

**La constitution d'un thésaurus
pour les pères de l'église latine**

par

Paul TOMBEUR

CETEDOC, Louvain-la-Neuve - BELGIQUE

869

Le projet de traitement systématique par ordinateur des œuvres des Pères latins se situe dans le cadre d'un vaste ensemble de projets en cours au Centre de Traitement Electronique des Documents (CETEDOC) de l'Université Catholique de Louvain à Louvain-la-Neuve. L'ensemble des projets concernant le latin a reçu l'appellation générale de *Vox latina*.⁽¹⁾ Rappelons rapidement, en dehors du domaine patristique, l'objet des plus importants d'entre eux : tous les textes de notre moyen âge latin belge jusqu'en 1200, les Conciles œcuméniques, les Sources franciscaines et les œuvres de saint Bonaventure, le Livre des Sentences de Pierre Lombard, les florilèges d'Aristote.⁽²⁾ Quantité de résultats ont été publiés ou sont dès à présent à la disposition des chercheurs. Ainsi pour les millions de mots que comptent toutes les œuvres recensées dans notre répertoire national belge, nous avons constitué une concordance complète, sur microfiches qui correspond à pas moins de 100 volumes de 1000 pages ; grâce à cette concordance on peut repérer immédiatement tout emploi, toute expression utilisés dans ces textes.⁽³⁾

Si précédemment nous avons déjà étudié quelques œuvres patristiques, l'entreprise que j'ai l'honneur de vous présenter concerne tous les textes qui ont été ou qui sont en cours de publication dans la collection *Corpus Christianorum, Series latina*, à laquelle il faut joindre la *Continuatio Mediaevalis*, car il est bien clair aujourd'hui que les limites chronologiques qu'on nous a imposées n'ont pas de sens scientifique et qu'il faut étudier comme un tout la patristique et le dit moyen âge. Un tel travail englobant cette masse d'œuvres est de nature à aboutir à la constitution d'un véritable *Thesaurus Patrum Latinorum*. Comme celui-ci sera intégralement traité par ordinateur, il sera susceptible de prendre les formes les plus diverses et d'être soumis aux questions les plus variées. Ce projet sera notamment mené grâce à l'aide financière du Fonds National de la Recherche Scientifique. Il est le résultat d'un accord conclu entre le *Corpus Christianorum*, que dirige à l'abbaye de Saint-Pierre de Steenbrugge le Révérend Père Dekkers, la maison d'édition Brepols, à Turnhout, et le CETEDOC.

Deux principes fondamentaux sont appliqués en vue de la réalisation de ce *Thesaurus* :

- 1) Chaque œuvre est traitée de façon intégrale, du premier mot au dernier. Les bandes magnétiques constituées contiennent l'ensemble du texte.
- 2) Chaque mot de chaque œuvre est étudié en fonction des deux réalités qu'il représente :
 - la *réalité textuelle*, c'est-à-dire la réalité formelle : la forme en tant que forme, isolée ou associée à d'autres (syntagmes, expressions, citations) ;
 - la *réalité lexicale* : le lemme auquel correspond chaque forme. Chaque œuvre est dès lors soumise à une analyse lexicologique fondamentale.

Voyons brièvement les implications de ces principes et la manière dont nous les mettons en œuvre.

1. La constitution des bandes magnétiques

Il faut que nous disposions de tout le texte, sans qu'aucune partie n'en soit privilégiée. L'un des buts principaux à atteindre est en effet de pouvoir fournir des concordances pour tous les mots, les

plus rares comme les plus fréquents, les plus signifiants comme ceux qui apparemment sont les plus banals. Il ne faut opérer au départ aucune sélection dans les données que nous enregistrons.

Pour toutes les œuvres parues dans le *Corpus Christianorum* jusqu'à ces dernières années, nous avons entrepris l'enregistrement manuel du texte. La référence de chaque œuvre, son code d'œuvre, est constituée par le numéro qui lui est attribué dans la *Clavis Patrum Latinorum*.⁽⁴⁾ Les références de chaque mot, calculées automatiquement grâce à notre programme de référencement tiennent compte des divisions de l'œuvre et de la numérotation des lignes dans l'édition du *Corpus*.

