

STUDIIIS	088	1	3	STVDIVM	12N00	6	00
	088	1	3	STVDIVM	12C00	6	00
QUID	088	1	4	QVIS	1 47A	6	00
	088	1	4	QVIS	1 47C	6	00
	088	1	4	QVIS	2 48A00	6	00
	088	1	4	QVIS	2 48C00	6	00
SENTIAM	088	1	5	SENTIO	54A13		00
	088	1	5	SENTIO	54A31		00
SCIRE	088	1	6	SCIO	5DB21		00
	088	1	6	SCIO	54071		00
DESIDERAS	088	1	7	DESIDERO	51B11		00
NULLUM	088	1	8	NVLLVS	48A00	6	00
	088	1	8	NVLLVS	48C00	5	00
SUSPICIO	088	1	9	SVSPICIO	1 13A00	2	00
	088	1	9	SVSPICIO	2 55A11		00
NULLUM	088	1	10	NVLLVS	48A00	6	00
	088	1	10	NVLLVS	48C00	5	00
IN	088	1	11	IN	70300		00
	088	1	11	IN	70600		00
BONIS	088	1	12	BONI	12N00	4	00
	088	1	12	BONI	12000	4	00
	088	1	12	BONVM	12N00	6	00
	088	1	12	BONVM	12000	6	00
	088	1	12	BONVS	21N00	1	00
	088	1	12	BONVS	21000	1	00
NUMERO	088	1	13	NVMERO	1 51A11		00
	088	1	13	NVMERO	2 60000		00
QUOD	088	1					00
	088	1					00
	088	1					00
	088	1					00
	088	1					00
	088	1					00
	088	1	14	QVOD	1 81000		00
	088	1	14	QVOD	2 820		00

l.a.s.l.a.

l'ordinateur et le latin

Techniques et méthodes

Morphologie, syntaxe, lexicologie

Stylistique

PREFACE

Lors du premier Colloque pour l'étude des problèmes de la didactique du grec et du latin, organisé à Gand en 1963, le Professeur L. Delatte, conscient de l'aide que l'ordinateur pouvait apporter à l'étude de la langue latine, tant au point de vue de l'apprentissage de la langue qu'en ce qui concerne le commentaire du texte, proposa de réaliser, en collaboration étroite avec l'Enseignement secondaire, une étude statistique de tous les auteurs latins qui figurent au programme des Humanités.

Ce travail vient de s'achever par la publication d'un volume consacré aux œuvres philosophiques de Cicéron. On trouvera en annexe, la liste des ouvrages que le Ministère de l'Education Nationale, Organisation des Etudes, a mis à la disposition des professeurs.

Au moment où l'Enseignement secondaire et l'Enseignement supérieur pédagogique entraînent en possession de cet instrument de travail complet, forgé au fil des années et selon des méthodes sans cesse affinées, il importait qu'une information scientifique fût donnée aux professeurs appelés à s'en servir. C'était aussi le moment de dire les conditions et les limites d'utilisation des informations statistiques et de décrire les progrès pédagogiques qu'elles favorisent.

Tous les professeurs de langues anciennes de l'Enseignement secondaire et de l'Enseignement supérieur pédagogique ont été invités à participer, au L.A.S.L.A., à des journées d'études au cours desquelles M. le Professeur L. Delatte et ses collaborateurs, Mademoiselle S. Govaerts, Chef de travaux et M. J. Denooz, Premier assistant, ont présenté les trois communications qui font l'objet de la présente publication.

Ces communications rendent compte des méthodes scientifiques du L.A.S.L.A. et de leurs prolongements didactiques. Elles rencontrent les objectifs de l'enseignement des langues anciennes, tels que le programme ministériel les définit, et leur apportent une aide précieuse ().*

* Programme 73/41, Première partie, M.E.N., 1978.

La communication de M. Denooz a pour objet de décrire les techniques et les méthodes du L.A.S.L.A. Son exposé fait ressortir les ressources immenses de l'informatique linguistique à différents points de vue: la rapidité d'exécution et la facilité d'accès aux données enregistrées, la rigueur de l'information scientifique. Un exemple caractéristique est fourni par l'étude des œuvres de la littérature grecque ancienne, dont l'analyse statistique a commencé récemment (*).

Après avoir décrit en détail le contenu des ouvrages réalisés au L.A.S.L.A., Mademoiselle Govaerts, dans une synthèse des relevés statistiques concernant la morphologie, la syntaxe et le vocabulaire, a proposé des changements méthodologiques visant un apprentissage rationnel de la langue latine, une compréhension et une étude plus profondes des textes. Un lexique de base des auteurs étudiés dans l'Enseignement secondaire et une grammaire fréquentielle du latin classique seront bientôt les auxiliaires pédagogiques irremplaçables d'un enseignement efficace.

Le Professeur Delatte montre des voies nouvelles et enrichissantes, tant pour les maîtres que pour leurs élèves. Sur le plan culturel, comme pour l'action pédagogique, sa communication donne un exemple de l'évolution et de l'adaptation indispensables de notre enseignement. Par le recours aux méthodes et aux techniques de diverses sciences humaines, le professeur de langues anciennes suscitera l'analyse du message linguistique: l'élève prendra conscience des réalités de l'œuvre ancienne et des aspects humains de sa composition; en même temps, il découvrira ce qu'il y a de permanent et d'évolutif dans la pensée occidentale et dans son expression verbale.

La lecture de ces communications provoquera des réflexions, voire des réactions diverses. On voudra bien que soit soulignée ici une des conclusions qui ponctueront les journées d'études.

Un enseignement secondaire moderne se caractérise notamment par des fondements scientifiques que les travaux du L.A.S.L.A. lui apportent dans le domaine des langues anciennes. Ces travaux, outre le fait qu'ils renforceront les objectifs et les méthodes du programme d'étude, augmenteront l'efficacité pédagogique des manuels de latin et la valeur formatrice des éditions scolaires et des anthologies.

* On trouvera, p. , la liste des œuvres grecques qui ont été analysées. Le choix des textes est dû à des objectifs scientifiques.

Le professeur devrait trouver dans l'emploi judicieux des documents statistiques plus de sécurité et, partant, plus d'autonomie pédagogiques.

L'élève sera associé à une démarche exemplaire, puisqu'elle sera conforme à des normes scientifiques et axée sur des objectifs qu'elles rendent possibles ou imposent. Il sera sollicité par un enseignement dont les matières grammaticales ne seront pas une fin en soi, mais lui donneront très souvent l'occasion de mettre en œuvre les moyens de son intelligence.

L'enseignement classique doit tous ces progrès à une collaboration suivie entre les chercheurs du L.A.S.L.A. et des professeurs de l'Enseignement secondaire. Tous ceux qui firent partie des équipes successives méritent l'estime et la reconnaissance de leurs collègues. Au demeurant, nous ne pourrions dire combien nous sommes redevables au Professeur Delatte et à ses collaborateurs de leur dévouement inlassable pour l'Enseignement secondaire, d'autant plus que l'exploitation didactique des recherches scientifiques fera l'objet de tous leurs soins.

Au Ministère de l'Education Nationale, il y eut nombre de personnes qui comprirent l'intérêt de nos recherches et qui nous apportèrent sans réserve leur soutien. Nous voulons saisir cette occasion de les remercier.

Nous exprimons notre vive reconnaissance à Monsieur le Secrétaire Général Vanbergen, qui fut Directeur Général de l'Organisation des Etudes et à ses collaborateurs, notamment Messieurs Dupont et Goffart, dont la générosité et les encouragements ne se sont jamais démentis.

Il conviendrait de citer encore bien d'autres responsables ministériels qui nous aidèrent: c'est le cas de Monsieur le Directeur Général de l'Enseignement Supérieur F. Dethier qui facilita la création de C.R.I.D.E.L.A., Centre de Recherches, Information et Documentation pour l'Etude des Langues Anciennes, association qui publie les documents destinés aux écoles.

Nous n'avons jamais fait appel en vain à Messieurs les Inspecteurs généraux. Monsieur l'Inspecteur général G. Dethier, accompagné de M. Goffart, nous faisait l'honneur de sa présence au cours de la première journée d'information.

Mais rien n'aura de pleine signification qu'au moment où tout sera fait.

En attendant, nous souhaitons aux professeurs qui enseignent les langues anciennes et dont nous connaissons bien les aspirations et les difficultés, de découvrir dans cette publication une source d'enrichissement personnel et des raisons de persévérer avec nous.

René LECROMPE

Docteur en Philosophie et Lettres

Inspecteur des Enseignements secondaire et supérieur.

TECHNIQUES ET METHODES

par

J. DENOZ,

Chef de Travaux, L.A.S.L.A.

Université de Liège

PRESENTATION DU L.A.S.L.A.

Le Laboratoire d'Analyse statistique des Langues anciennes a été créé pour apporter une solution à des problèmes que la philologie traditionnelle est impuissante à résoudre : je veux parler des problèmes d'authenticité et de chronologie relative des oeuvres d'un auteur. En outre, nous croyons que les éléments recueillis par des analyses formelles permettront de déceler les mécanismes profonds de la création littéraire à tous les niveaux, même les plus inconscients.

Pour atteindre ces objectifs, il faut se fonder sur l'étude statistique des mots d'un texte aux points de vue phonétique, lexicologique, syntaxique et stylistique. Nous avons donc résolu de faire une analyse morphologique et syntaxique complète de chaque forme ainsi qu'une lemmatisation, c'est-à-dire le renvoi du mot-forme au mot tel qu'il figure au dictionnaire. Une telle recherche suppose l'établissement de fichiers extrêmement complets dont la réalisation et l'exploitation exigent le recours aux machines mécano-graphiques et à l'ordinateur. L'emploi de ces machines entraîne l'utilisation

d'un langage simplifié qui permet de condenser certaines informations sous forme de codes.

Je commencerai par décrire brièvement la carte mécanographique (annexe 1).

Dans une carte mécanographique, les informations sont représentées sous forme de perforations. La carte se divise verticalement en 80 colonnes numérotées de 1 à 80 et horizontalement en 12 niveaux. Dans chaque colonne, on peut perforer un caractère d'imprimerie numérique ou alphabétique. La signification d'une perforation dépend du niveau où elle se situe. La position des colonnes et des niveaux est indiquée, sur les cartes de modèle standard, par des chiffres imprimés, sauf pour les niveaux 12 et 11 situés dans la partie supérieure de la carte : cette partie est souvent utilisée pour l'impression d'autres indications, notamment pour la traduction en clair du contenu des colonnes.

Une perforation à l'un des trois premiers niveaux (12, 11, zéro) est appelée perforation de zoning, une perforation à l'un des niveaux de 0 (zéro) à 9 est appelée perforation numérique ou de digit. La perforation zéro est mixte puisqu'elle peut jouer le rôle de zoning pour les signes alphabétiques et le rôle de digit pour les signes numériques.

Les chiffres sont représentés par une des dix perforations numériques. Quant aux caractères alphabétiques, ils exigent la présence dans une même colonne d'une perforation de zoning (12, 11 ou 0) et d'une perforation numérique (1 à 9).

Voici un aperçu général du dessin de la carte (Annexe 3 - 2e carte).

1	Code-carte
2-17	Lemme
18	Indice de lemme
19-38	Forme du texte
39-40	Code d'oeuvre Il s'agit de deux perforations permettant d'identifier l'oeuvre à laquelle le mot appartient.
41-53, 64,68-75	Référence du mot
41-43	Numéro du chapitre (001, 002 etc.)
44-47	Numéro du paragraphe ou du vers (0001, 0002 etc.)
48-50	Numéro d'ordre du mot dans le paragraphe ou dans le vers (001, 002 etc.)
51-53	Numéro d'ordre du mot dans la phrase (001, 002 etc.)
64	Code du rôle pour les oeuvres dramatiques
68-70	Indication éventuelle de la subdivision en livres
71-75	Numéro d'ordre du mot dans l'oeuvre (00001, 00002 etc.)
54-63	Analyse complète de la forme
65	Indications relatives à la tradition manuscrite
66	Ponctuation du texte
67	Colonne libre
76-80	Numéro d'ordre du mot dans l'index.

Il me faut maintenant, à propos d'un certain nombre d'informations contenues dans la carte, préciser les conventions que nous avons adoptées.

