



HAL
open science

Du Texte Latin à la Concordance Imprimée

Philippe Fleury

► **To cite this version:**

Philippe Fleury. Du Texte Latin à la Concordance Imprimée. Computers and the Humanities, 1986, Activities in France. Part II, 20 (4), pp.239-245. hal-03012950

HAL Id: hal-03012950

<https://hal.science/hal-03012950v1>

Submitted on 18 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VOLUME 20, NO. 4
OCTOBER-DECEMBER 1986

ISSN 0010-4817

Computers

AND THE

Humanities



PARADIGM PRESS, INC.
OSPREY, FLORIDA 33559

Du Texte Latin à la Concordance Imprimée

Philippe Fleury

Summary: The development of the technique of laser printing opens new ways for the writing of concordances to Latin texts, by means of electronic data processing. It is indeed possible to obtain as good a quality of printing as with traditional typography, even for a very important number of characters, which is always the case with concordances. A first step to obtaining perfection in the format of the final document consists in collecting data carefully and programming the software so as to provide scholars with a tool that will really facilitate their researches. Although the cutting out of the "useful context" is a very difficult problem to solve when using computers, as many data (punctuation, capital letters, diacritic signs . . .) as possible can be collected from the very beginning. These data will then make the text provided for each keyword more comprehensible. The programming must obviously be improved to take all the collected data into account. The reader of a concordance produced in such a way will have at his disposal a text easy to read not only because of its typographic quality, but also because of all its information on reference, punctuation, lemma, manuscript tradition; also he will be able to find in appendices pertinent frequency lists, lists of names and all sorts of lexical, morphological, syntactical information according to the work involved in the data collecting and programming.

Lorsqu'en Europe, sous l'impulsion du L.A.S.L.A.² de Liège, les techniques informatiques commencèrent d'être appliquées aux langues an-

Philippe Fleury is with the Centre d'Etudes et de Recherches pour l'Antiquité of the University of Caen.¹

ciennes, le domaine de la concordance connut une certaine désaffection. Ce fut même l'objet d'un débat animé, dont la *Revue des Etudes Latines*³ s'est faite l'écho, entre Louis Delatte, directeur à l'époque du L.A.S.L.A. et "promoteur" des indices informatisés, et Pierre Grimal, défenseur des concordances traditionnelles et éditeur lui-même de concordances de Sénèque.⁴ En fait le fond du débat n'a jamais porté sur une opposition proprement dite entre index et concordance: la supériorité de la concordance, présentant les mots à l'intérieur de leur contexte, sur l'index qui ne donne que les références est théoriquement indérivable.⁵ La question est de savoir s'il est possible et intéressant d'informatiser le travail d'élaboration d'une concordance.

Une des premières concordances informatisées publiées sur un texte latin fut celle de Tite-Live par D.W. Packard.⁶ Diversement accueillie dans le monde des latinistes, on la recomut comme un document utile pour la recherche livienne, mais on lui reprocha l'absence de lemmatisation et surtout l'extrême confusion de la présentation. Devait-on pour autant condamner définitivement toute perspective d'informatisation dans le domaine des concordances?

Depuis 1975 j'ai travaillé sur cette question pour faire en sorte que les reproches faits jusqu'ici aux publications informatisées devienrent sans fondements. Mes efforts ont porté dans trois directions: —modernisation des moyens de saisie des textes et des possibilités de mise à jour des fichiers. —affinage de la programmation pour prendre en compte le plus grand nombre possible de paramètres dans la constitution de la concordance (notamment en ce qui concerne la lemmatisation)

—utilisation des techniques de l'impression laser pour offrir une présentation rivalisant avec l'imprimerie traditionnelle.

I. Saisie des Textes et Mise à Jour des Fichiers

Il m'a paru nécessaire d'offrir au lecteur, dans le document final, une image absolument fidèle du texte de référence, en respectant non seulement la typographie (majuscules-minuscules⁷. . .) et la mise en page (pour les textes en vers notamment), mais aussi tous les signes de ponctuation simples (virgules, points-virgules, points . . .) ou complexes (parenthèses, guillemets . . .) ainsi que les signes diacritiques. Le lecteur de notre concordance de Vitruve, par exemple, peut reconnaître aussi bien au niveau des formes, du contexte, que des lemmes, les mots qui posent un problème de tradition manuscrite, celui de la concordance de Cyprien peut distinguer immédiatement les mots ou les passages considérés par les éditeurs du *Corpus Christianorum* comme des citations bibliques.