Plusieurs millions de mots ont été ainsi enregistrés et nous disposons dès à présent "on machine readable form" de quantité de textes : la plus grande partie des œuvres d'Augustin, les *opera omnia* de Tertullien et de Lucifer de Cagliari, des écrits d'Ambroise, Jérôme, Quodvultdeus, Léon le Grand, Prudence, Fulgence, Cassiodore, Césaire d'Arles, Grégoire le Grand, Bède, etc.

Pour les œuvres récemment éditées, comme pour toutes celles qui paraîtront dans les années à venir, nous recevons directement de l'éditeur des bandes magnétiques contenant le texte. Dorénavant, en effet, nos œuvres patristiques sont imprimées grâce aux procédés de photocomposition et les bandes réalisées pour la fabrication du livre contenant l'édition du texte, constituent pour nous les données d'entrée. L'avantage est immense : une seule transcription du texte, à partir du manuscrit de l'auteur, sert de base tant pour la publication de l'édition que pour la constitution des divers index et concordances. Notre banque de données patristiques va dès lors s'accroître automatiquement au fur et à mesure des éditions publiées. Il est d'ailleurs hautement souhaitable que désormais on ne publie plus de nouvelles éditions sans recourir à ce procédé permettant d'un seul coup la production du livre et celle de ses instruments de lecture.

Ces bandes magnétiques produites par l'éditeur sont soumises à divers programmes d'ordinateur afin d'établir des fichiers référencés selon les normes standardisées que nous avons mises au point (suppression des codes de photocomposition, réécriture automatique des mots coupés en fin de ligne, référencement automatique).

2. L'étude de la réalité textuelle et la publication des Instrumenta Lexicologica Latina - Series A - Formae

Une fois les textes enregistrés et référencés, nous disposons de toutes les occurrences et nous pouvons dès lors, dans un premier temps, créer des instruments de travail basés sur les formes. Mis à part les problèmes de correction — il y a toujours des erreurs à corriger tant pour les enregistrements manuels que pour les bandes réalisées pour la photocomposition —, et l'indication des codes de citation — nous codons en effet les formes faisant partie de citations sur la base du travail réalisé par l'éditeur —, la réalisation de ces instruments de travail est évidemment purement automatique.

La publication des résultats de nos traitements se réalise au sein d'une nouvelle collection qui vient d'être créée sous le sigle du *Corpus Christianorum* et qui s'intitule : *Instrumenta Lexicologica Latina (ILL)*. Tout instrument basé sur les formes paraîtra dans une *series A*. A chaque œuvre ou à chaque ensemble d'œuvres retenu correspondront en effet deux fascicules des *ILL* : une *Series A Formae* et une *Series B - Lemmata* dont nous parlerons ultérieurement. Comme on le verra, ces deux séries sont rigoureusement complémentaires.

La *series A* comprend les parties suivantes :

1) *Enumeratio formarum*

C'est la liste alphabétique de toutes les formes différentes du texte. Chaque forme est accompagnée de sa fréquence. Cette liste permet donc de connaître toutes les formes utilisées et de savoir si celles que l'on recherche se trouvent dans l'œuvre.

L'indication de la fréquence ne constitue généralement, à ce niveau formel, qu'une simple indication pratique. Quantité de formes sont en effet ambiguës : seul le contexte immédiat ou large (le texte même de l'œuvre) permet de savoir de quel vocable il s'agit.

Cette *enumeratio formarum* est publiée sous forme imprimée. Elle sert de manuel de référence pour l'instrument de travail suivant.

2) *Concordantia formarum*

La concordance des formes est rigoureusement exhaustive. Toutes les formes sont classées dans l'ordre alphabétique. Les formes non latines sont regroupées sous un code qui les rassemble (ex. G- pour toutes les formes grecques).