1. Nous appelons forme le mot tel qu'il figure dans le texte de l'édition que nous avons choisie. Nous reproduisons le texte de cette édition aussi fidèlement que possible, à quelques rares exceptions près sur lesquelles je n'insisterai pas ici.

2. Par lemme, nous entendons le mot tel qu'il figure au dictionnaire. Nous utilisons comme ouvrage de référence le *Lexicon Totius Latinitatis* de FORCELLINI. Ce dictionnaire nous a paru être l'instrument de travail le mieux adapté à nos recherches.

En ce qui concerne le lemme, nous avons mis au point un certain nombre de conventions qui sont indispensables pour lire un index et que le lecteur trouvera en annexe 4.

Je décrirai maintenant notre système d'analyse morphologique.

Chaque forme fait l'objet d'une analyse qui est perforée dans les colonnes 54-63 (annexes 5 et 6).

Colonne 54

On note, dans cette colonne, la catégorie grammaticale à laquelle le mot appartient. 1 substantif, 2 adjectif ...

Colonne 55

Selon la manière dont le mot a été analysé en colonne 54, on précise en colonne 55 :

- a) la sous-catégorie (classe, conjugaison, etc.);
- b) le degré;
- c) la voix.

Pour les substantifs, un chiffre indique la déclinaison à laquelle le substantif appartient.

Pour les adjectifs et les verbes nous avons utilisé, pour économiser de la place, un tableau à deux entrées.

Pour les adjectifs, la codification indique à la fois la classe de l'adjectif et son degré.

- 1 adjectif en -Vs (1ère classe)
- A adjectif 1ère classe au comparatif
- J adjectif 1ère classe au superlatif

Pour les verbes on indique en colonne 55 à la fois la conjugaison et la voix.

- 1 1ère conjugaison voix active
- A 1ère conjugaison voix passive
- J 1ère conjugaison déponent

La colonne 56 est réservée aux indications suivantes :

- 1) le cas et le nombre pour les formes déclinées;
- 2) la personne et le nombre pour les formes conjuguées;
- 3) le cas régi pour les prépositions.

Ex. A pour un subst. représente le nominatif singulier

A pour un verbe représente la 1ère pers. du singulier

Colonne 57

- 1) Pour les verbes, cette colonne est réservée à l'indication du mode.
- 2) Pour les mots introduisant un verbe de proposition subordonnée (adjectifs-pronoms, adverbess relatifs et interrogatifs, conjonctions de subordination), on note le mode du verbe régi.

Colonne 58

- 1) Pour les verbes, cette colonne contient l'indication du temps. Il va de soi que l'adjectif verbal, le gérondif et le supin ne portent aucune indication dans cette colonne.
- 2) Pour les mots introduisant un verbe de proposition subordonnée, on note

le temps du verbe régi.

Colonne 59

Les cartes de verbes reçoivent l'indication de la fonction :

Perforation 12 verbe de proposition indépendante ou principale

Perforation 11 verbe de proposition subordonnée

La colonne 61 est réservée à l'indication du genre pour les adjectifs, les numéraux, les adjectifs-pronoms et les formes déclinées du verbe.

Nous prenons en considération non pas le genre que le mot a effectivement dans le texte, mais le genre formel en dehors de tout contexte. Cette façon de procéder nous permet de réaliser aisément la mise en ordre des formes dans l'index.

Dès lors, nous distinguons les six catégories suivantes :

- 1 genre commun (masculin, féminin et neutre)
- 2 féminin
- 3 masculin et féminin
- 4 masculin
- 5 masculin et neutre
- 6 neutre

Ainsi, *bonum*, à l'accusatif, sera codifié 5, tandis que le nominatif *bonum* recevra le code 6.

Colonnes 62-63

Ces colonnes reçoivent une indication uniquement dans le cas des verbes subordonnés. Nous avons symbolisé chaque subordonnant par un code formé de 2 caractères alphabétiques. Ainsi, LN indique que l'on a affaire à un verbe dépendant du pronom relatif QVI.

La liste des subordonnants accompagnés de leur code se trouve en annexe 6. Cette liste n'est certes pas exhaustive. Lors de chaque étude de texte, nous sommes amenés à y faire des additions. C'est pourquoi, nous n'avons pas utilisé jusqu'à présent toutes les lettres de l'alphabet dans la seconde colonne.

Colonne 65

Les indications relatives à la tradition manuscrite sont codifiées en colonne 65 : par exemple, 5 représente une croix.

Colonne 66

La colonne 66 contient des renseignements qui permettent à l'ordinateur de calculer automatiquement la référence de chaque forme du texte.

Le code utilisé est alphanumérique; les lettres correspondent aux cas où deux renseignements relatifs à la ponctuation doivent être donnés simultanément.

11 fin de chapitre et du §

K fin du chapitre, du § et de la phrase etc.

Après avoir exposé les conventions relatives à l'enregistrement des textes latins sur cartes mécanographiques, je voudrais donner un bref aperçu historique qui montrera comment, progressivement, le Laboratoire s'est développé.

Au début, nous n'avions que des machines classiques : une perforatrice, une trieuse, une petite calculatrice et un système d'impression automatique.

Lorsque la dactylo avait perforé dans la carte la forme du mot telle qu'elle apparaît dans le texte (annexe 2, carte 1) et lorsque les cartes avaient été référencées sur la calculatrice (annexe 2, carte 2), chaque membre de l'équipe prenait une partie du fichier à sa table de travail et, à la main, écrivait sur chaque carte, dans des emplacements prévus à l'avance, le lemme (annexe 3, carte 3) et l'analyse complète du mot (annexe 3, carte 4).

On se rend compte aisément de la somme de travail que représentaient ces analyses. Elles exigeaient une attention sans défaillance jusque dans les plus petits détails; elles ne souffraient pas la moindre approximation. C'est la raison pour laquelle un philologue ne pouvait guère analyser plus de deux cents à deux cent cinquante mots par jour. Au surplus, nous nous étions astreints à de nombreuses vérifications, vérifications de la codification de l'analyse et de la perforation de cette analyse sur les cartes.

De 1961 à 1965, soit dans un espace de quatre ans, nous avons publié cinq volumes consacrés aux Dialogues de Sénèque. Pendant ces années, désespérés par la lenteur, la lourdeur de l'analyse, nous avons cherché des moyens d'accélérer le travail en l'automatisant au maximum. Une telle automatisation

supposait l'acquisition de moyens techniques plus puissants que les simples machines classiques, c'est-à-dire l'emploi d'un ordinateur. Nous avons obtenu du F.N.R.S. un petit ordinateur scientifique IBM 1620. Cet ordinateur est arrivé en janvier 1965. Après avoir suivi des cours de programmation et nous être familiarisés avec le maniement de la machine, nous avons réalisé en huit mois un programme d'analyse automatique de la morphologie du latin, et, plus récemment, un programme de lemmatisation du grec.

Peut-être faut-il rappeler brièvement ce qu'est un ordinateur.

Un ordinateur est, à l'origine, une simple machine comptable capable, d'une part, d'effectuer les opérations arithmétiques traditionnelles et, d'autre part, de procéder à des comparaisons entre des valeurs qu'elle a emmagasinées dans une mémoire.

A partir de cette conception de base, tout le progrès a été consacré à augmenter la puissance de la machine dans le domaine de la mémoire et dans le domaine de la rapidité des opérations. Pour augmenter la rapidité, il a été nécessaire de réduire la taille des éléments constitutifs de la machine où la seule longueur des câbles de connexion joue un rôle primordial dans le déroulement des opérations.

Dans ce but, l'effort des constructeurs a porté sur la miniaturisation. En effet, les premières machines, vers les années 50, étaient des monstres à lampes qui consommaient autant d'électricité pour fonctionner que pour se refroidir (240 kw par heure pour le Gamma Bull 60). La deuxième

génération, qui est apparue vers 1960 était équipée de transistors et avait une taille nettement moindre. Actuellement la troisième génération d'ordinateurs est équipée de circuits intégrés analogues à ceux qui équipent les montres digitales, où, sur un centimètre carré, sont groupés 30 à 40.000 transistors.

L'ordinateur du Centre de Calcul de l'Université de Liège est une machine de cette 3e génération (IBM 370/158). Il est très puissant; en fait, c'est une des machines les plus puissantes du monde. Sa mémoire centrale a une capacité de plus d'un million cinq cent mille caractères. Mais cette mémoire ne constitue qu'une faible partie de sa capacité totale. En effet, le 370 est doté de mémoires périphériques gigantesques sous forme de disques et de bandes magnétiques qui représentent plus de quarante milliards de caractères. Pour faire comprendre la puissance de cet instrument, on peut dire que les mémoires globales du 370 équivalent à une bibliothèque de 20.000 ouvrages de 1.000 pages chacun. En d'autres termes, un lecteur qui consacrerait 10 heures par jour, 365 jours par an à lire cette bibliothèque devrait vivre 456 ans et encore 200 jours.

L'ordinateur, quant à lui, lit cette bibliothèque en moins de 10 minutes.

Cette machine ne travaille pas uniquement dans le lieu où elle est installée. Elle possède des ramifications à travers tous les bâtiments universitaires. Ces ramifications sont en fait des terminaux travaillant simultanément. Un terminal est un appareil relié à un ordinateur central par une ligne téléphonique spéciale qui permet le transfert d'informations à grande vitesse et avec un maximum de sécurité. On apprendra avec étonnement qu'à l'Université de Liège, il y a 55 terminaux qui fonctionnent simultanément.

Le L.A.S.L.A. possède actuellement 4 terminaux. Il s'agit tout d'abord d'un terminal lourd, c'est-à-dire d'une machine qui possède un lecteur de cartes qui traite les cartes à la cadence de 600 à la minute, un perforateur de cartes (9.600 perforations à la minute ou 120 cartes) et une imprimante semi-rapide à 600 lignes/minute. On ne se rend pas compte de ce que signifie "600 lignes à la minute". Cela signifie qu'un volume de 1.000 pages, à raison de 60 lignes par page, peut être imprimé en moins de 2 heures. Par ailleurs, ce terminal possède une machine à écrire qui permet de dialoguer avec l'ordinateur lui-même.

Outre cet équipement, le L.A.S.L.A. possède 3 terminaux légers. L'un est constitué par une simple machine à écrire qui sert essentiellement au traitement du grec, le deuxième est un écran cathodique et enfin, le troisième est un terminal acoustique qui permet d'entrer directement en communication avec l'ordinateur, par l'intermédiaire du téléphone.

Ce dernier appareil préfigure l'avenir puisque dans un délai de trois ans maximum, on pourra parler à l'ordinateur et il répondra oralement.

Je vais décrire maintenant le programme d'analyse automatique du latin : le but du système est de faire produire par l'ordinateur la lemmatisation et l'analyse codée de toutes les formes d'un texte et, pour les formes ambiguës, tous les lemmes et toutes les analyses qui sont formellement possibles : soit, par exemple, la forme QUAE qui reçoit douze analyses différentes :

QUAE	QUI pronom relatif	nom. fém. sing. nom. fém. plur. nom. neutre plur. acc. neutre plur.
	QUIS adjectif-pronom interrogatif	nom. fém. sing. nom. fém. plur. nom. neutre plur. acc. neutre plur.
	QUIS adjectif-pronom indéfini	nom. fém. sing. nom. fém. plur. nom. neutre plur. acc. neutre plur.

Pour effectuer ces opérations d'analyse et de lemmatisation, l'ordinateur fait subir au texte cinq traitements successifs.

1. Dans une première phase, le texte entier, perforé sur cartes mécanographiques, est enregistré sur disques magnétiques. Chaque mot reçoit à ce

moment un numéro d'ordre séquentiel qui permettra de le reclasser dans l'ordre du texte à n'importe quel moment. En outre, chaque forme est éventuellement suivie d'un code de ponctuation qui servira au calcul automatique de la référence.