Ce souci de fidélité a pour contre-partie une complexité plus grande de la saisie des textes et de la mise à jour des fichiers: il ne s'agit pas seulement d'enregistrer les uns à la suite des autres les mots d'un texte, il faut aussi tenir compte de tous les éléments qui nous permettront une restitution exacte de l'édition de référence. Pour faire face techniquement à ces problèmes, j'ai mis au point un système d'intercommunication entre les microordinateurs du C.E.R.L.A.,⁸ l'ordinateur moyen du Centre de Calcul de l'Université de Caen⁹ et la puissante configuration du C.I.R.C.E. d'Orsay.¹⁰ Cette interconnexion me permet de concilier à la fois des impératifs d'économie (l'exploitation des micro-ordinateurs nécessite de faibles crédits de fonctionnement), de rapidité (la connexion en ligne directe avec le centre de calcul de notre université nous donne immédiatement accès à de puissants logiciels à partir des micro-ordinateurs transformés en terminaux) et de puissance (le C.I.R.C.E. nous offre toutes les capacités de calcul et de stockage mémoire dont nous pouvons avoir besoin). En effet les fichiers que nous manipulons sont parfois assez volumineux: pour la concordance des *Traité*s de Cyprien, par exemple, le fichier de départ représente 6.311.760 caractères, le fichier d'impression 12.308.400. Il est donc plus intéressant pour nous d'utiliser, à la limite de leurs possibilités, d'abord nos micro-ordinateurs, ensuite

l'ordinateur de Caen et en dernier lieu le C.I.R.C.E. Nos données sont stockées au C.I.F.C.E. sur bandes magnétiques, mais la saisie et les mises à jour sont préparées sur micro-ordinateurs, avec un logiciel de traitement de texte. J'ai écrit un programme de mise à jour en PLI permettant d'effectuer en temps différé (batch) les corrections sur bandes magnétiques.

Il faut ajouter que l'utilisation du MINITEL¹¹ me permet, comme à n'importe quel chercheur, d'intervenir à tout moment et à partir de n'importe quel local muni d'une prise téléphonique, sur mes fichiers, soit pour les interroger, soit pour les traiter directement, soit enfin pour contrôler l'exécution des programmes. Cette innovation, à mon avis très importante pour nos disciplines, détruit l'argument de ceux qui envisageaient l'informatique comme un instrument réservé à quelques privilégiés en raison du coût et de la rareté des appareils.

II. La Programmation

Une fois le texte saisi avec toutes les indications complémentaires dont je viens de parler, il s'agit de le traiter pour obtenir la concordance. J'aurais pu utiliser un programme KWIC (KeyWord In Context), comme il en existe depuis longtemps dans beaucoup de Centres de Calcul. Ce programme aurait présenté toutes les formes du texte classées en ordre alphabétique avec un contexte paramétrable avant et après le mot,¹² mais il m'a paru nécessaire d'aller plus loin et d'écrire moi-même des programmes exactement adaptés aux textes latins et aux exigences de notre discipline.

Aller plus loin signifie lemmatiser les formes, adopter le système de référencement usuel (qui n'est pas toujours le plus cohérent ni le plus logique informatiquement parlant . . .) pour l'auteur traité, délimiter le plus intelligemment possible le contexte, calculer les fréquences en tenant compte de la tradition manuscrite, mais triser totalement la mise en page.