Dans chaque contexte, les formes sont centrées au milieu d'une ligne de texte qui comprend un maximum de 142 caractères. Des références précises et d'éventuels codes de citation accompagnent chacune des formes.

Cette concordance doit permettre de saisir les réalités textuelles. Les attestations sont dès lors classées selon l'ordre alphabétique de tous les mots qui suivent, c'est-à-dire selon l'ordre des contextes. Ce classement permet de retrouver immédiatement tout syntagme, toute expression, toute reprise à l'intérieur même de l'œuvre ou toute citation, à condition, bien entendu, qu'il n'y ait pas de variation orthographique pour une même forme. L'examen de la liste des formes ainsi que celui du "reverse-index" dont nous allons parler, permettent une vérification rapide quant aux orthographes possibles. Dans le cas où les contextes qui suivent une forme sont rigoureusement identiques, l'ordre des attestations est celui du texte.

Cette concordance est publiée sous forme de microfiches. La microfiche constitue en effet, à notre avis, la forme de publication qui s'impose pour ces instruments de travail. Elle seule permet de diffuser, à un prix abordable, les instruments de travail que nécessite la recherche ;

elle nous met en mesure de fournir de larges contextes et ce pour tous les textes, sans privilégier quelques œuvres jugées maîtresses. (La publication, sous forme imprimée, de la seule liste des formes, nous a paru la solution la plus heureuse afin que le chercheur n'ait cependant recours aux microfiches, qu'après avoir constaté la présence des mots qu'il recherche ou qui l'intéressent.)

3) *Index formarum a tergo ordinarum*

A la suite de la concordance, nous publions, également sur microfiche, un index inverse de toutes les formes. Quantité d'informations peuvent être puisées dans un "reverse-index" des formes. Celui-ci rassemble toutes les désinences identiques. Il nous permet également de vérifier certaines variantes orthographiques : il classe en effet en un même lieu toutes les formes dispersées alphabétiquement dans la concordance ou la liste des formes en vertu des variations portant sur les lettres initiales (ex. *co-* et *quo-* ; on songe naturellement à *cotidie*, *cottidie*, *quotidie*, etc., mais on risque de ne guère penser à *coaeternus* et *quoeternus* ; ou *humerus* et *umerus*).

Cette *series A* sera fournie aux chercheurs en même temps que sortira de presse une nouvelle édition du *Corpus Christianorum, Series Latina* et *Continuatio Mediaevalis*. Sans plus attendre, les chercheurs disposeront donc d'outils précieux pour l'étude de la réalité textuelle. Quant aux textes déjà parus, nous espérons réaliser la *series A* dans l'espace de quelques années.

Ajoutons que, grâce à un accord conclu avec l'*Augustinus Lexikon* que dirige à Würzburg le Professeur Mayer, nous disposerons de ces instruments formels pour *tout* Augustin (y compris donc les textes qui n'ont pas encore paru dans le *Corpus Christianorum*) dès 1983. Ce sera la base du *Thesaurus Augustinianus* que nous publierons aux éditions Brepols, parallèlement à l'*Augustinus Lexikon* qui en sera le pendant sémantique. Dès la constitution de ces concordances formes pour saint Augustin, on pourra donc retrouver de manière immédiate toute citation de cet auteur.

3. L'étude de la réalité lexicale et la publication des *Instrumenta Lexicologica Latina : Series B - Lemmata*

L'observation des réalités textuelles est une chose, celle des réalités lexicales en est une autre. On ne peut s'en tenir en effet à un point de vue purement formel. Les réalités formelles sont ambiguës. Pour en dégager le lexique d'un auteur ou d'une œuvre, une analyse lexicologique s'impose, à savoir une lemmatisation. Pour une étude du vocabulaire d'une œuvre, on ne peut se baser sur les seules formes. Il faut regrouper selon une même entrée ou lemme les diverses formes orthographiques et grammaticales correspondant à un vocable. Dans bien des cas, le non-spécialiste (oserais-je dire le non-lemmatiseur de métier?) ne peut imaginer les multiples formes grammaticales et orthographiques qui correspondent à un lemme unique et en même temps se rendre compte de l'ambiguïté lexicale que représentent les formes, savoir par exemple que plusieurs lemmes peuvent correspondre à des formes telles que *condicio*, *contentus*, à un grand nombre de mots commençant par *in-* (où l'on constate l'existence d'un vocable et celui de son contraire!), etc. Ce n'est pas ici le

