence graphique correspond à une différence sémantique, nous ne sommes pas intervenus.

## 7. Les variantes morphologiques

Nous n'avons pas estimé légitime d'intervenir sur la disposition du texte quand les variantes ne reflètent pas seulement des réalités phonétiques mais elles concernent le niveau morphologique. La limite entre les deux catégories est souvent insaisissable et peut-être arbitraire si l'on considère l'interdépendance des différents niveaux de la langue; toutefois, puisqu'une limite même conventionnelle doit être tracée, on a inclus dans cette seconde classe toutes les variantes qui intéressent le morphème désinentiel et le thème flexionnel<sup>28</sup>.

Sous cet aspect également l'usage de Symmaque apparaît homogène dans l'ensemble; toutefois on y rencontre les anomalies les plus communes et les plus tolérées par la *consuetudo* dans la flexion nominale, pronominale et verbale.

7.1. Dans la flexion nominale on signale :

- 7.1.1. les nominatifs sigmatiques *colos* (p. 11,3) et *honos* (p. 269,10) à côté des nominatifs analogiques en *-or* beaucoup plus nombreux;
- 7.1.2. le génitif archaïque en *-as* dans le rapport *mater familias* (pp. 275,35 et 338,34) et *matrem familias* (p. 78,20);
- 7.1.3. le génitif en *-i* dans de nombreux noms propres avec thème en *-io-*;
- 7.1.4. l'ablatif en *-i* dans les participes présents soit avec fonction de substantif (ex *abundanti* pp. 98,2 et 184,8) soit avec fonction d'adjectif : *insequenti* (die) (p. 297,33), *neglegenti* ... *celeritate* (p. 7,28), *de praesenti fortuna* (p. 123,15), *pro ratione praesenti* (p. 304,1);
- 7.1.5. le génitif pluriel en *-um* soit de thèmes en *-ā-* (*caelicolum* p. 3,22) soit de thèmes en *-o-* (*modium* p. 312,16) soit de thèmes en *-ti-* (*Laurentum* p. 31,24 et *optimatum* pp. 247,9 et 302,30) soit de participes (*cogentum* p. 67,9);

<sup>28</sup> Ces variantes sont examinées et discutées aux pp. XVIII–XXVIII de la préface à la concordance de Symmaque.

- 7.1.6. les morphèmes grecs dans des noms d'origine grecque;
- 7.1.7. les variations de genre entre singulier et pluriel dans le cas de *bal(i)neum* : *bae* et de *locus* : *loca*;
- 7.1.8. les traces de l'ancienne apophonie entre les suffixes :
  - 7.1.8.1. *-ondo-* et *-endo-* dans les formes du gérondif et du participe futur passif;
  - 7.1.8.2. *-es-* et *-os-* dans *pignus*, dont on trouve le génitif pluriel *pignerum* (pp. 48,4 et 49,3) à côté des formes en *-or-*;
- 7.1.9. les cas d'étréoclisie dans *luxuria/luxuries* et dans *plebes/plebs*;
- 7.1.10. les anomalies de degrés apophoniques dans la flexion de *deus*.
- 7.2. Dans la flexion pronominale on signale :
  - 7.2.1. les formes avec *-cum* enclitique;
  - 7.2.2. les formes avec le suffixe *-ce* intensif (*huiusce* p. 5,4) et les formes redoublées (*memet* p. 337,18 et *sese* pp. 84,12; 142,9; 156,14; 337,9);
  - 7.2.3. les oscillations dans la flexion de *is* (le datif-ablatif pluriel *eis* apparaît seulement à la p. 242,26), de *ille* (le datif archaïque *olli* à la p. 3,21 est isolé), de *qui* (l'ablatif pluriel *quis* ne paraît qu'aux pp. 14,19 et 133,3).
- 7.3. Dans la flexion verbale on signale :
  - 7.3.1. la deuxième personne du singulier en *-re* de verbes déponents;
  - 7.3.2. la troisième personne pluriel en *-ēre* du parfait;
  - 7.3.3. les imparfaits en *-ibam* de thèmes en *-ī-* *aperibam* (p. 9,9) et *desilibat* (p. 325,15);
  - 7.3.4. les formes contractées dans le système du perfectum des verbes à voyelle longue;
  - 7.3.5. les archaïsmes : l'optatif *ausim* (p. 44,34), les conjonctifs radicaux *fuat* et *foret*, le parfait redoublé *tetulistī* (p. 3,37).