1. La lemmatisation

L'intérêt de la lemmatisation n'est pas à démontrer,¹³ surtout pour une langue telle que le latin. Il s'agit cependant d'un travail long et délicat pour le philologue auquel l'ordinateur n'apporte qu'une assistance. Nous utilisons pour cela le programme de lemmatisation de Liège¹⁴ qui analyse automatiquement un grand nombre de

formes et propose un choix de solutions pour les autres. Deux reproches ont été faits à la lemmatisation de Liège. Ils concernent d'une part le choix du dictionnaire de référence: le *Lexicon Totius Latinitatis* de Forcellini, ouvrage vieilli et peu utilisé aujourd'hui (il est pourtant le seul complet à ce jour), d'autre part l'utilisation d'indices numériques pour distinguer les homographes; il est évident que pour un lecteur qui ne dispose pas des codes d'analyse du L.A.S.L.A. la distinction entre ET 1 et ET 2 ou entre SVSPICIO 1 et SVSPICIO 2 n'est pas immédiatement déchiffrable.

Pour remédier à cela j'ai conçu un lexique de transformation des lemmes qui remplace certaines graphies du Forcellini¹⁵ (par exemple ADSVM au lieu de ASSVM, INQVAM au lieu de INQVIO. . .) et qui substitue aux indices numériques des indications plus claires (ET (*adu.*)) et "ET (*coni.*)" au lieu de "ET 1" et "ET 2"; "SVSPICIO, onis" et "SVSPICIO, ere" au lieu de "SVSPICIO 1" et "SVSPICIO 2"). Un programme PLI effectue la transformation automatiquement.

2. La référencement

Chaque mot cité dans la concordance doit pouvoir être localisé très précisément dans le texte de référence. Or notre discipline connaît divers types de référencement suivant les auteurs et suivant les oeuvres: par livres, chapitres et paragraphes, par chapitres et lignes, par livres et vers etc. Certaines oeuvres comportent une préface, d'autres non. A l'intérieur d'une même oeuvre, un livre peut avoir une préface et le suivant non, etc. J'ai donc été amené à écrire un programme PLI tenant compte de tous ces paramètres et susceptible d'être modifié lorsque de nouveaux cas de figure se présenteront.¹⁶

Ainsi nos concordances ont une référencement semblable à celle de nos dictionnaires usuels, avec la même précision qu'une référencement transcrite "manuellement."

3. La constitution du fichier concordance

Un élément essentiel dans la constitution d'une concordance est le choix du "contexte utile" que l'auteur proposera pour accompagner chaque forme. Disons-le clairement tout de suite, jamais la machine ne pourra rivaliser avec le philologue dans ce domaine: d'un point de vue informatique, il s'agit donc de faire le mieux possible en proposant

un contexte à la fois suffisamment long pour être utile et sans mots superflus pour ne pas charger la présentation. C'est une question difficile, mais elle l'est aussi pour le philologue qui travaille manuellement. En effet ou bien le lecteur de la concordance sera intéressé par les associations de mots et alors un contexte restreint lui sera suffisant, ou il s'attachera à la sémantique et dans ce cas, c'est parfois une page qu'il faudrait résumer pour éclairer le sens d'un mot.

Les concordances de textes latins publiées à l'aide de moyens informatiques ont généralement recours à deux solutions simples: pour les textes en prose, le contexte est délimité par la longueur de la ligne d'impression (la forme étant placée au milieu), pour les textes de poésie, le contexte choisi est souvent le vers. L'inconvénient de la première solution est d'offrir une présentation trop massive, encombrée de mots inutiles. La seconde solution est certes satisfaisante pour les mots placés en milieu de vers, mais elle conduit à ces aberrations pour les mots placés en début ou en fin de vers.¹⁷

Pour les textes en prose j'utilise donc comme limites du contexte les signes choisis par l'éditeur lui-même pour marquer les unités de sens, c'est-à-dire la ponctuation "forte" (./!/?/;). Jamais je ne fais "descendre" ou "monter" le contexte après ou avant une ponctuation forte. Si cette solution ne permet pas de résoudre la question du contexte minimal, elle permet au moins d'obtenir une présentation plus aérée, sans mots assurément superflus. Je centre la forme au milieu de la ligne d'impression, puis je donne un contexte à droite et à gauche, en lui fixant pour limite maximale la longueur de la ligne d'impression¹⁸ ou la rencontre d'une ponctuation forte. Pour les textes de poésie, j'ai adopté présentation vers par vers qui permet au lecteur d'avoir immédiatement connaissance de la métrique et de la place du mot dans le vers. La forme-clef est en majuscules afin d'être repérée plus facilement. Lorsque cette forme est la première ou la deuxième du vers, je donne également le vers précédent; lorsqu'elle est la pénultième ou la dernière, je donne le vers suivant, cela toujours dans les limites de la rencontre d'une ponctuation forte.