lieu de m'étendre sur ces exemples. Pour saisir le vocabulaire latin, il faut lemmatiser les formes. Au premier stade seul, une statistique lexicale prend du sens.

Oui ... mais quel type de lemmatisation va-t-on appliquer ? Il importe en effet de se livrer tout premier lieu à une critique des lemmatisations pratiquées. On ne peut travailler dans ce domaine pendant vingt ans sans découvrir de nouvelles perspectives. La question fondamentale dans la mise au point des méthodes me semble celle-ci : qu'est-ce qui, dans nos pratiques, correspond à des habitudes et qu'est-ce qui correspond à des réalités scientifiques ?

L'utilisateur des divers documents lemmatisés que tous ceux qui ont pratiqué la lemmatisation lui ont présentés, ne peut que difficilement se rendre compte des implications découlant des divers types de lemmatisation. Bien souvent, il sera pris au piège sans le savoir.

Pour plusieurs textes on dispose de lemmes et de fréquences, voir même de lemmes et de résultats de recherches statistiques portant sur ces lemmes. Ces lemmes sont étrangement semblables, mais que regroupent-ils ? *Sum* = ? A-t-on distingué ou non l'auxiliaire ? *Qui* = ? A-t-on distingué le relatif, l'indéfini, l'interrogatif ? Et qu'en est-il pour tous les cas adverbiaux, comme *quo* ou *quod* ? *Ante* = ? A-t-on distingué les emplois adverbiaux et les emplois prépositionnels ? Même *et* = ? puisque les uns distinguent l'emploi adverbial de la conjonction de coordination, et non les autres. *Amicus* = ? A-t-on distingué les adjectifs des substantifs ? *Bonus* = ? A-t-on de même distingué les divers dérivés grammaticaux et a-t-on consacré des entrées différentes aux formes comparatives et superlatives et à leurs dérivés substantivés ? Je pourrais continuer longuement cette liste. Nous vivons à vrai dire en pleine anarchie lexicale, le plus souvent sans nous en rendre compte, et le risque d'erreur d'interprétation est immense. (Je crois d'ailleurs qu'il en est ainsi pour la plupart des langues.)

Je me permettrai de citer un exemple parmi des textes que nous avons nous-mêmes lemmatisés : le *de Trinitate* d'Hilaire de Poitiers et *l'Ethica* de Spinoza. Si je compare les listes de fréquences décroissantes établies pour ces deux œuvres (lesquelles, soit dit en passant, n'ont de sens au point de vue lexical qu'établies sur la base des lemmes), il y a plusieurs ressemblances frappantes dans l'ordre des lemmes. Ces ressemblances sont totalement trompeuses, car les méthodes de lemmatisation sont radicalement différentes. (Il en est question dans les introductions des concordances publiées, mais qui lit les introductions ?...) Si j'adapte le système de lemmatisation appliqué à Hilaire au fichier de Spinoza, sur les 60 lemmes les plus fréquents, 37 lemmes changeront de fréquence presque à coup sûr et tout l'ordre se trouvera perturbé. Ainsi à la fréquence de *sum* de Spinoza faut-il ajouter les cas du verbe *esse* utilisé comme auxiliaire, examiner s'il n'y a pas eu de *esse* substantivé qui aurait été distingué ou des cas de *ens* ; de même pour *qui* : quantité de distinctions au niveau des lemmes ont été opérées pour Spinoza alors qu'elles ne l'ont pas été pour Hilaire. Peut-on attendre de celui qui réceptionne nos résultats autant d'attention ? Sûrement non.