2. La deuxième phase du travail consiste à classer les formes du texte en ordre alphabétique. Cette opération a pour but de diminuer la durée de la phase suivante. En effet, partant de l'ordre alphabétique intégral, l'ordinateur analyse la première forme qu'il rencontre et reproduit ensuite les analyses autant de fois qu'il y a de formes identiques dans le texte. Ainsi, si la forme EST se trouve cent fois dans un texte, l'ordinateur procède une seule fois à la recherche de ses analyses puis il les recopie cent fois à très grande vitesse sans consultation nouvelle du lexique. L'économie de temps que l'on réalise ainsi est considérable puisque, pour un texte de 10.000 mots, ce procédé réduit presque de moitié le nombre de formes à analyser.
3. Au cours de la troisième phase s'effectuent la lemmatisation et l'analyse.
4. La quatrième phase consiste à reclasser les formes en ordre du texte à l'aide du numéro séquentiel qui leur a été attribué lors de la première phase.
5. Finalement, dans une cinquième phase, chaque analyse est imprimée et est enregistrée sur bande magnétique. C'est également au cours de cette phase que chaque mot du texte reçoit sa référence complète.

Les deux premières phases du système ainsi que les deux dernières ont un rôle secondaire puisqu'elles consistent, d'une part, à classer les mots d'un texte dans l'ordre le mieux adapté au traitement et, d'autre part, à présenter les résultats d'une manière telle que le philologue puisse facilement les examiner. En revanche, la troisième phase occupe une place prépondérante puisque c'est à ce moment que chaque forme est lemmatisée et analysée. Pour effectuer ces opérations, l'ordinateur doit au préalable, avoir reçu tous les éléments qui lui permettront d'agencer ses réponses.

Notre système d'analyse se fonde sur la distinction entre la partie du mot qui reste constante à travers toute la flexion et la lettre ou le groupe de lettres susceptibles d'affecter tous les lemmes appartenant à une même flexion. Nous désignons ces deux parties respectivement par les termes *radical* et *désinence* en donnant à ces mots une signification purement conventionnelle sans aucun lien avec la grammaire. Si l'on prend, par exemple, la forme AMABAT, on dira que AM- est le radical de AMO, AMARE et que -ABAT est la désinence de la troisième personne du singulier de l'indicatif imparfait actif des verbes de la première conjugaison. C'est pourquoi il est nécessaire d'enregistrer dans les mémoires de l'ordinateur, deux Thesaurus, c'est-à-dire un lexique et une table des désinences. Le programme les consultera constamment pour identifier les formes du texte. En plus des données linguistiques, ces fichiers contiennent des informations d'ordre technique qui guident le travail de l'ordinateur. On trouvera en annexe 7 et 8, des échantillons du lexique et de la table des désinences. En voici une brève description.

Le lexique doit contenir les éléments qui permettront à l'ordinateur de

reconnaître un mot quelconque de la langue, de le rattacher à un lemme et enfin de déterminer pour ce mot, la ou les analyses possibles.

Dans le dictionnaire de la machine, une unité lexicale est composée du radical d'un mot et de son lemme. Elle intervient dans l'analyse de toutes les formes de ce mot. Dès lors, avec un lexique relativement restreint (environ 16.000 unités), l'ordinateur peut lemmatiser et analyser plus de 90 pourcents des formes de n'importe quel texte latin.

Le principal intérêt du processus de décomposition réside dans l'économie appréciable d'unités lexicales qu'il permet de réaliser. Il existe toutefois certains mots pour lesquels cette méthode est soit inutile, soit inapplicable.

Ce sont, par exemple, les mots invariables qui sont représentés dans la langue par une forme unique : les conjonctions, les prépositions et certains adverbes.

Ce sont aussi toutes les formes anormales. Elles sont traitées de la même manière que les mots invariables : chaque unité lexicale de forme irrégulière conduit à une seule analyse : c'est le cas des diverses formes du verbe ESSE.

Enfin, pour terminer la description du lexique, il convient de noter que pour des raisons d'économie, les verbes composés ne font pas l'objet d'unités lexicales séparées. En effet, leur analyse et leur lemmatisation sont réalisées à partir des verbes simples et d'une table des préverbes mise en mémoire.

Un mot maintenant sur *la table des désinences*.

Après avoir décomposé une forme en radical et terminaison et après avoir découvert une unité lexicale compatible avec le radical, l'ordinateur consulte la table des désinences pour y découvrir l'analyse du mot.

Dès lors, la table des désinences contient obligatoirement toutes les terminaisons des mots à flexion et, parallèlement, toutes les analyses morphologiques possibles; c'est pour cette raison que les désinences formellement identiques figurent dans la table autant de fois qu'elles donnent lieu à des analyses différentes. En procédant de la sorte, nous avons dénombré quelque 3.100 désinences différentes. (cf. annexe 8).

Les trois premières colonnes sont rigoureusement identiques aux colonnes 3 à 5 des unités lexicales. Cette identité est fondamentale dans le processus d'analyse puisque c'est elle qui permettra de préciser la série de désinences qui est compatible avec chaque radical.

La désinence occupe les positions 6 à 14, les colonnes 17 à 22 contiennent l'analyse morphologique complète de la désinence. Ces colonnes correspondent aux colonnes 54 à 58 de notre code d'analyse.

Tels sont les deux thesaurus qui permettent d'organiser le programme proprement dit de l'analyse automatique du latin. C'est ce programme que je voudrais décrire maintenant.

J'ai dit précédemment que, pour produire le lemme et l'analyse d'un mot latin, le programme exécute essentiellement deux opérations :

- 1- la décomposition de la forme,
- 2- la consultation du lexique et de la table des désinences.

La décomposition de la forme s'effectue à partir de la fin du mot en remontant progressivement lettre par lettre vers le début. Chaque nouvelle décomposition est suivie d'une consultation de la table des désinences.

	natura	-e	rien	} de NATURA
naturae	natur-	-ae	Génitif singulier	
			Datif singulier	
			Nominatif pluriel	
	natu	-rae	rien	
	nat-	-urae	nascor	-participe futur de nasci
	nat-	-urae	no	-participe futur de nare

Le lecteur trouvera en annexe 9 un échantillon d'analyse corrigée par le philologue. A ce stade, en effet, le philologue reprend les listings d'analyse et il indique par un signe quelle est la bonne analyse. Comme les différentes analyses sont numérotées, il suffit à la dactylo d'indiquer à la machine le numéro de l'analyse correcte.

On remarquera l'extraordinaire degré d'amphibologie de la langue puisque chaque forme est susceptible de recevoir en moyenne trois analyses.

On le comprend d'autant mieux si l'on se rappelle que les formes QVAE et NATURAE, peuvent recevoir chacune 12 analyses.

Pour terminer cette partie de l'exposé, je voudrais dire un mot de l'efficacité du programme. Notre lexique comporte environ 16.000 unités, ce qui représente quelque 8.500 mots différents. Malgré cette faible étendue du dictionnaire, la machine parvient à analyser plus de 90 pourcents de n'importe quel texte latin, qu'il s'agisse de prose ou de poésie, de latin classique, de latin de la Renaissance ou de l'époque moderne : c'est ainsi que nous avons analysé les Regulae de Descartes et le Tractatus Politicus de Spinoza avec le même succès.

En ce qui concerne la rapidité d'exécution, un ordinateur de la 2e génération analysait 10.000 mots en une nuit. L'ordinateur que nous utilisons actuellement traite ces mêmes 10.000 mots en moins de 12 minutes.

Par ailleurs, il est très important de noter que, lorsque l'analyse a été corrigée par le philologue, tout le reste du travail, c'est-à-dire la constitution de l'Index, des listes de fréquence, des concordances, des relevés statistiques, est entièrement automatique.

Un Laboratoire qui s'intitulait des Langues anciennes se devait de traiter aussi les textes grecs.

Le traitement du grec par machines mécanographiques et par ordinateur présente certaines difficultés liées, d'une part, à l'alphabet et, d'autre part, à l'existence d'accents, d'esprits et d'iota souscrit etc... En ce qui concerne l'alphabet, les codes de perforations du latin pouvaient difficilement répondre à nos besoins en raison du fait qu'ils ne suivaient pas l'ordre alphabétique de la langue grecque. C'est ainsi que γ se plaçait non pas entre β et δ mais avant la lettre η . C'est pourquoi nous avons adopté une codification des lettres grecques qui répond parfaitement à leur succession dans l'alphabet.

Le problème des esprits et des accents était plus difficile à résoudre. Cependant, il était impossible de les ignorer, comme le font les philologues américains, parce qu'ils sont porteurs d'informations linguistiques. Nous avons donc établi un code de perforation capable de représenter tous les graphèmes du grec ancien. Ce code comporte, outre les caractères alphabétiques, les divers accents et esprits, l'iota souscrit, le tréma et éventuellement le digamma.

Parallèlement à ces difficultés de perforation liées à l'enregistrement, nous avons dû rechercher une solution au problème de l'impression des résultats : les imprimantes d'ordinateur ne possèdent pas de chaîne de caractères pour le grec ancien; nous avons donc demandé à la firme IBM d'adapter pour nos besoins deux systèmes 870 qui se composent chacun d'une perforatrice connectée à une machine à écrire automatique, pour perforer et imprimer le grec.

Les méthodes que nous utilisons pour le grec sont très simples. Nous nous sommes limités, jusqu'à présent, à lemmatiser les formes sans procéder à la

décomposition en radical et terminaison, en indiquant uniquement la catégorie grammaticale des formes (cf. annexe 10). Nous avons procédé de façon purement empirique : voici comment.

Au départ, nous disposions de deux textes lemmatisés de façon artisanale : une tragédie d'Euripide, Electre et Saint Jean Chrysostome, la Lettre d'exil. Ce sont ces deux textes que nous avons mis en mémoire pour constituer le premier dictionnaire automatique, avec lequel nous avons ensuite traité les lettres à Olympias de Saint Jean Chrysostome. Toutes les formes de cette oeuvre qui ne se trouvaient pas dans le lexique initial ont été lemmatisées par un philologue et ajoutées au dictionnaire de l'ordinateur.

Cette même technique a été utilisée pour le Corpus Hermeticum que nous venons de publier et pour la Métaphysique d'Aristote qui est en cours de réalisation ainsi que pour les orateurs attiques que nous étudions en collaboration avec l'Université de Besançon.

Actuellement, notre dictionnaire grec automatique comprend approximativement 36.000 formes différentes, accompagnées de leur lemme, ce qui nous permet de lemmatiser entre 80 et 85 pourcents des occurrences de n'importe quel texte.

Encore faut-il ajouter que, pour une oeuvre donnée, une fois les premières pages analysées et les mots inconnus introduits en machine, le pourcentage de formes analysées automatiquement, pour cette oeuvre, atteint 95 %.

Enfin, pour terminer cet exposé, nous avons repris dans les annexes 10, 11, 12 et 13, la liste des textes latins et grecs que le L.A.S.L.A. a étudiés en utilisant les méthodes et les programmes que nous avons décrits dans cet article.

J. DENOOZ

AMO AMABAT 04002024058017 2 05875

A	LEMME	S-L	FORME	COE	REFERENCE	ANALYSE	P	N° D'ORDRE
0	0		0	0	0	0	0	0
1	1		1 A J / 1	1	1	1	1	1 (OEUVRE) (INDEX)
2	2		2 K S Z	2	2	2	2	2
3	3		3 C L T 3	3	3	3	3	3
4	4		4 D M U 4 (12) (11) (0)	4	4	4	4	4
5	5		5 E N V 5	5	5	5	5	5
6	6		6 F O W 6	6	6	6	6	6
7	7		7 G P X 7	7	7	7	7	7
8	8		8 H Q Y 8	8	8	8	8	8
9	9		9 I R Z 9	9	9	9	9	9

IBM BELGIUM S.A. - B 375C

LEMMES S-L

3

AMO AMABAT 04002024058017 BN 2 05875

A	LEMME	S-L	FORME	COE	REFERENCE	ANALYSE	P	N° D'ORDRE
0	0		0	0	0	0	0	0
1	1		1 A J / 1	1	1	1	1	1 (OEUVRE) (INDEX)
2	2		2 K S Z	2	2	2	2	2
3	3		3 C L T 3	3	3	3	3	3
4	4		4 D M U 4 (12) (11) (0)	4	4	4	4	4
5	5		5 E N V 5	5	5	5	5	5
6	6		6 F O W 6	6	6	6	6	6
7	7		7 G P X 7	7	7	7	7	7
8	8		8 H Q Y 8	8	8	8	8	8
9	9		9 I R Z 9	9	9	9	9	9

IBM BELGIUM S.A. - B 375C

LEMMES S-L

4

ANNEXE 3

CONVENTIONS DU L.A.S.L.A.