4. La documentation complémentaire

Grâce au programme d'analyse automatique de C.I.P.L. de Liège, nous disposons dans notre fichier

de départ d'informations concernant les catégories grammaticales, les genres, les cas, les voix, les modes, les temps... de chaque mot. Cela me permet d'éditer, à la suite de la concordance, plusieurs tableaux grammaticaux donnant la répartition des lemmes et des occurrences entre chaque catégorie grammaticale, et, à l'intérieur des catégories, entre les sous-catégories. Ces tableaux sont accompagnés de calculs de pourcentages.

La lemmatisation et l'enregistrement des signes diacritiques me permettent aussi de fournir des listes de fréquence pertinentes, éliminant par exemple du comptage les formes rejetées par l'éditeur (présentées cependant pour information à l'intérieur de la concordance avec un astérisque), distinguant le comptage des formes propres à l'auteur et celui des formes propres aux citations éventuellement. J'ai également écrit un programme pour comparer des listes de fréquence d'œuvres ou d'auteurs différents et isoler ainsi des vocabulaires particuliers.

En marge de la documentation lexicographique et grammaticale, je me suis attaché aussi à informatiser la documentation bibliographique concernant chaque auteur traité. J'utilise à cet effet depuis 1984 le logiciel documentaire TEXTO¹⁹ qui offre toutes les facilités de création, de mises à jour, d'interrogation et d'impression. Pour Perse et Vitruve, nous avons imprimé ces bibliographies au début de la concordance, mais elles continuent d'être mises à jour.

III. L'Impression Laser

En impression classique (à impact), l'auteur d'un ouvrage volumineux se trouve toujours devant un dilemme: il faut choisir entre la rapidité (donc l'économie) et la qualité.²⁰

Pour la troisième de nos concordances, celle de Cyprien, j'ai utilisé la technique de l'impression au laser, qui, à mon avis, est en train de modifier actuellement de façon très sensible les perspectives de publication des travaux informatisés, en conciliant la rapidité des imprimantes classiques les plus performantes avec une qualité comparable à celle de l'impression traditionnelle.

Le principe de cette technique est le suivant: un faisceau laser se déplaçant dans les deux dimensions de la page²¹ est activé ou désactivé tous les 1/300 inch sur le matériel utilisé au C.I.R.C.E. (XEROX 9700). A chaque activation correspond un

point noir sur la page. La capacité totale d'une page, exprimée en points, est théoriquement de 2478 x 3507 points. En termes de résolution on peut parler de 300 x 300 = 90.000 points par inch,² ce qui donne une excellente qualité graphique. En fonctionnement standard (ligne à ligne), l'imprimante laser du C.I.R.C.E. a un débit nominal d'une page par seconde, quel que soit le taux de remplissage de cette page.

Le premier avantage de cette technique est donc l'association de la qualité et de la rapidité. Pour avoir un aperçu de la qualité obtenue, on consultera en Annexe une page de la concordance du *Quod Idola*, une œuvre contestée de Cyprien, que nous avons éditée à la suite de la concordance générale de ses traités. Cette concordance générale comporte près de 80.000 occurrences; à raison de 83 lignes par page, elle occupe 1.118 pages et il a fallu à peine 19 minutes pour l'imprimer (soit plus de 4.800 lignes à la minute).

Le deuxième avantage, et non le moindre, est l'entière liberté laissée à l'auteur de choisir la forme et la taille des caractères, leur espacement ainsi que celui des lignes, la dimension des marges, l'orientation de la page, etc. Les perspectives ouvertes par ce type d'impression sont donc les mêmes que celles de l'imprimerie classique. L'auteur peut mêler divers types de caractères: romains et italiques par exemple, utiliser le gras et le souligné, imprimer les caractères grecs... tout cela sans rien changer à la vitesse de l'impression.