Il importe dès lors que l'on mette en place une nouvelle méthode de lemmatisation qui nous préserverait au maximum des erreurs d'interprétation, des méprises dues à la simple lecture des

instruments lexicaux. La nécessité d'une méthodologie de lemmatisation unique s'impose quel que soit, comme nous le verrons, le degré d'analyse que chacun désire développer. Cette lemmatisation opérerait des regroupements identiques et fournirait dès lors en mineure, s'il y a lieu, les diverses distinctions dans lesquelles un chacun désire entrer.

La lemmatisation de base ne peut être qu'une lemmatisation globalisante. Elle correspond aux méthodes choisies pour établir les entrées lexicales par le *Thesaurus Linguae Latinae*, et plus encore à celles du *Latin Dictionary* de Lewis-Short, utilisé comme instrument de base par presque tous les anglophones. Dans une moindre mesure elle correspond au *Lexicon Totius Latinitatis* de Forcellini. Il est piquant de constater que le plus souvent les non-spécialistes n'ont pas conscience des entrées lexicales attestées dans ces instruments de travail ; c'est dire notamment la force des habitudes, suivies le plus souvent sans références sérieuses.

Je ne puis entrer ici dans le détail et je consacrerai ultérieurement un rapport spécifique à la question de la lemmatisation. Il fallait ici souligner ce problème : parler de lemmatisation ne suffit pas et il faudrait s'interdire d'employer désormais ce substantif sans ajouter immédiatement, du moins dans nos commentaires, un qualificatif qui explique de quelle lemmatisation il s'agit.

Il faut en effet distinguer plusieurs niveaux de lemmatisation. On peut concevoir divers types d'entrées lexicales selon les critères de base retenus : une lemmatisation de type formel, une lemmatisation morphologico-syntaxique et une lemmatisation sémantique.

La lemmatisation de type formel rassemble sous une même unité toutes les variantes orthographiques d'un même mot ainsi que toutes les formes grammaticales dérivées d'un même vocable. La lemmatisation morphologico-syntaxique opère une distinction entre les formes grammaticales dérivées d'un même vocable. Ainsi, par exemple, distingue-t-elle un emploi adjectivé ou substantivé d'un même participe, lequel se trouvera, dans son emploi de participe, classé sous le verbe ; on y séparera de même les emplois adverbiaux ou prépositionnels d'un même mot. La lemmatisation sémantique opère un pas de plus : non seulement elle entre dans toutes les considérations morphologico-syntaxiques, mais elle distingue les divers sens d'un même mot.

On voit dès lors que les critères retenus pour le choix du lemme dans chacune de ces trois lemmatisations sont divergents. Dans une lemmatisation de type formel, le critère de base ne reposera ni sur des considérations morphologico-syntaxiques ni sur des évolutions de sens. C'est la lemmatisation la plus générale ; c'est celle que nous appliquons parce qu'elle nous paraît offrir les meilleures possibilités d'observation en regroupant au maximum les diverses attestations d'un même mot, sans entrer dans des distinctions qui peuvent être intéressantes, sans doute, mais discutables. Nous voulons offrir des outils pour une réflexion sémantique et linguistique ; nous ne présageons pas des interprétations que les divers chercheurs désirent faire. Il importe de ne distinguer les points de vue morphologico-syntaxiques et sémantiques qu'à l'intérieur du regroupement maximal ; on travaillerait dès lors avec la notion de sous-lemmes. Cette technique du sous-lemme est celle du *Thesaurus linguae latinae*.