Les lemmes homographes sont distingués par un indice numérique selon un système de classement qui nous est propre. La connaissance de ces conventions est nécessaire pour lire avec fruit les index des auteurs du programme.

1. D'une manière générale, l'ordre adopté pour l'attribution des indices de lemme correspond à la succession numérique des catégories et des sous-catégories grammaticales dans notre codification. Nous attribuons l'indice 1 au lemme qui se range le premier dans cette succession.

EXEMPLES: TAMEN, taminis recevra l'indice 1 parce que les substantifs sont codifiés 1;
TAMEN conjonction recevra l'indice 2 parce que les conjonctions sont codifiées 8.

DICO, -are recevra l'indice 1 parce que les verbes de la première conjugaison sont codifiés 51,
DICO, -ere recevra l'indice 2 parce que les verbes de la troisième conjugaison sont codifiés 53.

2. Lorsque deux lemmes appartiennent à la même catégorie grammaticale, nous les distinguons d'abord en fonction de l'ordre alphabétique des radicaux ou, si les radicaux sont identiques, en fonction de la quantité.

EXEMPLES: OS, oris sera affecté de l'indice 1,
QS, ossis sera affecté de l'indice 2,
les composés en -CIDO de CADERE recevront l'indice 1,
les composés en -CIDO de CAEDERE recevront l'indice 2.

3. En l'absence de tout autre élément distinctif, c'est le lemme le plus fréquent qui reçoit l'indice 1.

EXEMPLES: TEMPVS 1, le temps; TEMPVS 2, la tempe;
MOROR 1, s'attarder; MOROR 2, être fou.

4. Nous avons également distingué par un indice, les emplois de ESSE et de IRE comme auxiliaires.

L'indice 1 est réservé aux cas où ces verbes sont employés en tant que verbes pleins, l'indice 2 aux cas où ils sont auxiliaires.

5. Tous les noms propres sont suivis d'un sigle alphabétique, en général la lettre N, tandis que l'adjectif reçoit le sigle A, signifiant adjectif dérivé de nom propre.

EXEMPLES: LATINVS N désigne le roi Latinus,
LATINVS A, l'adjectif LATINVS, -A, -VM.

Lorsque deux homographes appartiennent à la même catégorie grammaticale, nous les distinguons en faisant précéder le sigle N d'un indice numérique choisi en fonction de leur fréquence.

EXEMPLE: GALLI 1N, les Gaulois; GALLI 2N, les Galles.

ANNEXE 4

54	55	56	57	58	59	65	66
Catégorie grammaticale	Sous-catégorie Degré Voix	Cas Personne Nombre	Mode	Temps	Fonctions	Tradition	Ponctuation
Substantif 1	1e décl. 1 2e décl. 2 3e décl. 3 4e décl. 4 5e décl. 5 Anomal 6 Décl. gr. 7				Verbe principal 12 Verbe subord. 11	Leçon contestée 1 Mot absent dans certains mss. et adopté par édit. 2 Conjecture reçue 3 Lacune comblée CRUX 4 Leçon rejetée 5 Mot présent dans certains mss. et rejeté par édit. 6 Mot impossible à analyser en fonction d'une lacune 7	K dernier mot du chap. du paragraphe ou du vers et de la phrase S dernier mot du paragraphe ou du vers et de la phrase 2 dernier mot de la phrase 11 dernier mot du chapitre et du paragraphe ou du vers 0 dernier mot du paragraphe ou du vers
Adjectif 2	P C S 1e classe 1 A J 2e cl. cons. 2 B K -er 3 C L -is 4 D M imp. 5 E N Anomal 6 F O Décl. gr. 7				60 Emplois		
Numéral 3	P C S Cardinal 1 Ordinal 2 B K Distributif 3 Multipl. 4 Adv. ord. 5 E N Adv. mult. 6	S P Nominatif A J Vocatif B K Accusatif C L Génitif D M Datif E N Ablatif F O Locatif G P Indéclinable Z			Emploi subst. 1 adject. 2 adverb. 6		
Adjectif-pronom 4	Personnel 1 Possessif 2 Réfléchi 3 Possessif réfléchi 4 Démonstratif 5 Relatif 6 Interrogatif 7 Indéfini 8		Mode	Temps			
Verbe 5	A P D S-D 1e conj. 1 A J 2e conj. 2 B K S 3e conj. 3 C L T 4e conj. 4 D M 4e bis 5 E N Anomal 6 F O	S P 1e A J 2e B K 3e C L	du verbe subordonné régi par un pronom relatif ou interrogatif (cf. verbe)	Indicatif 1 Impératif 2 Subjonctif 3 Participe 4 Adj. verbal 5 Gérondif 6 Infinitif 7 Supin en -UM 8 Supin en -U 9	Présent 1 Imparfait 2 Futur simple 3 Parfait 4 Plus-que-parfait 5 Futur antérieur 6 -us fui/uerim/fuisse 7 -us fueram/fuissem 8 -us fuero/urus fuisse 9	61 Genre - Tmèse	
Adverbe 6	Relatif 6 Interrogatif 7 Négatif 8 Interrogatif-négatif 9 Comparatif 12 Superlatif 11		Mode	Temps	a) Genre commun 1 féminin 2 masc. et fém. 3 masculin 4 masc. et neutre 5 neutre 6		
Préposition 7	Type MECUM 1	Cas régi 3, 4 ou 6			b) Tmèse 12		
Conjonction 8	Coordination 1 Subordination 2		Mode	Temps			
Interjection 9			du verbe subordonné régi par une conjonction de subordination (cf. verbe)				

CODE ALPHABÉTIQUE			
	12	11	Zéro
1	A	J	/
2	B	K	S
3	C	L	T
4	D	M	U
5	E	N	V
6	F	O	W
7	G	P	X
8	H	Q	Y
9	I	R	Z

La lettre O sera notée Ø
Le zéro sera noté 0
La conjonction du digit 1 et du zoning zéro est représentée non par une lettre mais par le signe /

ANNEXE 5

62 - 63

CODES DE SUBORDINATION

subjonctif seul complément	AA	qualiscumque	GD	quomodo (rel.)	PN
ablatif absolu	AD	quam (rel.)	GG	quomodo (int.)	PS
proposition infinitive	AG	quam (int.)	GK	quoniam	PX
ac	AK	quamdiu (adv. int.)	GN	quoquo	RA
proinde ac si	AN	quamdiu C-S	GS	quot (rel.)	RD
perinde ac si	AS	quamdudum	GX	quot (int.)	RG
an	AX	quamlibet	HA	quotcumque	RK
an an	BA	quamobrem (rel.)	HD	quotiens (rel.)	RN
anne	BD	quamobrem (int.)	HG	quotiens (int.)	RS
antequam	BG	quamquam	HK	quotienscumque	RX
antequam	BK	quamvis	HN	quotquot	SA
cum	BN	quando ADV	HS	quotus	SD
cumcumque	BS	quando C-S	HX	quotuscumque (rel.)	SG
cur	BX	quandocumque	JA	quotusquisque	SK
donec	CA	quandoque	JD	quousque	SN
dum	CD	quandoquidem	JG	seu	SS
dummodo	CG	quantopere (rel.)	JK	si	SX
dumtaxat	CK	quantopere (int.)	JN	sicut	TA
etiamsi	CN	quantulus (rel.)	JS	simulac	TD
etsi	CS	quantulus (int.)	JX	sin	TG
licet	CX	quantuluscumque	KA	siquidem	TK
modo	DA	quantus (rel.)	KD	sive	TN
ne C-S	DD	quantus (int.)	KG	tametsi	TS
-ne	DG	quantuscumque	KK	tamquam	TX
-ne an	DK	quare (rel.)	KN	tamquam si	WA
-ne anne	DN	quare (int.)	KS	ubi (rel.)	WD
-necne	DS	quasi	KX	ubi (int.)	WG
nedum	DX	quatenus (rel.)	LA	ubi C-S	WK
neve, neu	EA	quatenus (int.)	LD	ubicumque	WN
ni	ED	quemadmodum (rel.)	LG	unde (rel.)	WS
nisi	EG	quemadmodum (int.)	LK	unde (int.)	WX
nonne	EK	qui (rel.)	LN	undecumque	XA
num	EN	qui (int.)	LS	ut (adv. rel.)	XD
postquam	ES	qui (abl. rel.)	LX	ut (adv. int.)	XG
posteaquam	EX	qui (abl. int.)	MA	ut C-S	XK
priusquam	FA	quia	MD	ut si	XN
proquam	FD	quicumque	MG	utcumque	XS
prout	FG	quin C-S	MK	uter (rel.)	XX
qua (rel.)	FK	quippe	MN	uter (int.)	YA
qua (int.)	FN	quis	MS	utercumque	YD
quacumque	FS	quisnam	MX	uti (adv. rel.)	YG
qualis (rel.)	FX	quisquis	NA	uti (adv. int.)	YK
qualis (int.)	GA	quo (rel.)	ND	uti C-S	YN
		quod	NG	utqui	YS
		quod (non) quo(d)	NK	utrum ...an	YX
		quoad (adv. int.)	NN	utrum ...necne	ZA
		quoad	NS	velut	ZD
		quocumque	NX	velut si	ZG
		quod	PA		
		quominus	PD		
			PG		
			PK		

ANNEXE 6

9CVSTODIO	CVSTODIT	01548 0- -
9CVSTODIO	CVSTODIV	01544 0- -
9CVSTOS	CVSTOS	0113 2- -
9CVTIS	CVT	01133 0- -
9CVTIO	CVT	0255 00 -
9CVTIS	CVTIS	0113 2- -
9CYCNVS	CYCN	01123 0- -
9CYCNVS	CYGN	01123 0- -
9CYMA	1CYM	0111 0- -
9CYMA	2CYMA	0113 3- -
9CYMA	2CYMAT	01136 0- -
9CYMBA	CYMB	0111 0- -
9CYMBALVM	CYMBAL	01126 0- -
9CYNICI	NCYNIC	01123-0- -
9CYNICVS	NCYNIC	0221 0- -
9CYPVVS	NCYPR	01123 0- -
9CYRVS	NCYR	01123 0- -
9DEVS	D	01123-0- -
9DO	D	0251 0# -
9DO	D	0353 00 -
9DAEDALVS	DAEDAL	0121 0- -
9DEMIVRGVS	DAMIVRG	01123 0- -
9DAMNVM	DAMN	01126 0- -
9DAHND	DAMN	0251 0- -
9DAMNO	DAMNAT	01518 0- -
9DAMNATIO	DAMNATI	0113 0- -
9DAMNO	DAMNAV	01514 0- -
9DAMNOSVS	DAMNOS	0121 0- -
9DAPS	DAP	01133 0- -
9DAPS	DAPS	0113 2- -
9DAREVS	NDARE	01123 0- -
9DAREVS	NDARI	01123 0- -
9DO	DAT	01518 0# -
9DEA	DE	0111 0- -
9DEVS	DE	02123 0- -
DE	DE	0370600-1 -
DEA	DEABVS	0111000- -
DEA	DEABVS	0211N00-1 -
9DEBEO	DEBE	0152 0- -
9DEBILIS	DEBIL	0124 0- -
9DEBILITAS	DEBILIT	0113 0- -
9DEBILITO	DEBILIT	0251 0- -
9DEBILITO	DEBILITAT	01518 0- -
9DEBILITO	DEBILITAV	01514 0- -
9DEBITVM	DEBIT	01126 0- -
9DEBITOR	DEBIT	0213 0- -
9DEBEO	DEBIT	03528 0- -
9DEBEO	DEBV	01524 0- -
9DECOR	DEC	0113 0- -
9DECET	DECF	0152 0- -
DECEM	DECEM	0131700-1 -
9DECEMBER	NDECEMBER	0113 2- -
9DECEMBER	ADECEMBER	0223 2- -
9DECEMBFR	NDECEMBFR	01133 0- -
9DECEMBER	ADECEMBFR	0223 0- -
9DECENS	DECENS	0125 3- -
9DECENS	DECENT	0125 0- -
9DECEPTORIVS	DECEPTORI	0121 0- -
DECIENS	DECIENS	0136000-1 -
DECIENS	DECIES	0136000-1 -
9DECIMVS	DECIM	0132 0- -
9DECLIVVS	DECLIV	0121 0- -

ANNEXE 7

Tusculanes, V, 1: Quintus hic dies, Brute, finem faciet Tusculanarum disputationum, quo die est a nobis ea de re quam tu ex omnibus maxime probas disputatum.