A ce matériel de pointe est associé, au C.I.R.C.E., un système de traitement de texte, qui permet de rédiger introductions, en-têtes, notes, tables des matières et de les imprimer au laser avec les possibilités définies ci-dessus.

Pour peu que l'on veuille s'en donner la peine, il paraît donc possible aujourd'hui d'imprimer avec les moyens informatiques des concordances de qualité, au même titre que celles produites "artisanalement," mais avec un gain de temps très appréciable et une grande sûreté. Certes le contexte est découpé moins "intelligemment" par la machine que par le philologue et le recours au texte s'avère parfois nécessaire, comme il s'avère aussi nécessaire (un peu moins souvent il est vrai) dans les concordances "manuelles." Mais là nous rejoignons le débat concordance-index: la concordance est sûrement un instrument plus perfectionné,

cependant elle occupe un volume environ trois fois plus important²² que celui d'un index. Le choix de l'outil dépend finalement des objectifs de l'utilisateur; l'idéal étant de pouvoir disposer des deux pour chaque auteur. Néanmoins, pour les deux types de travaux, il faut avoir les mêmes exigences de qualité pour un travail produit par les moyens informatiques que pour un livre classique. J'ai essayé de montrer que les techniques actuelles le permettent et si les précurseurs avaient quelques excuses dans ce domaine, on ne devrait plus pouvoir reconnaître aujourd'hui, en l'ouvrant en son milieu, un livre sorti des "presses de l'ordinateur."

NOTES

1. Dans le cadre du Centre d'Etudes et de Recherches pour l'Antiquité (anciennement C.E.R. de Langues Anciennes) de l'Université de Caen, un groupe de travail, dirigé par Louis Callebat, Professeur de Latin, s'applique depuis 1975, en collaboration avec le L.A.S.L.A. de Liège (cf. note infra), au traitement automatique des textes anciens. Les travaux entrepris tendent vers un double objectif:

I. Contribuer à fournir (pour les recherches lexicales, grammaticales, stylistiques) des instruments utiles d'analyse et d'interprétation: concordances, vocabulaires sélectifs, relevés grammaticaux, analyse des énoncés.

II. Constituer et exploiter une documentation spécialisée touchant un domaine particulier: celui des langages scientifiques et techniques latins.

Les oeuvres de Perse, Vitruve, Apulée (*Apologie, Florides*), Cyprien (*Traité*), Tertullien (*De Pallio*), Lactance, Duden de St. Quentin, Guillaume de Jumièges, Guillaume de Poitiers sont actuellement en mémoire au C.E.R.L.A.

Trois concordances ont été éditées:

—P. Bout, L. Callebat, Ph. Fleury, M. Zuinghedau, *Konkordanz zu den Satiren des Persius Flaccus*, Hildesheim/New York, Olms, 1978.

—id., *Vitruve, De Architectura. Concordance*, Hildesheim/Zürich/New York, Olms/Weidmann, 1984.

—P. Bouet, Ph. Fleury, A. Goulon, M. Zuinghedau (avec la collaboration de P. Dufraigne), *Cyprien. Traité. Concordance*, ibid., 1985.

2. On ne saurait trop insister sur le rôle de promotion joué par le Laboratoire d'Analyse statistique des Langues anciennes de Liège, au début des années 60, dans le domaine de l'étude des langues anciennes par ordinateurs. Sur l'histoire du L.A.S.L.A., ses objectifs, ses méthodes, voir par exemple E. Evrard *Le L.A.S.L.A.: "RELO"* 1970, 4, p.1-16. C'est dans ce laboratoire, devenu depuis 1983 le C.I.P.L. (Centre Informatique de Philosophie et de Lettres) que je me suis initié en 1973 à l'application des méthodes informatiques au latin, et son actuel directeur, Joseph Denooz, a guidé avec beaucoup de bienveillance et de compétence mes premiers pas dans le domaine de la programmation en PLI. Je tiens ici à l'en remercier.

3. Cf. P. Grimal, *Index et Concordances: "REL"* XLIV, 1966, p.108-116.

4. L. *Annaei Senecae Operum Moraliu Concordantia*, P. Grimal cura edita, Publications de la Faculté des Lettres et Sciences Humaines de Paris, Série "Textes et documents" VIII, Paris, P.U.F., fasc.1 (*Ad Marciam. De consolatione.*) paru en 1965.