Les regroupements sous un même en-tête, à la manière du *Thesaurus* (sans cependant, comme lui, établir des sous-lemmes), peuvent présenter des inconvénients aux yeux de certains chercheurs ; nous pensons cependant que les avantages l'emportent sur les inconvénients. Il faut en effet bien convenir que dans plusieurs cas les distinctions en sous-lemmes sont discutables, et que dès lors il y a un réel inconvénient à classer séparément le verbe, le participe adjectivé et le participe substantivé, par exemple. Il peut paraître au contraire plus utile de grouper tous les emplois, précisément en vue d'une réflexion sémantique donnée. Nous sommes en effet bien souvent victimes de nos traductions : il faut laisser parler les divers emplois dans leur ambiguïté morphologico-syntaxique, laquelle correspond au système même de la langue latine. Dans le langage théologique et philosophique, il me semble aujourd'hui dangereux de cristalliser, par exemple, les formes participiales tantôt sous un lemme verbe, tantôt sous un lemme adjectif. L'expérience montre que notre lecture peut varier et qu'à tel moment nous optons dans certains cas pour l'interprétation verbale, et qu'à tel autre moment nous optons pour l'interprétation adjectivale. Cette réflexion doit demeurer ouverte. Que l'on songe à des exemples tels qu'*absolutus*, *absconditus*, ou d'autres. Bien sûr, il y a des cas qui paraissent clairs ; des distinctions peuvent être faites dans tel texte, mais pas dans tel autre. Notre outil sera valable dans la mesure où il est systématique.

Est-ce à dire que le sens n'intervient pas dans notre lemmatisation de type formel ? Nullement. Le sens n'est pas notre critère de base pour la lemmatisation, mais le sens du texte, comme, auparavant, les considérations morphologico-syntaxiques interviennent quand on se trouve face à des formes homographes. Pour donner quelques exemples, c'est la connaissance des règles morphologico-syntaxiques et la compréhension du texte qui nous permettent de distinguer l'adjectif *mundus* et le substantif *mundus*, *conditio* provenant de *condicio* (*condicere*), de *conditio* (*condere*) ou de *conditio* (*condire*), *ingenitus*, de *ingigno* ou de *ingenitus* (*in* privatif) — cf. de même tous les autres cas où *in-* correspond soit à l'*à-* privatif grec, soit à *év-*, et renvoie à des lemmes de sens contraire —, *oblitus* de *obliuiscor* ou de *oblino*, etc.

Le critère de base permettant de décréter l'homographie d'un lemme est l'histoire du mot, son étymologie. Voyez le cas cité précédemment du lemme *conditio*. Le *Dictionnaire étymologique de la langue latine* d'Ernout-Meillet est ici notre guide principal.

Tout lemme qui peut être homographe est expliqué de manière formelle, en donnant un élément flexionnel qui permet de le distinguer (ex. notation de la forme génitive du lemme — *senectus, utis* —, de l'infinitif ou du parfait d'un verbe — *dico, are* ou *dico, ere* ; *serere, serui* ou *serere, seui* ou l'étymologie (*incidere - cado* ou *incidere - caedo*) ou la catégorie grammaticale (ex. *a, praep.* ; *cum, coni.* ou *cum, praep.*) ou la quantité (*leuis, e, breuis* ou *e longa*) ou, en dernier recours, quand aucune autre possibilité précitée ne peut être utilisée, un synonyme (ex. *limus, i - caenum*). Ces annotations ont un but purement pratique ; elles servent à faire connaître aux chercheurs de quelle unité lexicale il s'agit, tout en évitant au maximum le recours à une traduction précise, nécessairement dépendante du contexte. — C'est bien la raison pour laquelle nous ne mentionnerons un synonyme qu'en tout dernier lieu. —

Cette lemmatisation est nécessairement semi-automatique. Un dictionnaire mis en mémoire nous permet de lemmatiser automatiquement dans un premier temps l'ensemble des formes ; dans un second temps, il faut revoir chacune des formes et décider dans les cas d'ambiguïtés, du lemme qu'il convient de choisir.

*

A la suite de cette lemmatisation de type formel, nous diffusons la *series B* de nos *Instrumenta Lexicologica Latina* qui comprend les parties suivantes :

1) *Enumeratio lemmatum*

C'est la liste alphabétique de tous les lemmes, conçus ainsi de manière globalisante. On trouve donc rassemblées sous une même unité lexicale toutes les variantes orthographiques d'un même mot ainsi que toutes les formes grammaticales dérivées d'un même vocable. Chaque lemme est accompagné de sa fréquence.