5	1	1	1	QUINTUS	QVINTVS	32A00	4	1
5	1	1	2	HIC	HIC	1 45A00	4	1
					HIC	2 63000		2
5	1	1	3	DIES	DIES	15A00		1
					DIES	15J00		2
					DIES	15L00		3
5	1	1	4	BRUTE	BRUTE	60000		1
5	1	1	5	FINEM	BRUTE VS N FINIS	12B00 13C00		1
5	1	1	6	FACIET	FACIU	55C13	0	1
5	1	2	1	TUSCULANARUM	TUSCVLANVS N	21M00	2	1
5	1	2	2	DISPUTATIONUM	DISPVATIO	13H00		1
5	1	2	3	QUO	QVI	1 46F14	5	1
					QVIS	1 47F	5	2
					QVIS	2 48F00	5	3
					QVO	1 660		4
					QVO	2 660		5
					QVO	3 670		6
					QVO	4 60J00		7
					QVO	5 820		8
5	1	2	4	DIE	DIE	60000		1
					DIES	15F00		2
5	1	2	5	EST	EDD	2 56C11		1
					SVM	1 56C11		2
					SVM	2 E6C11		3
5	1	2	6	A	AB	70600		1
					AH	90000		2
5	1	2	7	NOBIS	NOS	41N00		1
					NOS	41000		2
5	1	2	8	EA	EA	60000		1
					IS	45A00	2	2
					IS	45F00	2	3
					IS	45J00	6	4
					IS	45L00	6	5
5	1	2	9	DE	DE	70600		1
5	1	2	10	RE	RES	15F00	0	1
5	1	3	1	QUAM	QVAM	1 660		1
					QVAM	2 670		2
					QVI	1 46C11	2	3
					QVIS	1 47C	2	4
					QVIS	2 48C00	2	5
5	1	3	2	TU	TV	41A00		1
					TV	41B00		2

TEXTES LATINS

Apulée, Florides

Le traité de l'Asclepius dans le Corpus Hermeticum (8748 mots)

Catulle, Carmina (13108 mots)

César, Bellum Gallicum (45748 mots)

Cicéron, Catilinaires (12488 mots)

Pro Caecina (11980 mots)

Pro Rabirio (3552 mots)

Pro Lege Manilia (6698 mots)

De Natura deorum (± 25000 mots)

Cornelius-Nepos (21000 mots)

Descartes, Compendium Musicae

Regulae (20707 mots)

Horace, Art Poétique (1300 mots)

Chant Séculaire (339 mots)

Epîtres (10450 mots)

Epodes (3093 mots)

Odes (13660 mots)

Satires (14694 mots)

Juvénal, Satires (26402 mots)

Ovide, Métamorphoses (extraits-10259 mots)

Perse, Satires (4646 mots)

Plaute, Amphitruo (9447 mots)

Asinaria (7807 mots)

Aulularia (6662 mots)

Bacchides (9551 mots)

ANNEXE 10

Casina (7328 mots)
 Captivi (8514 mots)
 Quinte-Curce (72497 mots)
 Saint Cyprien, Quod Idola
 Ad Donatum
 De Immortalitate } (± 7000 mots)
 Salluste, Catilina (10950 mots)
 Bellum Jugurthinum (21453 mots)
 Sénèque, La Consolation à Helvia (6775 mots)
 La Consolation à Marcia (8384 mots)
 La Consolation à Polybe (5688 mots)
 De la Clémence (8283 mots)
 De la Constance (5354 mots)
 Des Bienfaits (45591 mots)
 De la Brièveté de la Vie (6213 mots)
 De la Colère (22612 mots)
 Des Loisirs (1962 mots)
 De la Providence de Dieu (4142 mots)
 De la Tranquillité de l'âme (7629 mots)
 De la Vie Heureuse (7313 mots)
 Lettres à Lucilius (119698 mots)
 Agamemnon (5719 mots)
 Hercule Furieux (7818 mots)
 Hercule sur l'Oeta (11617 mots)
 Médée (5809 mots)
 Oedipe (6088 mots)
 Phèdre (7355 mots)

ANNEXE 11

Les Phéniciennes (4226 mots)
Thyeste (6448 mots)
Les Troyennes (7073 mots)
Pseudo-Sénèque, Octavie (5326 mots)
Spinoza, Tractatus Politicus (21081 mots)
Tacite, Agricola (6880 mots)
 La Germanie (5867 mots)
 Le Dialogue des Orateurs (6391 mots)
 Les Annales (90682 mots)
Tertullien, De Pallio (3463 mots)
Tibulle, (12635 mots)
Tite-Live, (extraits-17598 mots)
Virgile, Bucoliques (5713 mots)
 Enéide (1 à 6 - 32210 mots)
 Géorgiques (14834 mots)
Vitruve, Livres 5 (6700 mots)
 7 (5716 mots)
 8 (5896 mots)
 10 (± 8000 mots)

ANNEXE 12

TEXTES GRECS

- Saint Jean Chrysostome : La lettre d'exil (8523 mots)**
Les lettres à Olympias (29485 mots)
La Providence de Dieu (20535 mots)
- Hesychius de Jerusalem : (37753 mots)**
- Corpus Hermeticum : (36566 mots)**
- Aristote : Métaphysique (3 livres - 20054 mots)**

ANNEXE 13

MORPHOLOGIE, SYNTAXE, LEXICOLOGIE

par

S. GOVAERTS,

Chef de Travaux, L.A.S.L.A.

Université de Liège

DE L'UTILISATION DES INDEX ET DES RELEVÉS GRAMMATICaux ET SYNTAXIQUES

Le but du présent article est de donner une description détaillée des ouvrages publiés par C.R.I.D.E.L.A., relatifs aux auteurs latins étudiés dans les Humanités. Cette description montrera l'usage qu'on peut faire de ces ouvrages pour rationaliser l'étude de la langue, rationalisation indispensable à une époque qui exige d'épargner au maximum le temps des élèves et d'éviter de leur part des efforts de mémoire inutiles.

L'étude sera divisée en deux parties. La première concernera la morphologie et la syntaxe. La seconde sera consacrée à la lexicologie. L'exposé reposera sur l'examen du dernier ouvrage qui vient d'être publié et qui concerne des extraits des oeuvres philosophiques de Cicéron : il s'agit du livre III du *De natura deorum* et du livre V des *Tusculanes* jusqu'au paragraphe 72. Etant construit exactement sur le même plan que les publications précédentes, ce volume nous servira d'exemple.

I. LA MORPHOLOGIE ET LA SYNTAXE

En gros, nos ouvrages contiennent un Index Verborum, une liste de fréquence et divers relevés statistiques.

1° L'INDEX VERBORUM

Exemples.

1. Un substantif : *DEVS* (v. annexe 1). Le lemme écrit en capitales est précédé d'un chiffre, 160, qui représente sa fréquence d'apparition dans l'oeuvre. Il est suivi des diverses formes attestées en caractères minuscules. Ces formes sont classées dans l'ordre grammatical traditionnel et accompagnées de leur référence.

Dans l'ordre, on trouvera *DEUS, DEUM*. Le génitif singulier n'est pas attesté. Viennent ensuite deux formes homographes *DEO*, datif et ablatif singulier. Pour les distinguer l'une de l'autre, la forme a été réimprimée. Ensuite vient le pluriel, *DEI*, la variante *DI*, la forme *DI* au vocatif distinguée du nominatif de la même manière que *DEO* datif et ablatif, par une réimpression de la forme.

La référence se compose de 4 chiffres. Le premier représente le livre et permet ici d'identifier en même temps l'oeuvre dont il s'agit : 3 pour le *de natura deorum*, 5 pour les *Tusculanes*. Le second chiffre désigne le paragraphe, le 3e, le numéro de la ligne dans le paragraphe, le 4e le numéro du mot dans la ligne.

2. Un verbe : *DICO*, *-ERE* (v. annexe 2). On constatera d'abord que *DICO* est accompagné d'un indice de lemme. Cet indice permet de distinguer les homographes. Dans le cas présent, *DICO 1*, non attesté, se référerait à *dico*, *-dicare*; *DICO 2* se réfère à *dico*, *dicere*. Les différentes occurrences sont classées selon 4 critères hiérarchisés :

- 1- la voix : toutes les formes actives, puis le passif;
- 2- le mode, dans l'ordre suivant : indicatif, impératif, subjonctif, participe, adjectif-verbal, gérondif, infinitif, supin en *-um*, supin en *-u*;
- 3- le temps dans l'ordre traditionnel (présent, imparfait, futur simple, etc.);
- 4- et enfin, la personne.

3. Les mots invariables.

Quand la fréquence d'un mot invariable est supérieure à 100, nous n'avons pas jugé utile d'alourdir l'ouvrage en imprimant les références. Ainsi, pour *ET* conjonction, on trouvera uniquement la fréquence (410) et le lemme (annexe 3).

Si l'on désirait étudier l'emploi de cette conjonction, des listes de références pourraient être fournies. En effet, l'étude des mots-outils peut conduire à caractériser ou à différencier un style, aussi bien que l'examen des mots significatifs. Des travaux de ce genre ont été réalisés pour le grec, par exemple l'ouvrage de Lutoslawski sur les particules dans l'œuvre de Platon.

2°. L'index verborum est suivi d'une liste des mots rangés en ordre de fré-

quence décroissante (v. annexe 4). L'intérêt d'une telle liste est évident. Elle fournit des informations précieuses sur le vocabulaire de l'auteur et permet de nombreuses exploitations, dont je parlerai à propos du vocabulaire de base et des microlexiques d'auteurs.

3° Viennent ensuite divers relevés statistiques concernant la morphologie et la syntaxe. Le premier tableau (v. annexe 5) donne la distribution des occurrences et du vocabulaire, en valeur absolue et en % selon les diverses catégories grammaticales. Ce tableau prête déjà à bien des observations. Ainsi on constatera que chez Cicéron, les deux catégories les plus importantes sont le substantif (23,99 % des occurrences) et le verbe (22,16 %) : cet exemple est représentatif de la langue latine et l'on peut ajouter de toutes langues; cela correspond parfaitement à la fonction de chaque catégorie grammaticale. C'est surtout le substantif qui est porteur d'information, c'est par lui que s'expriment les concepts : il est donc normal qu'il soit toujours le plus fréquent.

Les 5 tableaux suivants concernent le groupe nominal, c'est-à-dire le substantif, l'adjectif et l'adjectif-pronom. Trois tableaux sont consacrés au substantif. Le premier donne la répartition des occurrences et du vocabulaire selon les déclinaisons (v. annexe 6); les deux autres, la répartition des occurrences selon les déclinaisons, les cas et les nombres, en valeur absolue (v. annexe 7) et en pourcentage (v. annexe 8). Le tableau consacré aux adjectifs donne la distribution des occurrences et du vocabulaire selon les classes et les degrés de comparaison. Les adjectifs-pronoms sont classés selon leur nature.

La répartition des occurrences telle qu'elle apparaît dans chacun de ces tableaux impose, pour ainsi dire, un ordre d'apprentissage des notions morphologiques.

Il faudrait pouvoir s'arrêter à chaque tableau; chacun d'entre eux, en effet, mériterait un commentaire.

Un simple exemple cependant suffira. Le premier tableau relatif aux substantifs (annexe 6) montre que c'est la troisième déclinaison qui est la plus importante (37,21 %) et le 3^e tableau (annexe 8) montre que le cas le plus fréquent dans la langue latine est l'accusatif (27,80), suivi de près par l'ablatif (27,51) et que le singulier (71,82 %) l'emporte de loin sur le pluriel (28,17 %).

Dès lors, il serait logique de commencer l'apprentissage de la morphologie par la troisième déclinaison et dans cette déclinaison, par l'accusatif singulier.

Ceci présente un grand avantage. L'accusatif singulier permet de mettre en relief le radical et la désinence : ainsi, pour prendre un exemple simple, *de regem*, on tire *reg-* qui permet d'une part d'assimiler toute la déclinaison avec un effort de mémorisation moindre et une automatisation plus rapide et qui permet, d'autre part, à l'aide de quelques règles phonétiques élémentaires, de reconstituer le nominatif singulier.