Quelques oscillations s'observent aussi dans les invariables : *alioqui* (p. 16,8), *donecum* (p. 8,12) et *rursum* (p. 17,27) sont isolés par rapport aux fréquents *alioquin*, *donec* et *rursus*; *fortasse* et *sat* sont moins fréquents que *fortassis* et *satis*.

## 8. Le programme de lemmatisation automatique

Les interventions laborieusement effectuées à la main sur la concordance de Symmaque, sur la base des index produits par l'ordinateur, ont été possibles à cause de l'extension limitée du texte (environ 100 000 mots) et de sa relative homogénéité sur le plan linguistique. Il aurait été impensable d'appliquer le même procédé aux grammairiens, non seulement à cause du volume du matériel (plus d'un million de mots), mais aussi à cause de son caractère : en effet l'objet privilégié des *artes*, à côté de la description du système du latin, consiste précisément à discuter les formes alternatives sur le plan graphico-phonétique et morphologique. Il s'agirait surtout d'un travail inutile, si l'on considère d'un côté la mobilité du *corpus*, du moment que de nouvelles éditions continuent à être publiées, et de l'autre la possibilité de consultation vidéo du matériel de dépouillement. Pour ces raisons il s'est avéré nécessaire de chercher de nouveaux procédés qui permettent à l'utilisateur d'obtenir toutes les occurrences d'un mot quelle qu'en soit la graphie ou quel que soit le schéma de la flexion.

Le problème a été affronté au moyen du programme de lemmatisation automatique du latin<sup>29</sup>, qui en ce qui concerne la transcription et l'analyse des données est parvenu désormais à la phase de la révision finale.

Du fait que l'ordinateur considère comme autonome n'importe quelle unité graphiquement différente, les cas d'assimilation et de dissimilation sont enregistrés séparément, de même que les mots admettant une graphie unie ou divisée, les variantes graphiques et ainsi de suite, comme les pages précédentes l'ont montré. La première difficulté se présente donc dans l'opération qui semblerait la moins complexe, l'ordonnance des formes.

Pour trouver une solution il a été nécessaire de réunir les variantes graphiques dans un même ensemble de formes. En adoptant ce critère on obtient ce qu'on peut définir comme une lemmatisation graphique, c'est-à-dire que l'on crée un lemme multiple, mais unitaire, une « constellation lemmatique », au sein de laquelle aucune forme n'est privilégiée. Par ex. *adfero* et *affero*, *existo* et *existo*, *harena* et *arena* deviennent respectivement les deux formes auxquelles

---

<sup>29</sup> Cf. MARINONE (Nino) : 1983, « A Project for a Latin Lexical Data Base », *Linguistica Computazionale*, III, pp. 175-178; 1990 : « A Project for Latin Lexicography : 1. Automatic Lemmatization and Word — List », *Computers and the Humanities*, XXIV, pp. 417-420; BOZZI (Andrea) : 1982, « Progetto di organizzazione di un vasto repertorio lessicale automatico della lingua latina », *Maia*, XXXII, pp. 167-172; BOZZI (Andrea) et CAPPELLI (Giuseppe) : 1990, « A Project for Latin Lexicography : 2. A Latin Analyser », *Computers and the Humanities*, XXIV, pp. 421-426.

l'ordinateur reconduit de façon unitaire. Le chercheur peut partir de l'une ou de l'autre et il obtiendra toujours le même résultat d'analyse.

Mais par le même mécanisme qui rend distinctes les variantes graphiques, les homographies dans les domaine morphologique ne sont pas séparées : par ex. *uis* substantif est confondu avec *uis* forme verbale, l'ablatif *tristi* ne fait qu'un avec la forme syncopée du parfait *trivi*. En conséquence nous avons pris en considération la possibilité d'introduire dans l'analyse une lemmatisation morphologique qui agisse conjointement avec la lemmatisation graphique. Autrement dit, le système rapporte chaque forme à un lemme (ou à une constellation lemmatique, dans le sens donné plus haut) qui correspond à un article figurant dans un dictionnaire latin.