5. V. Lomanto (*Concordantiae in Q. Aurelii Symmachi Opera*, Hildesheim/New York, Olms/Weidmann, 1983, p.V.) oppose à tort concordances non lemmatisées et indices lemmatisés. En réalité, on peut très bien avoir des concordances lemmatisées et des indices non lemmatisés. La question de la lemmatisation est indépendante, à mon avis, du choix entre index et concordance (cf. infra). Mais son ouvrage est un exemple intéressant d'association des techniques informatiques avec un travail de finition classique permettant d'obtenir un outil remarquable au point de vue de la forme et du contenu.

6. D.W. Packard, *A Concordance to Livy*, Cambridge (Mass.), Har-

vard Un. Pr., 1968 (4 vd.). A chaque mot de texte correspond une ligne complète dans la concordance (une vingtaine de mots en moyenne). Les formes ne sont pas lemmatisées et les groupes de formes identiques ne sont pas isolés les uns des autres. Il n'y a donc pas de comptage de fréquence. L'absence de lemmatisation est justifiée par D.W. Packard dans son introduction par l'importante intervention humaine qu'elle aurait demandée, et qui est exact: l'oeuvre complète de Tite Live représente tout de même 505.000 mots et il a certainement été plus utile aux spécialistes de Tite Live de disposer rapidement d'une concordance non lemmatisée que de devoir attendre des dizaines d'années un travail lemmatisé. Il faut reconnaître aussi que depuis, des travaux non lemmatisés de qualité bien inférieure ont été publiés: la concordance de J. Klecka (*Concordantia in Publum Papinium Statium*, Hildesheim, Olms, 1983) par exemple allie contenu insuffisant, présentation et typographie médiocres.

7. Conformément aux conventions de l'édition française et de beaucoup d'éditions internationales, nous respectons la majuscule pour la première lettre des mots commençant une phrase. Ce détail est absent de beaucoup de travaux informatisés. Un travail imprimé uniquement en majuscules comme celui d'E. Siedschlag, *Martial-Konkordanz*, Hildesheim, Olms, 1979, est très désagréable à consulter.

8. Un Zenith 100 et un Zenith 150 compatible IBM PC.

9. MINI 6/43 BULL.

10. Un NAS9080 bprocesseur et un ADAHL470/V7. (C.I.R.C.E. = Centre Inter Régional de Calcul Electronique.)

11. Certes ce terminal, distribué par les Postes et Télécommunications Françaises, offre un clavier et un écran peu confortables, mais des améliorations sont promises pour l'avenir. L'absence de traitement direct des caractères accentués, qui est admissible pour les matériels conçus aux USA, est tout de même un peu surprenant pour un matériel et un système de conception française.

12. Cf. R. Busa, *Concordantiae Senecae*, Hildesheim, Olms, 1975 (2 vol.).

13. Il existe dans nos disciplines un courant de pensée refusant la lemmatisation pour des raisons d'objectivité. Je ne reprendrai pas ici l'ensemble de ce débat qui est développé ailleurs (voir par exemple "RELO" *passim*); je me contenterai simplement de souligner trois points:

—il est vrai que la lemmatisation introduit une certaine dose de subjectivité par les choix que l'auteur est amené à faire pour quelques regroupements (cas par exemple des participes adjectifs en "to" qu'un philologue rattachera à un lemme d'adjectif, qu'un autre rattachera au lemme du verbe...). Mais la subjectivité ne commence-t-elle pas dès le choix d'une édition de référence et n'est-ce pas au fond la nature de tout travail de recherche d'être l'ouvrage d'un auteur ou d'un groupe d'auteurs?

—par contre, je partage l'opinion de ceux qui disent que, pour une oeuvre volumineuse, il vaut mieux disposer rapidement d'un travail non lemmatisé que d'attendre des années le résultat d'une lemmatisation. Ceci étant je ne pense pas que l'on puisse considérer un index ou une concordance non lemmatisé comme une fin en soi, mais comme une étape vers un travail plus élaboré.