Au surplus, il faut noter que l'accusatif singulier est toujours en *-m* quelle que soit la déclinaison (sauf, bien entendu, pour certains neutres). Ceci facilitera, par analogie, le passage aux autres déclinaisons. C'est ici que l'automatisme prend toute sa valeur.

Tous les autres tableaux concernent le verbe :

tableaux de distribution des occurrences et du vocabulaire selon les conjugaisons et les voix (v. annexe 9);

tableaux de distribution selon les modes, les temps, les voix;
tableaux de distribution selon les modes, les temps, les voix, les conjugaisons.

Ici aussi, un examen attentif des données permettra de rationaliser l'enseignement.

La seconde partie des relevés concerne proprement la syntaxe.

- 1° Le premier tableau concerne la distribution des occurrences des verbes selon les fonctions et les modes (v. annexe 10). Ce tableau se prête à bien des considérations. Pour ne donner qu'un exemple, on constatera que Cicéron, comme tous les auteurs latins, emploie plus de subordonnées (48,54 %) que de principales (30,58 %), ce qui correspond à une expression fortement hiérarchisée de la pensée.
- 2° Vient ensuite la liste des verbes subordonnés classés selon le mode : indicatif, subjonctif, participe et infinitif. A l'intérieur de chaque mode, les verbes sont rangés sous la rubrique du mot subordonnant, classés dans l'ordre des temps, pour chaque temps dans l'ordre du texte. Chaque forme est accompagnée de sa référence. Par exemple, on trouve sous la rubrique *CVM*, d'abord l'indicatif présent, puis l'imparfait, le futur etc. (v. annexe 11).

Cette liste est suivie de deux séries de tableaux.

- 1- d'un tableau qui en constitue la synthèse pour les modes personnels dans

1- d'un ordre alphabétique des codes de subordination donc, des mots subordonnants;

2- d'un tableau qui synthétise l'emploi des modes impersonnels;

3- de 2 tableaux qui concernent tous les verbes subordonnés, qu'ils soient à un mode personnel ou à un mode impersonnel.

Parmi toutes les possibilités de synthèses qu'offrent les différentes données contenues dans ces ouvrages, j'ai choisi, à titre d'exemple, d'étudier la répartition des différents types de subordination. Voici comment j'ai procédé.

J'ai rassemblé en une liste unique les chiffres relatifs à César, *Bellum Gallicum*, à Salluste, Catilina et Jugurtha, à des extraits de Tite-Live et de Tacite, à des extraits des *Métamorphoses* d'Ovide et à Virgile, *Enéide*, *Bucoliques* et *Géorgiques* (v. annexe 12).

Ici, les différents types de subordinations sont rangés en ordre de fréquence décroissante de leur apparition dans le corpus constitué par tous ces auteurs. Chaque total général est suivi du pourcentage par rapport à l'ensemble des verbes subordonnés et des pourcentages cumulés. On voit que la proposition relative représente 3223 occurrences des verbes subordonnés, ce qui équivaut à 21,72 % de toutes les subordinations et que la proposition infinitive représente 3164 occurrences, ce qui équivaut à 21,32 %. Ces deux types de subordination constituent un pourcentage cumulé de 43,04 % du total des subordinations, ce qui est déjà étonnant. Ce qui est plus extraordinaire encore, c'est

qu'avec 22 subordinations, on couvre 92 % de tous les verbes subordonnés et que les 8 % restant sont répartis entre 90 subordonnants différents.

Cette liste est illustrée par un graphique (v. annexe 13) qui fait ressortir, beaucoup plus nettement encore, les faits observés. Ce graphique porte en abscisse les différents types de subordination, et, en ordonnée, les % cumulés. La courbe est parlante. Elle montre qu'avec 4 subordonnants seulement, le relatif, la proposition infinitive, l'ablatif absolu et VT conjonction, on couvre déjà 62 % des verbes subordonnés.

Au surplus, l'allure de la courbe, dont la montée est très raide, fait apparaître qu'à partir d'un taux de compréhension de 90 %, on n'a guère d'intérêt à mémoriser de nouveaux subordonnants.

On peut tirer de cet examen des subordinations, deux conclusions.

1. L'abondance des propositions infinitives et de l'ablatif absolu exige une connaissance parfaite de l'infinitif, du participe présent et du participe parfait passif.
2. Il est évident que les 15 subordonnants les plus fréquents couvrent pratiquement toute la syntaxe verbale, ce qui permet de synthétiser les mécanismes fondamentaux de la subordination.

Ainsi, par exemple, après avoir étudié la seule proposition relative introduite par *QVI*, l'assimilation de tous les autres types de relatives ne sera plus qu'une question de vocabulaire. Il en va de même de l'interrogation indirecte dont

l'essentiel tient en un seul subordonnant, *QVIS*, ce qui permet d'utiliser, pour la compréhension de ce type de construction, l'énergie et le temps qui étaient exigés par la mémorisation d'une série impressionnante de subordonnants variés.

Est-il nécessaire d'ajouter que, pour toutes les subordinations, l'extrême richesse des listes des verbes subordonnés aidera le professeur dans sa tâche, puisqu'il pourra choisir, parmi d'innombrables exemples, ceux qui illustreront le mieux la règle grammaticale étudiée.

Ce seul exemple de synthèse des données statistiques recueillies remet en question la façon de procéder des grammaires traditionnelles et met en évidence la nécessité de constituer une grammaire fréquentielle où les phénomènes seront hiérarchisés en fonction de leur fréquence d'apparition dans la langue, grammaire qui répondra, dès lors, beaucoup mieux, aux exigences de notre époque. D'ailleurs, M. l'Inspecteur Lecompe, conscient de ce problème depuis de longues années, a proposé récemment à M. l'Inspecteur général et à l'Organisation des Etudes de poursuivre nos recherches pour construire une telle grammaire.

II. LA LEXICOLOGIE

Considérant l'énorme documentation que constituent nos ouvrages, nous avons

cru nécessaire d'en faire une synthèse au point de vue du vocabulaire:

Nous avons donc publié en 1972, un Vocabulaire de Base fondé sur un corpus de 323.085 occurrences regroupant les oeuvres suivantes :

- un choix des Fables de Phèdre
- le *Bellum Gallicum* de César
- des extraits des Métamorphoses d'Ovide
- le *De Coniuratione Catilinae* et le *Bellum Jugurthinum* de Salluste
- les Bucoliques, les Géorgiques et les six premiers chants de l'Enéide de Virgile
- des extraits de Tite-Live
- des extraits des oeuvres philosophiques de Cicéron
- les *Carmina* de Catulle
- les Odes d'Horace
- les Lettres à Lucilius de Sénèque
- des extraits des Histoires et des Annales de Tacite

Un vocabulaire de base valable pour tous les écrivains d'une même langue est utopique. En effet, si les mots-outils, c'est-à-dire les adjectifs-pronoms, les prépositions, les conjonctions et les mots-thèmes comme *RES*, *FACIO*, *POSSVM*, *FERO*, sont stables à travers toute la littérature, il n'en va pas de même des mots de caractérisation. Ceux-ci sont en dépendance absolue avec le sujet traité. Sans doute, ce sont les mots les plus rares, mais, comme ils sont, par définition, porteurs de l'information, il est difficile de comprendre un texte sans les connaître. Il est donc nécessaire d'adopter la méthode préconisée par les théoriciens de la Traduction automatique. Elle consiste à créer, à côté d'un

vocabulaire général relativement restreint et consacré aux mots les plus fréquents, une série de vocabulaires particuliers ou microlexiques où apparaîtront les mots qui caractérisent un auteur.

Ces microlexiques offrent un double intérêt :

- a) ils permettent d'épurer le vocabulaire de base de tout vocable dont l'aire d'utilisation est trop limitée;
- b) ainsi que je le montrerai tout à l'heure, ils préparent à merveille la lecture d'une oeuvre particulière.

Quelles techniques avons-nous utilisées et sur quels critères nous sommes-nous fondés pour constituer ces divers relevés ?

Pour les différents auteurs, nous disposons, sur bande magnétique, d'un fichier-fréquence où chaque lemme est enregistré avec sa fréquence d'apparition dans le texte et son analyse morphologique.

C'est ce fichier qui permet d'établir la liste de fréquence en ordre décroissant dont un échantillon se trouve en annexe 4.

Il permet aussi d'établir le tableau de distribution du vocabulaire d'un auteur donné, tableau où les différents mots de même fréquence sont regroupés. Ce tableau indique, dans l'ordre décroissant, l'effectif de chaque fréquence, c'est-à-dire le nombre de mots ayant la fréquence 1, 2, 3 etc. et le nombre d'occurrences correspondant à chacun de ces effectifs.

Prenons un exemple théorique simple. Supposons que, sur un total de 10.000 occurrences, il y ait chez un auteur 1 mot, *ET* conjonction, attesté 1000

fois, 2 mots, *VT* et *QVI*, attestés 500 fois, 3 mots, *RES*, *POSSVM* et *FACIO*, attestés 200 fois.

Le tableau de distribution se présentera de la manière suivante :

fréquences	effectifs	nombre d'occurrences
1000	1	1000
500	2	1000
200	3	600

On voit immédiatement l'intérêt que présente un tel tableau. En effet, si l'on cumule les effectifs d'une part, le nombre d'occurrences, d'autre part, en valeur absolue et en pourcentage, la manière dont s'accroît le vocabulaire d'un auteur apparaît clairement.

fréquences	effectifs	effectifs cumulés	nombre d'occurrences	%	% cumulés
1000	1	1	1000	10	-
500	2	3	1000	10	20
200	3	6	600	6	26

Nous avons établi, pour chaque auteur, un tableau de ce type (v. annexe 14). En voici la description.

La première colonne numérote simplement les lignes.

La seconde colonne donne la fréquence d'emploi des lemmes de l'oeuvre.

La troisième colonne indique le nombre de mots affectés de la même fréquence, tandis que la quatrième donne le nombre d'occurrences par fréquence.

La cinquième colonne cumule les lemmes employés par l'auteur, de sorte que la dernière ligne donne le nombre total de mots différents de l'oeuvre.

La sixième colonne cumule les occurrences : ainsi, enfin de cette colonne, lit-on le nombre total des mots-formes de l'oeuvre, déduction faite des noms propres.

La septième colonne donne le pourcentage de la totalité de l'oeuvre que représentent les occurrences de chaque fréquence ou, si l'on veut, le pourcentage de compréhension de l'oeuvre que représente l'effectif de la colonne 4.

La huitième colonne cumule ces pourcentages.

La neuvième colonne calcule le pourcentage de compréhension de l'oeuvre que l'on obtient pour chaque lemme et la dixième colonne cumule ces pourcentages.

Un exemple concret mettra mieux les choses en lumière.

La 27e ligne du tableau relatif à César montre, en colonne 3 et 2, qu'il y a deux lemmes, dont chacun est employé 244 fois, ce qui fait 488 occurrences pour cette fréquence. La colonne 5 montre que l'auteur a employé jusqu'ici

28 lemmes différents qui représentent (colonne 6) 14.675 occurrences. La colonne 7 fait apparaître que pour cette fréquence seule, 488 occurrences représentent une compréhension de 1,13 % de l'oeuvre. La colonne 8 montre que les occurrences employées jusqu'ici représentent 34,26 % de l'oeuvre entière. Les colonnes 9 et 10 présentent les pourcentages, non plus pour les occurrences, mais pour les lemmes : les deux lemmes de cette ligne représentent 0,07 % de l'oeuvre et les lemmes employés jusqu'ici représentent 1 %.

Les données essentielles, c'est-à-dire les effectifs cumulés et le pourcentage de texte couvert par chacun de ces effectifs, sont contenues dans les colonnes 5 et 8.

Ce sont ces données qui permettent de déterminer le nombre de mots nécessaires pour obtenir le taux de compréhension que l'on souhaite atteindre.

Naturellement, la consultation de tels tableaux n'est pas aisée. Mais il existe une technique qui permet de synthétiser et de visualiser les données : ce sont les graphiques. C'est pourquoi nous avons établi, pour chaque distribution du vocabulaire, la courbe correspondante construite de la manière suivante :

- en abscisse, nous portons les effectifs cumulés des différentes fréquences des lemmes, c'est-à-dire le contenu de la colonne 5;
- en ordonnée, figure le pourcentage de texte couvert par chacun de ces effectifs, c'est-à-dire le contenu de la colonne 8.