Cette liste est publiée sous forme imprimée et sert de manuel de référence pour les autres instruments de travail diffusés.

2) *Concordantia lemmatum et formarum*

Il s'agit ici d'une concordance de même type que la *concordantia formarum*, mais ici les formes sont groupées sous leurs lemmes, lesquels figurent dans l'ordre alphabétique. De plus les attestations sont classées rigoureusement selon l'ordre du texte afin de situer l'emploi d'un vocable dans le texte et de pouvoir en étudier l'évolution.

Sont éliminés de cette concordance tous les contextes relatifs à des entrées lexicales qui correspondent à des mots-outils (adjectifs-pronoms, prépositions, conjonctions, etc.). Pour ces mots, un renvoi est fait à la *concordantia formarum* de la *series A*, tout en donnant cette fois les diverses graphies ou les diverses formes qui correspondent à chacun de ces lemmes. Cette concordance, ainsi d'ailleurs que les autres résultats qui suivent, est diffusée sous forme de microfiches.

3) *Index formarum et lemmatum*

Cet index est une liste de renvoi systématique donnant pour chaque forme, munie de sa fréquence, le lemme auquel elle correspond. Comme on y ajoute également la fréquence du lemme, on peut se rendre compte immédiatement si une forme "épouse" un lemme, si donc toutes les attestations d'un lemme correspondent à une forme donnée. Cet index permet, en cas de difficulté pour le chercheur, de savoir quel a été le lemme retenu. Il permet également diverses observations concernant les graphies d'un texte ou ses particularités grammaticales.

4) *Index lemmatum a tergo ordinatorum*

Le lexique inverse des lemmes permet d'étudier les différents types de formation lexicale attestés.

5) *Tabula frequentiarum*

Cette *Tabula* contient la liste des lemmes classés dans l'ordre des fréquences décroissantes. Elle met en évidence les termes les plus employés et donne dès lors un premier aperçu sur la thématique de l'œuvre.

*

Tels sont les *Instrumenta Lexicologica Latina* que nous avons entrepris d'établir et de diffuser pour tous les textes patristiques latins et leurs continuations médiévales.⁽⁵⁾ Ils ont été conçus de telle manière que les chercheurs puissent disposer des instruments de travail fondamentaux que nécessitent l'étude scientifique d'un texte ou la recherche des phénomènes donnés dont on veut savoir s'ils sont attestés dans une œuvre précise (formes, ensembles de formes ou expressions, citations, lemmes, caractéristiques orthographiques, morphologiques, syntaxiques et stylistiques).

C'est l'expérience de longues années de recherche ainsi que la mise à profit de toutes les techniques disponibles aujourd'hui qui nous ont permis de concevoir cet ensemble d'instruments de lecture, grâce auxquels nous pouvons enfin échapper aux divers dilemmes auxquels nous étions assujettis jusqu'aujourd'hui : le dilemme index - concordance, le dilemme formes - lemmes, le dilemme de présentation des contextes (ordre alphabétique des expressions ou ordre du texte). On peut résolument s'engager dans la publication de larges contextes ; il faut présenter tant les entrées formes que les entrées lemmes : elles ont toutes deux leur signification et leur importance et l'on peut enfin classer un instrument de travail selon l'ordre des contextes et un autre selon l'ordre du texte. On sait en effet que les créateurs même du genre de la concordance, les exégètes, ont considéré que les deux types de présentation leur étaient indispensables, selon le type de recherche poursuivi.