Cette solution ayant été choisie, un autre problème s'est présenté : quel dictionnaire adopter comme référence. Normalement on juge un dictionnaire soit d'après le nombre des mots-vedettes par rapport à la tranche prise en considération dans l'évolution d'une langue, soit et surtout d'après la méthode adoptée pour l'élaboration des lemmes, d'après la clarté avec laquelle sont distinguées les différentes acceptions et l'abondance d'exemples appropriés, munis de références exactes. Dans notre cas, du fait que l'ordinateur n'a pas accès aux valeurs sémantiques, l'évaluation quantitative du patrimoine lexical est de loin ce qui prévaut. Par conséquent il est opportun de prendre comme base la liste qui contient le plus grand nombre de mots, plutôt que le dictionnaire le meilleur dans le sens traditionnel, du fait qu'en réalité pour l'analyse automatique il suffit d'une simple liste des lemmes. Cette assertion peut sembler paradoxale, mais elle correspond exactement à la situation.

Comme le *Thesaurus Linguae Latinae* est encore inachevé, en général on estime que le *Lexicon totius Latinitatis* d'Egidio Forcellini<sup>30</sup> soit le plus complet parmi les dictionnaires latins et en effet il a été adopté dans diverses élaborations par ordinateur. Toutefois cette conviction est basée sur l'ensemble des 92 052 lemmes contenus dans le Forcellini, parmi lesquels sont incluses aussi les 29 624 entrées de l'index onomastique, qui pour de multiples raisons ne nous rend aucun service; en outre de fréquents renvois augmentent le nombre des lemmes, mais certainement pas la quantité et la qualité de l'information, qui est de plus infirmée par le peu de crédibilité des éditions utilisées. Après un contrôle précis il est manifeste que l'index des lemmes de Gradenwitz<sup>31</sup>, qui compte 52 687 entrées à l'exclusion totale des noms propres, offre une plus

<sup>30</sup> Dans la quatrième édition aux soins de CORRADINI (Francesco) et PERIN (Giuseppe), Padova 1864-1887 [Bologna 1965].

<sup>31</sup> GRADENWITZ (Otto), *Laterculi uocum Latinarum*, Leipzig 1904 [Hildesheim 1966].

grande richesse quantitative; par conséquent dans notre cas, il est sans aucun doute à préférer. Toutefois cette décision ne résout pas le problème, elle ne fait que le poser.

En effet l'ordinateur a cette autre caractéristique (ou qualité ou défaut selon les points de vue) déjà signalée : pour produire des résultats valables il exige une organisation des données rigoureusement formalisée. Mais aucun dictionnaire latin n'est structuré de la sorte. Une étude comparative a mis en évidence les divergences dans l'organisation des lemmes entre le Forcellini, le Georges<sup>32</sup> (avec les suppléments du Gradenwitz) et le *Thesaurus*<sup>33</sup>. Le même mot est souvent traité de façon différente : par ex. un participe à valeur nominale est présenté chez l'un comme un lemme autonome, chez l'autre comme un sous-lemme du verbe, chez l'autre encore à l'intérieur du verbe lui-même; tandis qu'un autre participe qui admet une fonction semblable est traité de façon différente. Cette anomalie est justifiée par des motifs que la compétence de l'auteur de chaque lemme a considérés comme valables; mais du point de vue de l'ordinateur ce n'est pas admissible.

La seule issue possible nous a semblé être la rédaction d'un nouveau répertoire lexical qui tienne compte de cette exigence particulière. Pour sa réalisation nous avons rempli environ 57 000 fiches lexicales en adoptant des critères de catalogage particuliers et en utilisant, à côté du Gradenwitz, le Georges et le récent dictionnaire d'Oxford<sup>34</sup>. L'enregistrement des données, effectué sur une grille de 80 colonnes, est fondé sur la segmentation des formes répartie en six secteurs de la façon suivante :

8.1. *Références* : lettre initiale du domaine morpho-lexical auquel appartient le lemme, indiqué par les majuscules N(ominal), P(ronominal), V(erbal), I(nvariable), suivie de cinq chiffres pour la numérotation progressive à l'intérieur de chaque domaine.

8.2. *Indicateurs* :

8.2.1. pour le lemme : la lettre V signale que le lemme doit être inséré dans une constellation lemmatique, et vice versa Y interdit cette insertion;

8.2.2. pour les segments initiaux : des codes alphanumériques indiquent la présence d'altérations dans les préfixes variables (par ex. les pré-verbales) dans le but d'obtenir une lemmatisation graphique; de façon

<sup>32</sup> Dans la dernière édition : GEORGES (Karl Ernst et Heinrich), *Ausführliches Lateinisch-Deutsches Handwörterbuch* (Hannover — Leipzig 1913–1918 [Basel 1951]).

<sup>33</sup> Cf. LOMANTO (Valeria) : 1980, «Lessici latini e lessicografia automatica», *Memorie dell'Accademia delle Scienze di Torino*, Classe II, 5.4, pp. 111–270.