On trouvera, en annexe 15, la courbe relative au *De Bello gallico* de César. Cette courbe est représentative, grosso modo, de la manière dont s'accroît le vocabulaire d'un auteur.

D'une manière générale, on constate que les courbes, qui, au début, sont en pente raide, s'aplatissent de plus en plus et tendent à l'horizontale. En clair, ceci signifie qu'un petit nombre de mots de haute fréquence couvre un pourcentage important du texte et qu'il faut un nombre de plus en plus important de mots pour couvrir un pourcentage de texte de plus en plus faible.

Pour rendre les choses plus claires encore, nous avons regroupé sur un même graphique, les courbes relatives à César, Salluste, Tite-Live, Tacite, Ovide, Virgile, Enéide, Horace, Odes (v. annexe 16).

Ce graphique fait apparaître très nettement la différence entre la prose et la langue poétique. La pente de la courbe, plus faible dans ce dernier cas, montre que le poète emploie une langue plus riche, plus diversifiée, comportant moins de mots-thèmes, plus de mots de caractérisation. A partir de là, on atteint aux structures mêmes de chaque type de langue. A cet égard, la courbe de Tacite est très révélatrice. Elle montre que la langue de cet auteur est beaucoup plus proche du langage poétique que de la prose des historiens.

Ce sont les courbes de distribution qui nous ont permis de déterminer de manière empirique le niveau de fréquence auquel il faut s'arrêter pour atteindre un taux de compréhension efficace.

L'allure des courbes montre nettement qu'au-delà de 85 %, le gain de com-

préhension est très faible comparativement à l'effort de mémorisation qu'il exigerait.

Nous appelons fréquence-limite le niveau de fréquence correspondant approximativement à ce taux de compréhension. Naturellement, cette fréquence-limite varie d'auteur à auteur. C'est elle qui a servi de critère pour établir d'une part le lexique de base, d'autre part les microlexiques.

1. LE LEXIQUE DE BASE

Après avoir compilé les fichiers-fréquence des différents auteurs, nous avons établi une liste de fréquence unique où chaque mot est suivi de sa fréquence chez chaque auteur, ainsi que de sa fréquence dans l'ensemble des textes étudiés.

De cette liste, nous n'avons retenu que les mots répondant aux deux conditions suivantes :

- 1.- Ces mots doivent être présents chez tous les auteurs étudiés.*
- 2.- Ces mots doivent y apparaître avec au moins la fréquence-limite.

La première condition répond à un critère de répartition homogène, la seconde à un critère d'économie.

* Hormis Phèdre et les Bucoliques, en raison de la brièveté de ces textes.

On trouvera en annexe 17 trois échantillons de ce lexique. Dans le premier, les mots sont classés en ordre de fréquence décroissante, dans le second dans l'ordre alphabétique des lemmes, dans le troisième ils sont rangés par catégories grammaticales. Le lexique de base ainsi constitué ne comprend que 144 mots. Cependant, la proportion de texte qu'il permet de couvrir est étonnante puisqu'elle se situe aux environs de 50 % pour la prose et de 36 % pour les oeuvres poétiques. Cette proportion relativement élevée est due au fait que ce lexique est constitué, en majeure partie, de mots-outils (*ET, QVE, QVI*) et de mots-thèmes (*MAGNVS, RES, FACIO, POSSVM etc.*).

II. LES MICROLEXIQUES

Ceux-ci ont été traités d'une manière différente selon qu'ils concernent les auteurs du cycle d'apprentissage de la langue ou les auteurs du cycle supérieur. L'idée qui nous a guidés dans l'élaboration du premier groupe est qu'un apprentissage progressif doit trouver son reflet dans le vocabulaire mémorisé.

Par conséquent, on ne retrouvera plus dans le microlexique de César les mots qui figurent dans le lexique de base; le microlexique de Phèdre ne contiendra plus les mots faisant partie du vocabulaire de César et ainsi de suite.

Nous avons condensé dans un tableau les taux de compréhension atteints pour les auteurs du premier cycle, en tenant compte de la progression chronologique de la mémorisation (v. annexe 18).

Les microlexiques ont une autre utilité. Comme les mots y sont repris en

ordre de fréquence décroissante, ces listes font apparaître avec une clarté étonnante, non seulement les mots porteur de *realia* qu'il conviendra d'expliquer, d'expliciter, pour introduire le cours proprement dit, mais encore les préoccupations conscientes et même inconscientes de l'auteur.

Prenons par exemple les mots les plus fréquents du microlexique de César (v. annexe 19). Le mot de caractérisation qui vient en tête est *HOSTIS*. On ne jugera cette fréquence naturelle que si l'on n'a une vue superficielle des choses. *hostis* permet bien des commentaires. D'abord *hostis* signifie l' "étranger". Le sens d'ennemi est dérivé directement de ce sens. Curieuse dérivation, et symptomatique de la mentalité d'un peuple conquérant, pour qui l'étranger est l'ennemi. Ensuite, l'emploi de *hostis* représente un véritable gauchissement de la vérité, car les Gaulois, pas plus que les Belges en 1940 par rapport aux Allemands, n'étaient les ennemis de Rome. Mais il était capital pour César de parler d'ennemis car cela constituait une justification de sa politique de conquête. Les mots suivants sur la liste sont des termes militaires : *OPPIDVM*, *LEGIO*, *MILES*, *EXERCITVS*, *CASTRA*, *LEGATVS*, *EQVES*. Illustrés par des diapositives, ils permettront, avant même l'explication proprement dite, de se faire une idée de l'organisation militaire du peuple romain, du déroulement de la campagne des Gaules et des difficultés de l'entreprise. Bref, ils apporteront un éclairage précieux pour la lecture du texte.

On peut en dire autant du microlexique de chaque auteur. Prenons, par exemple, Salluste, *Catilina*. Ici viennent en tête des mots relatifs à la vie politique de la cité, *CONSVL*, *CIVIS*, *DECERNO*, *CONSCRIBO* (dans l'expression *PATRES CONSCRIPTI*), *PRAETOR*, *PLEBS*, *PATRIA*, *CONSVLATVS*

etc., autant de *realia* se prêtant à une explication technique approfondie et intéressante au point de vue institutionnel.

D'autre part, des mots tels que *FACINVS*, *INGENIVM*, *SVPERBIA*, *INVIDIA*, *LVXVRIA*, mettent en relief la psychologie de Catilina, telle que Salluste la concevait. Sur ce point, une comparaison avec le microlexique des Catilinaires de Cicéron serait du plus grand intérêt.

Ainsi, à un double point de vue, la mémorisation et l'intelligence du texte, les microlexiques apportent au professeur une aide précieuse.

L'exposé qui précède montre combien les quelques 20 volumes publiés jusqu'ici constituent une mine inépuisable d'informations, tant au point de vue de la morphologie, de la syntaxe et du vocabulaire qu'au point de vue de l'explication historique, institutionnelle et stylistique des textes.

Tel est le résultat de la collaboration entre l'enseignement secondaire et l'Université, collaboration souhaitée depuis tant d'années par les responsables de l'enseignement.

Dans la mesure où les documents seront judicieusement utilisés par les professeurs, on aboutira à ce renouvellement des études exigé par l'économie de moyens et de temps que réclame notre époque.

S. GOVAERTS

dissignata (esse)	3,49, 2, 3	3,79,11, 5
3,85, 6, 2	3,76,10, 4	3,86, 1, 4
1 DESINO	3,91,15, 7	3,90, 2, 2
desierit	3,92, 4, 1	di
3,41, 9, 1	deum	5,19, 9, 4
2 DESPICIO	3,25, 3, 2	deos
despiciens	3,38, 1, 3	3, 3, 5, 7
5, 4, 6, 7	3,38,15, 6	3, 3, 7, 3
despicere	3,39, 2, 1	3, 6, 8, 4
3,39, 3, 4	3,41, 6, 4	3, 7, 4, 3
1 DESTINO	3,43,12, 6	3, 9,12, 2
destinatam	3,44,10, 5	3,11, 3, 7
5,63, 5, 5	3,49, 7, 6	3,11, 9, 7
7 DESVM 1	3,50,12, 3	3,15, 5, 3
deest	3,51, 3, 2	3,15,15, 7
5,39, 4, 11	3,76, 5, 8	3,17, 2, 2
5,40, 2, 2	deo	3,18,13, 3
defuit	3,24, 2, 3	3,19, 3, 2
3,68, 1, 8	3,38, 2, 8	3,38,10, 3
desit	3,38, 8, 5	3,39, 7, 6
5,23, 4, 9	3,61,12, 8	3,40, 6, 5
5,53, 6, 3	3,86,11, 4	3,40, 8, 6
deesse	3,90, 4, 1	3,41, 7, 5
5,23,11, 3	deo	3,44, 4, 3
5,42,17, 6	3,20, 7, 1	3,45, 2, 5
1 DETESTABILIS	3,24, 9, 2	3,47,13, 2
detestabilius	3,71, 6, 4	3,51,13, 8
5,64, 2, 2	3,87, 3, 9	3,53, 2, 3
1 DETORQVEO	3,88, 8, 2	3,64,10, 4
detorquebat	3,91, 5, 5	3,83, 3, 1
5,59, 7, 6	3,91, 7, 7	3,84,13, 8
3 DETRAHO	5,38,15, 11	3,89, 5, 5
detraxit	5,70, 2, 5	3,93, 5, 6
3,83,12, 3	dei	3,93,13, 6
detraxerat	3,26, 3, 5	3,93,15, 2
5,58, 7, 6	3,45, 5, 7	deorum
detracto	3,47, 1, 4	3, 5,11, 8
5,45,13, 3	3,51,16, 3	3, 5,27, 1
1 DEVERSOR 2	di	3,16, 3, 4
deuersarer	3,10, 3, 5	3,17, 7, 1
5,22, 3, 7	3,17, 2, 5	3,23,20, 3
1 DEVOCO	3,20, 4, 7	3,26,12, 6
deuocauit	3,43, 7, 6	3,28, 7, 1
5,10,13, 2	3,43,11, 4	3,44, 7, 7
1 DEVOTIO	3,43,15, 6	3,45, 8, 8
deuotionibus	3,44,11, 7	3,47,15, 6
3,15, 5, 1	3,45, 8, 4	3,48, 4, 8
160 DEVS	3,45,13, 10	3,49, 3, 7
deus	3,45,15, 5	3,50, 4, 6
3,20, 8, 3	3,47, 5, 10	3,50, 7, 1
3,23,16, 2	3,47,10, 6	3,50,10, 5
3,23,19, 4	3,48,11, 5	3,51, 6, 2
3,25,12, 6	3,49, 6, 2	3,51, 7, 7
3,25,13, 1	3,63, 7, 9	3,61, 4, 8
3,38,14, 2	3,64,10, 8	3,61, 9, 6
3,40, 3, 4	3,65, 2, 7	3,63, 8, 7
3,43,13, 5	3,75, 3, 5	3,63,10, 3
3,44, 9, 7	3,75, 5, 4	3,65, 1, 6
3,45,11, 6	3,75,10, 5	3,67, 1, 6
3,45,13, 4	3,77, 3, 1	3,71, 9, 4