En concluant, il importe de citer un Père et pourquoi pas — *ex oriente lux* — un Père grec ? De commencement en commencement, par des commencements sans fin, note Grégoire de Nazianze. D'encodage en encodage, de concordance en concordance, de lemmatisation en lemmatisation. Car il est bien certain que quand nous aurons publié les fascicules A et B pour toutes les œuvres d'un Père, nous ferons des *index indicum* et des concordances de concordances. L'opération est sans fin, puisque le fichier des Pères de l'Eglise est appelé à se compléter et à s'enrichir sans cesse, sinon de nouveaux textes, du moins de nouveaux types d'analyse. Mais le chemin même que nous parcourons est passionnant, parce que grâce aux multiples textes enregistrés et aux lemmatisations systématiques, nous entrons à chaque coup plus avant dans la découverte des structures générales de la langue et des structures particulières à la parole individuelle de chacun des Pères.

NOTES

- * La fondation du LASLA en 1961 m'a permis d'entamer, dès le 6 décembre de cette même année, l'enregistrement du premier texte latin que j'ai pu étudier à l'aide des machines. Depuis ce premier texte de Raoul de Saint-Trond dont à l'époque je commençai la perforation, que de chemin parcouru ! La fondation du Laboratoire d'Analyse Statistique des Langues Anciennes par le Professeur Louis Delatte m'a permis d'en entreprendre le début et je l'en remercie vivement.
1. On en trouvera un sommaire dans P. TOMBEUR, *Vox Latina : Belgian Initiatives in Data Processing the Intellectual Language of Europe, A.D. 197-1965*, dans *Computers and the Humanities*, 12 (1978), pp13-18.
Cf. également, ID., *Recherches en cours au Centre de Traitement Electronique des Documents*, dans *Res. III^o Colloquio Internazionale del Lessico Intellettuale Europeo*, Atti a cura di M. FATTORI et M. BIANCHI, Rome, 1982, pp. 567-578.
 2. Cf. notamment P. TOMBEUR, *Informatique et étude de textes. Pour une meilleure connaissance du vocabulaire médiolatin*, dans *Archivum Latinitatis Medii Aevi*, 40 (1977), pp. 124-138.
 3. Cf. *Index Scriptorum Operumque Latino-Belgicorum Medii Aevi. Nouveau Répertoire des œuvres médiolatines belges*, publié sous la direction de L. GENICOT et P. TOMBEUR. Première partie : *VIIIe-Xe siècle*, par A. STAINIER, Bruxelles, Académie Royale de Belgique, 1973, 203 pp. Seconde partie : *XIe siècle*, par P. FRANSEN et H. MARAITE, Bruxelles, Académie Royale de Belgique, 1976, 279 pp. Troisième partie : *XIIe siècle*, volume 1, *Oeuvres hagiographiques*, par M. McCORMICK, Bruxelles, Académie Royale de Belgique, 1977, 157 pp. Troisième partie : *XIIe siècle*, volume 2 : *Oeuvres non hagiographiques*, par M. McCORMICK, Bruxelles, Académie Royale de Belgique, 1979, 421 pp.
En 1982 nous publierons le premier tome de notre *Thesaurus Linguae Scriptorum Operumque Latino-Belgicorum Medii Aevi*. Ce *Thesaurus* permettra de connaître toutes les formes et tous les lemmes utilisés dans nos textes, ainsi que les œuvres qui attestent chacun des lemmes (cf. à ce propos P. TOMBEUR, *L'inventaire linguistique du latin médiéval belge*, dans *Proceedings of the International Conference on Literary and Linguistic Computing*, éd. Zvi MALACHI, Tel-Aviv University, 1979, pp. 347-353.
Ce projet de recherches est mené par le Comité National du Dictionnaire du Latin Médiéval qui en a confié l'exécution au CETEDOC. Il est subsidié par le Fonds National de la Recherche Scientifique et l'Académie Royale de Belgique, dont fait partie le Comité.
 4. E. DEKKERS et A. GAAR, *Clavis Patrum Latinorum*, 2e éd., Steenbrugge, 1961.
 5. Le lecteur peut obtenir auprès de la maison d'édition Brepols (8, Baron du Fourstraat, B-2300 Turnhout) tout renseignement quant à la diffusion des *Instrumenta Lexicologica Latina*. Une trentaine de fascicules paraîtront au cours de l'année qui vient.