—l'idéal serait que les travaux non lemmatisés soient simplement enregistrés sur supports magnétiques et consultés par tout chercheur qui en ferait la demande. Cette idée de banques de données consultables "à la carte" a été exprimée il y a déjà longtemps, par E. Evrard notamment qui la rappelle à la fin d'un compte-rendu récent (*Trois concordances récentes d'auteurs latins: "RELO"* 1983, p.225).

14. Cf. J. Denooz, *Recherches sur le traitement automatique de la langue latine: "RELO"* 1973, 1.

15. Notre base de référence en matière de graphes des lemmes est l'*Oxford Latin Dictionary*, fasc.1-8, Oxford, Clarendon Press, 1968-82. Les lemmes non cités dans ce dictionnaire sont choisis, dans tous les cas possibles, en référence au *Thesaurus Linguae Latinae*.

16. Lors de la saisie du texte, des codes sont enregistrés qui permettent de repérer: les fins de chapitre, de paragraphe, de ligne etc. Le programme lui-même est paramétré pour tenir compte des préfaces, des lignes ou des paragraphes qui ne commencent pas au numéro 1, etc.

17. Admettons par exemple que le mot vedette soit le premier du vers et qu'il soit suivi d'une ponctuation forte le seul contexte

intéressant sera le vers précédent. Cf. Virg. *Aen.* 2,13:

Incipiam. Fracti bello fatisque repulsi

Seul le contexte de 2,12 est utile pour interpréter *incipiam*:

Quamquam animus meminisse horret luctuque refugit,

18. Je me limite à une seule ligne par forme pour des raisons de volume global de l'ouvrage, mais aussi pour des questions de clarté. Il est à noter que ce principe n'est pas propre aux concordances informatisées: cf. H. Quellet, *Concordance verbale du DE CORONA de Tertullien*, Hildesheim/New York, Olms, 1975. La longueur de la ligne est paramétrée dans le programme PLI de constitution de la concordance et elle peut être fixée selon le format choisi. Pour la concordance de Vitruve, j'ai utilisé des lignes de 118 caractères: 12 caractères pour la référencement, 106 pour le contexte, ce qui représente en moyenne une quinzaine de mots.

19. Le logiciel TEXTO est le fruit de la collaboration d'une équipe de recherche de CHEMDATA (Lyon) avec le laboratoire du G.I.S. (Groupement d'Intérêt Scientifique) de la Maison de l'Orient (C.N.R.S.) à Lyon.

20. Il est vrai que des progrès ont été réalisés par l'impression rapide classique: nos concordances de Perse et de Vitruve ont été sorties (en double impression) sur les imprimantes rapides du L.A.S.L.A. et elles offrent, je crois, une typographie acceptable. Cependant, en poussant le souci de la qualité à l'extrême, la meilleure impression à impact s'obtient aujourd'hui avec les imprimantes dites "qualité courrier," mais elles n'impriment pas plus de 20 caractères à la seconde: à ce rythme il nous aurait fallu près de 3000 heures pour imprimer une concordance de la taille de celle de Vitruve!

21. La page standardisée au C.I.R.C.E. a le format A4: 21 x 29,7 cm (8,26 x 11,69 inches).

22. L'index comporte le même nombre de lignes théoriques qu'une concordance (une pour chaque forme, plus une pour chaque lemme dans le cas de travaux lemmatisés), mais elles peuvent être présentées sur trois colonnes.

ANNEXE

A (AB) 11

20 4 uisitur et sepulcrum eius ostenditur, et AB eo Saturnum fugatum manifestum est: inde
 20 15 uaria deorum religio mutatur, dum non unus AB omnibus Deus colitur, sed propriaeque
 20 18 potestatis proditum sibi de diis hominibus A sacerdote secretum, quod maiorum e regum
 21 10 Mox A nescio quo Febris dedicata et Accaet
 23 13 posteaquam terrenis uitiis inmersi sunt et A uigore caelesti terreno contagio
 24 16 cessat iniuria: nec aliud studium est quam A Deo homines auocare et ad superstitionem
 25 1 Deo homines auocare et ad superstitionem sui AB intellectu uerae religionis auertere, cum
 29 10 Itaque cum Christus Iesus secundum A prophetis ante praedicta uerbo et locis
 30 11 spiritum sponte dimisit et die tertio rursus A mortuis sponte surrexit.
 31 3 ut hominem, quem dilexit, quem induit quem A morte protexit, ad Patrem uictor
 31 7 monente diffusis praecepta in salutem dare, AB errore tenebrarum ad uiam lucis adducere.