<sup>34</sup> Cf. GLARE (P.G.W.) : 1982, *Oxford Latin Dictionary* (Oxford).



analogue un chiffre signale les préfixes pronominaux séparables *ne num si*;

- 8.2.3. pour les segments médians : le signe + produit automatiquement le thème du parfait, du supin et des participes passé et futur des verbes réguliers, et le signe – interdit la formation de toutes les formes nominales de n'importe quel verbe;
- 8.2.4. pour les segments qui suivent la finale : des chiffres signalent la présence des suffixes pronominaux et des enclitiques *-ne, -que, -ue*.
- 8.3. *Segment base* : il représente la partie invariable du lemme dans la flexion (qui ne s'identifie pas nécessairement avec la racine ou avec le thème ou avec le lexème).
- 8.4. *Segment final* : des codes alphanumériques indiquent le type ou les types de la flexion et les éléments finals (qui ne s'identifient pas nécessairement avec la désinence ou avec la terminaison) et pour chaque type de flexion sont mentionnés les groupes présentés ci-dessous.
- 8.4.1. Dans le domaine N(ominal) on signale :
- noms en *-ā-* (1<sup>re</sup> déclinaison)
    - flexion normale – anormale;
  - noms en *-o-* (2<sup>e</sup> déclinaison)
    - masculins/féminins – masculins en *-ius*,
    - neutres – neutres en *-ium*,
    - gérondifs,
    - flexion anormale;
  - noms en *-o-/-ā-* (2<sup>e</sup>/1<sup>re</sup> déclinaison)
    - adjectifs – adjectifs en *-ius*,
    - adjectifs pronominaux – numéraux – superlatifs,
    - participes passés – futurs – futurs passifs;
  - noms en *-i-* et en consonne (3<sup>e</sup> déclinaison)
    - avec gén.pl. en *-um/-ium* masculins/féminins, neutres,
    - avec gén.pl. en *-um* masculins/féminins, neutres,
    - avec gén.pl. en *-ium* masculins/féminins, neutres,
    - adjectifs – comparatifs,
    - participes présents,
    - flexion anormale;
  - noms en *-u-* (4<sup>e</sup> déclinaison)
    - flexion normale,
    - supins;
  - noms en *-ē-* (5<sup>e</sup> déclinaison);
  - noms indéclinables.

8.4.2. Dans le domaine V(erbal) on signale :

- infectum en  $\bar{a}$ - (1<sup>re</sup> conjugaison)
- infectum en  $\bar{e}$ - (2<sup>e</sup> conjugaison)
- infectum en  $\bar{e}$ - (3<sup>e</sup> conjugaison)
- infectum en  $\bar{i}$  (4<sup>e</sup> conjugaison)
- infectum en  $\bar{y}$ - (conjugaison en  $\bar{e}$ - /  $\bar{i}$ -)
- infectum anomal;
- perfectum – perfectum syncopé,
- flexion impersonnelle,
- flexion non déterminable.

Pour chaque type il est possible de faire suivre le code d'une lettre alphabétique, dont la fonction est de limiter l'analyse à certaines tables de segments finals dans le but d'éviter de faux résultats; en outre dans l'infectum on a réuni séparément les groupes du transitif, de l'intransitif, du seul déponent, du seul actif, de l'impersonnel.

8.4.3. Dans le domaine P(ronominal) on a signalé les segments qui s'unissent aux pronoms, dont toutes les formes sont enregistrées.

8.4.4. Dans le domaine I(nvariable) sont regroupés les adverbes (avec éventuellement le code nominal auquel ils se rapportent), les prépositions, les conjonctions, les interjections sans distinction spécifique.

8.5. *Lemme* : le lemme ou la constellation lemmatique auxquels la forme analysée se rapporte sont produits automatiquement (sauf dans le cas des formations irrégulières) avec éventuellement la présence simultanée d'alternatives dans le type de flexion et en particulier avec l'adjonction pour les comparatifs et les superlatifs de la forme de l'adjectif ou de l'adverbe dont ils dérivent; pour le perfectum et les noms verbaux, de la forme du présent auquel ils sont assignés.

8.6. *Classification* : la catégorie morphologique à laquelle appartient chaque lemme produit par l'analyse est indiquée automatiquement au moyen des sigles suivants :

N	V	PR(onom)
N1	V1	I(nvariable)
N2	V2	ENC(litique)
N3	V3	
N4	V4	
N5	V5	
N2/1	VA(nomal)	
N2G(érondif)	VP(erfectum)	
N4S(upin)		

Les premiers essais de ce système d'analyse ont déjà donné des résultats positifs<sup>35</sup>. Mais le moment est arrivé de se demander si ces résultats méritent une telle dépense d'énergie et d'argent. Tant de travail pour obtenir finalement de l'ordinateur la réponse que *patrem* est une forme de *pater* de la troisième déclinaison, que *inermis* peut se rapporter soit à *inermis* soit à *inermus*, que *arator* est un substantif, mais qu'il est aussi l'impératif futur de *arare*, que *suaue* peut être le neutre de l'adjectif ou équivaloir à *uel sua*, que *quinque* représente un numéral ou peut être analysé comme *et quin* (qui est un *hapax* de Plaute) et ainsi de suite, à la manière d'un jeu électronique divertissant.

Ce n'est certes pas là l'objectif final, qui réside plutôt dans la recherche des mots. Actuellement l'interrogation est limitée à une ou à plusieurs formes univoques; ou bien, si l'on demande le segment invariable d'une forme, la réponse embrasse tous les mots qui contiennent la séquence demandée. On peut le faire comprendre par un exemple. On centre la recherche sur le secteur choisi (c'est-à-dire un groupe d'auteurs ou même le patrimoine littéraire tout entier), on tape sur le clavier *rosarum* : on lira sur l'écran toutes les occurrences de ce génitif, chacune étant insérée dans un contexte et accompagnée de références. Autre type d'approche : on tape *ros*. On obtient toutes les formes commençant par ce segment, mais non seulement celles qui se rapportent au substantif *rosa* : on verra apparaître toutes les formes de *ros rosaceus rosarius rosatus roscidus/rosidus rosetum roseus rosio rosmarinum rosor rostellum rostralis rostrans rostratus rostrum rosulentus*, du parfait *rosi* et enfin de *rosus* aussi bien participe que substantif. Un résultat irréprochable, mais en pratique inutilisable si le domaine de la recherche a une certaine extension. Il est superflu d'ajouter à quel point les choses se compliquent dans le cas d'un verbe à paradigme supplétif. Nous croyons qu'il est possible de remédier à ces inconvénients en appliquant à la recherche des mots le système de lemmatisation que nous avons élaboré.

Actuellement nous examinons plusieurs hypothèses de travail. Nous pensons à un système d'interrogation où la recherche serait limitée par différents codes alphanumériques; par ex. en tapant *rosa NI* on n'obtient que les formes du substantif; en tapant *fero V* on obtient toutes les occurrences du verbe dans les textes étudiés, mais en ajoutant *VP* on n'obtiendra que celles du parfait, en

<sup>35</sup> Tous les travaux décrits jusqu'ici ont été exécutés avec la collaboration du personnel technique de l'Istituto di Linguistica Computazionale de Pise sous la direction du Dr. A. Bozzi, responsable de la sphère de recherche «Filologia computazionale». En particulier les concordances de Symmaque et des grammairiens latins et l'index des grammairiens ont été confiés aux programmeurs MM. R. Bindi et D. Brogna; le lemmatiseur fonctionne moyennant l'application d'un procédé complexe réalisé par le programmeur M. G. Cappelli.

ajoutant *VN* seulement les nominales, et ainsi de suite. En tenant compte des critères selon lesquels les entrées ont été enregistrées dans le but de parvenir à la plus grande formalisation possible, pour faciliter leur repérage il pourrait être utile de rédiger un index des lemmes où chaque mot-vedette serait accompagné du code alphanumérique conçu pour la lemmatisation : en ce cas il suffit que l'utilisateur tape la référence (éventuellement suivie de codes délimitateurs, comme il est indiqué plus haut) pour obtenir avec une extrême précision le résultat désiré : par ex. R0795 pour *rosa*, F0369 pour *fero* (avec ou sans *P* ou *N*), etc.

Pour le moment il ne s'agit que d'un projet<sup>36</sup>. Le chemin est encore long et les difficultés ne manquent pas, mais il vaut la peine de se mettre à l'épreuve pour atteindre dans ce secteur une véritable rigueur philologique, qui malheureusement n'est pas toujours l'aspiration des travaux linguistiques sur ordinateur.

---

<sup>36</sup> Son exécution est confiée à la collaboration du prof. M. Lana au Centro Interdipartimentale di Servizi Informatici de l'Université de Turin.