AC, c. conl.; & uid. ATQVE 3

22 1 utique si quid potestatis habuissent, sua AC suorum regni seruassent.
 30 5 crucem eius et mortem suffragiis uiolentis AC pertinacibus flagitarent.
 31 9 AC ne esset probatio minus solida et de

ACCA N 1

21 11 Mox a nescio Febris dedicata et ACCA et Flora meretrices.

ACCENDO 1

30 2 et doctrina ille et sapientia reuincebat, ACCENSA ira et indignatione prouocati, ut

ACCIDO, (cado) 1

22 5 Regna autem non merito ACCIDVNT, sed sorte uariantur.

ACCIPIO 4

20 3 nec mercedem operis infelix instructor ACCEPIT.
 20 5 est: inde Latium de latebra eius nomen ACCEPIT.
 23 3 aut auguriis Romana regna creuerunt, sed ACCEPITVM tempus certo fine custodiunt.
 28 4 de diuinis muneribus haurire: quam ACCEPITAM Iudae contemptis religionibus

AD 22

19 4 Inde illis instituta templa, inde AD defunctorum uultus per imaginem
 20 17 ita esse Alexander magnus insigni uolumine AD matrem suam scripsit, metu suae
 22 8 Ceterum si AD originem releas, erubescas.
 22 11 nunc ut rex ipse principatum habeat AD crimina, filii Romulus parricida, quae ut
 22 13 Rapiunt, ferociunt, fallunt AD copiam ciuitatis augendam: nuptiae sunt
 23 16 norunt et Socrates instrui se e regi AD arbitrium diemonii praedicabat, et magis
 24 1 daemonii praedicabat, et magis inde est AD perniciose uel ludicra potentatum: quorum
 24 13 ualitudinem frangunt, morbos lacessunt, ut AD cultum sui cogant, ut nidore altarium et
 25 1 studium est quam a Deo homines auocare et AD superstitionem sui ab intellectu uerae
 25 2 cum sint ipsi poenales, quaerere sibi AD poenam comites, quos ad crimen sum
 25 3 quaerere sibi ad poenam comites, quos AD crimen suum fecerint errore participes.
 25 16 AD diuinum imprium etiam de terris mutuemur
 28 11 duorum hominem induit, quem perducatur AD Patrem.
 30 8 mortem sed ut uinceret, et cum passus esset, AD superos denique regredi, ut uim diuinae
 30 14 substantiae corporalis firmitate conspicuus AD dies quadraginta remoratus est, et de eo
 31 1 ad dies quadraginta remoratus est, ut de eo AD praecepta uitalia instrui possent et
 31 3 dilexit, quem induit, quem a morte protexit, AD Patrem uictor inponeret, iam uenturus e
 31 4 uictor inponeret, iam uenturus e caelo AD poenam diaboli et ad censuram generis
 31 4 iam uenturus e caelo ad poenam diaboli et AD censuram generis humani uictoris rigore et
 31 7 in salutem dare, ab errore tenebrarum AD uiam lucis adducere, caecos et ignaros ad
 31 8 ad uiam lucis adducere, caecos et ignaros AD agnitionem ueritatis oculare.
 31 12 ut Christus Dei Filius, qui hominibus AD uitam datus creditur, non tantum

ADDVCO 1

31 7 dare, ab errore tenebrarum ad uiam lucis ADDVCERE, caecos et ignaros ad agnitionem

ADFINIS 1

26 3 Pompeius et Caesar ADFINES fuerunt: nec tamen necessitudinis

ADFLATVS 1

24 8 adque imaginibus consecratis delitescunt, hi ADFLATV suo uiaum pectora inspirant extorum