ANNEXE 1

3,73,14, 3	3,47,10, 4	3,18, 9, 1
3,76, 2, 2	3,64, 5, 4	3,57, 9, 8
3,76, 9, 1	dicit	3,83, 4, 7
3,82,10, 6	3,11,14, 5	3,84, 9, 8
3,84, 3, 4	3,43, 5, 3	5,12, 6, 6
3,88,13, 7	3,72, 9, 1	5,52, 8, 5
3,90, 5, 7	5,24,10, 7	5,65, 6, 4
3,90, 8, 4	5,24,17, 2	5,67, 7, 5
3,92, 7, 1	5,24,17, 6	dixti
3,93,17, 1	5,26, 2, 5	3,23, 9, 2
3,94, 3, 7	5,26,11, 7	diximus
3,94, 7, 7	dicimus	3,16, 9, 1
3,95, 7, 2	3,24,11, 3	3,20, 3, 1
5,48, 1, 3	3,41, 4, 2	3,59, 9, 1
5,70, 5, 7	5,28, 8, 8	3,59,10, 8
dis	5,29, 2, 6	5,33, 6, 7
3,87, 6, 7	5,33, 5, 3	dic
3,87, 8, 8	5,55,16, 3	3, 8, 1, 1
3,89, 3, 3	dicitis	3,43, 6, 6
5,31, 5, 3	3,35, 5, 2	dicam
dis	3,38,11, 3	3,82, 9, 2
3, 3, 4, 7	3,44, 3, 1	5,12,19, 6
3, 5, 8, 4	3,66, 9, 4	5,28, 6, 7
3, 7, 9, 4	3,78, 9, 3	dicas
3,11, 6, 4	3,79, 6, 6	3,54, 7, 8
3,16,11, 8	3,86, 5, 2	dicat
3,43, 2, 7	3,90, 5, 4	5,26, 3, 4
3,44, 5, 7	3,92, 8, 1	5,28, 3, 8
3,64, 2, 7	3,93, 5, 3	5,31, 9, 3
3,69, 8, 6	3,93, 9, 2	dicamus
3,70, 5, 5	3,94, 9, 3	3,53, 1, 1
3,70, 8, 4	dicunt	3,64, 3, 3
3,75, 1, 5	3,13, 6, 2	3,94, 6, 4
3,78, 2, 4	3,53, 3, 7	dicatis
3,79, 6, 4	3,58, 9, 4	3,11, 9, 2
3,83, 7, 8	5,23, 6, 6	3,14, 6, 5
3,86,10, 5	5,44, 2, 4	dicant
3,88,11, 7	5,47, 2, 4	5,23, 8, 1
3,93, 6, 3	5,47, 2, 8	5,30, 4, 8
2 DIAGORAS N	dicebas	dicerem
Diagoras	3,15, 2, 1	3, 1, 3, 7
3,89, 3, 6	3,16, 2, 3	diceres
Diagoram	3,18,11, 5	3, 8, 6, 4
3,89,15, 8	3,20, 6, 5	diceret
3 DIANA N	3,28, 4, 1	3,35, 4, 2
Dianam	3,52, 2, 1	3,83,17, 2
3,51, 4, 4	3,52, 4, 2	5,18,13, 3
Diana	dicebat	5,26, 5, 3
3,60, 1, 5	3,84, 4, 2	5,27, 4, 9
Dianae	3,84, 7, 3	5,31, 7, 1
3,58, 1, 1	3,88,13, 4	dicerent
183 DICO 2	dicam	3,89,12, 4
dico	3,91,11, 3	dixerim
5,12,19, 8	dices	3,91, 2, 5
dicis	3,45, 2, 7	5,12, 4, 3
3,12, 6, 2	dicemus	5,33, 2, 3
3,15,13, 6	5,23,11, 4	dixeris
3,21, 2, 8	dixi	3, 8, 5, 1
3,21, 4, 9	3, 4, 3, 9	5,16, 9, 2
3,41, 7, 3	3,16, 4, 7	dixerit

ANNEXE 2

3,11,20, 3	errori	3,43, 8, 5
equo 3,15,11, 1	3,76, 8, 8	3,43,13, 1
equos 3,47,13, 7	errore 3,25,13, 9	3,45, 5, 2
1 EREBOS N 5,11, 7, 2	5,43, 4, 5	3,45, 9, 5
Erebo 5,55,15, 6	erroribus 3,51, 7, 4	3,51, 5, 4
3,44,16, 6	5,39, 3, 1	3,51, 6, 5
2 ERECHTHEVS N 3,51,12, 6	3,52, 3, 6	3,51,12, 6
Erechtheus 3,53, 1, 7	3,52, 3, 6	3,52, 3, 6
3,49, 6, 5	3,53, 1, 7	3,53, 1, 7
3,50, 6, 2	3,53, 7, 3	3,53, 7, 3
1 ERECTVS 3,53,11, 4	3,60, 7, 8	3,53,11, 4
erectus 3,61, 8, 4	3,61, 8, 4	3,60, 7, 8
5,42, 8, 2	3,62, 5, 5	3,61, 8, 4
28 ERGO 2 3,63, 9, 7	3,63, 9, 7	3,62, 5, 5
3,11,18, 3	3,63,10, 6	3,63, 9, 7
3,29, 9, 6	3,70, 9, 4	3,63,10, 6
3,33,12, 3	3,74,12, 2	3,70, 9, 4
3,43,11, 3	3,75, 5, 3	3,74,12, 2
3,43,12, 8	3,80,14, 3	3,75, 5, 3
3,44, 3, 2	3,83,14, 5	3,80,14, 3
3,45, 5, 5	3,84, 1, 1	3,83,14, 5
3,47, 7, 3	3,89,15, 4	3,84, 1, 1
3,48, 6, 6	3,93, 3, 4	3,89,15, 4
3,50,13, 2	3,93, 8, 6	3,93, 3, 4
3,51, 5, 3	3,93,10, 4	3,93, 8, 6
3,51,15, 5	3,95, 5, 1	3,93,10, 4
3,52, 4, 3	5, 6, 3, 9	3,95, 5, 1
3,52, 8, 2	5, 7, 4, 3	5, 6, 3, 9
3,52,10, 3	5, 7,10, 4	5, 7, 4, 3
3,70, 6, 3	5,10, 2, 3	5, 7,10, 4
3,73,12, 2	5,10,14, 2	5,10, 2, 3
3,93,10, 8	5,12,17, 4	5,10,14, 2
5,16,11, 6	5,13, 2, 1	5,12,17, 4
5,35, 8, 3	5,14, 8, 5	5,13, 2, 1
5,40, 7, 6	5,16, 1, 9	5,14, 8, 5
5,45,17, 2	5,22, 7, 7	5,16, 1, 9
5,47,12, 7	5,24,14, 7	5,22, 7, 7
5,51, 6, 2	5,26,13, 1	5,24,14, 7
5,52, 4, 7	5,27, 1, 4	5,26,13, 1
5,53, 6, 6	5,29,10, 1	5,27, 1, 4
5,53, 8, 7	5,31, 5, 1	5,29,10, 1
5,67, 5, 6	5,37, 3, 4	5,31, 5, 1
1 ERRATVM 5,38, 1, 3	5,39, 9, 5	5,37, 3, 4
erratis 5,42,12, 6	5,43,10, 7	5,38, 1, 3
5,62,15, 1	5,43,12, 10	5,39, 9, 5
2 ERRO 2 5,43,10, 7	5,43,14, 4	5,42,12, 6
errantes 5,45, 3, 6	5,45,17, 3	5,43,10, 7
3,51, 5, 8	5,49, 1, 4	5,43,12, 10
errare 3,51, 5, 8	5,51, 6, 7	5,43,14, 4
5,36, 5, 2	5,51, 9, 9	5,45, 3, 6
11 ERROR 5,52, 9, 1	5,53, 2, 5	5,45,17, 3
error 3,63, 9, 3	5,53, 4, 1	5,49, 1, 4
3,64, 2, 3	5,53, 8, 8	5,51, 6, 7
5, 6, 8, 2	5,55,14, 2	5,51, 9, 9
errorem 5, 6, 8, 2		5,52, 9, 1
5, 4,10, 5		5,53, 2, 5
5, 8, 4, 6		5,53, 4, 1
		5,53, 8, 8
		5,55,14, 2

ANNEXE 3

SVM 1	646	NOS	49	FORTVNA	24
QVI 1	573	CVM 2	48	IMMORTALIS	24
ET 2	410	VOLO 3	48	PHILOSOPHVS 1	24
IS	346	MVLTVS	45	PRIMVS	24
IN	275	TVM	44	BENE	23
NON	222	ATQVE 1	43	SAPIENTS 1	23
QVE	201	ITA	43	VIR	23
HIC 1	194	SIC	43	VNVS	23
DICO 2	183	DO	42	CAVSA	22
DEVS	160	ISTE	42	DOLOR	22
OMNIS	156	PVTO	42	NOMEN	22
QVIS 1	156	ANIMVS	40	QVAM 2	22
POSSVM 1	152	NASCOR	40	QVIA	22
SI 2	151	NAM 2	37	AN	21
AB	135	BONVM	36	BEATE	21
SVM 2	135	INQVIO	35	IAM	21
ILLE	132	QVIDAM	35	LOCVS	21
ENIM 2	130	SVVS	35	MAGIS 2	21
CVM 3	126	ALTER	34	MENS	21
EX	113	IVPPITER N	34	MODVS	21
EGO	108	VIVO	34	NEQVE	21
AVTEM	103	FACIO	33	QVISQVAM	21
VT 4	101	MAGNVS	33	SINE	21
SED	99	NE 2	33	TAM	21
TV	97	NE 3	32	VTOR	21
BEATVS	95	EFFICIO	31	AGO	20
VIDEO	93	MALVM 1	31	APPELLO 1	20
SVI 1	91	NISI	31	MISER	20
AD	90	CORPVS	30	NECESSE	20
AVT	89	QVAERO	30	NEGO	20
ETIAM	84	QVOD 2	30	NOSTER	20
IGITVR 2	81	TAMEN 2	30	SENTIO	20
IPSE	81	MVNDVS 1	29	VLLVS	20
DE	80	AT 2	28	FIO	19
NEC 2	80	ERGO 2	28	ORATIO	19
RES	80	INTELLIGO	28	VEL 2	19
VT 1	76	SEMPER	28	COTTA N	18
NIHIL	75	VIS	28	DIVINVS 2	18
RATIO	75	GENVS 1	27	PARVM 2	18
BONVS	73	MODO 1	27	TVVS	18
HABEO	72	SVPERS	27	APVD	17
IDEM	70	TERTIVS	27	CERTE	17
VITA	70	ACCIPIO	26	DISPVTO	17
VIRTVS	68	ALIQVIS	26	FILIVS	17
HOMO	65	BALBVS N	26	ITAQUE 2	17
QVIDEM	56	CVR 2	26	PARS	17
ALIVS	53	QVISQVE 2	26	QVOQVE	17
NVLLVS	52	PATER	25	TALIS	17
QVAM 1	52	ANIMAL	24	TEMPVS 1	17
NATVRA	49	FERO	24	VOS	17

ANNEXE 4

1. TABLEAU GENERAL DE DISTRIBUTION DES OCCURRENCES
ET DU VOCABULAIRE SELON LES CATEGORIES GRAMMATICALES

Catégories	Nombre de mots *		Nombre de mots (vocabulaire)		Moyenne d'emploi
	Valeurs absolues	%	Valeurs absolues	%	
<i>Substantifs</i>	3.805	23,99	1.191	43,62	3,19
<i>Adjectifs</i>	1.151	7,25	407	14,90	2,82
<i>Numéraux</i>	156	0,98	27	0,98	5,77
<i>Adjectifs-pronoms</i>	2.687	16,94	54	1,97	49,75
<i>Verbes</i>	3.515	22,16	714	26,15	4,92
<i>Adverbes</i>	1.651	10,40	259	9,48	6,37
<i>Prépositions</i>	861	5,42	23	0,84	37,43
<i>Conjonctions de coordination</i>	1.480	9,33	25	0,91	59,20
<i>Conjonctions de subordination</i>	542	3,41	26	0,95	20,84
<i>Interjections</i>	12	0,07	4	0,14	3,00
TOTAUX	15.860	99,95	2.730	100,00	5,80
<i>ESSE employé comme auxiliaire</i>	135	-	1		
<i>IRE employé comme auxiliaire</i>	0	-	0		
<i>Mots dépourvus d'analyse</i>	31*	-	0**		
	16.026	-	2.731		

* Il s'agit de EGO, dans l'expression ME... HERCVLE, de QVIDEM dans l'expression NE... QVIDEM et de la *crux* TYPANIDIS.

** Ici n'interviennent ni EGO ni QVIDEM qui ont déjà été comptés, le premier avec les adjectifs-pronoms, le second avec les adverbes.

ANNEXE 5

2. TABLEAUX DETAILLES RELATIFS AUX PRINCIPALES CATEGORIES GRAMMATICALES

a.- Distribution des substantifs et de leurs occurrences selon les déclinaisons

NOMBRE D'OCCURRENCES

	Valeurs absolues	% du nombre total d'occurrences	% des occurrences de substantifs
<i>1ère déclinaison</i>	674	4,24	17,71
<i>2e déclinaison</i>	1.233	7,77	32,40
<i>3e déclinaison</i>	1.416	8,92	37,21
<i>4e déclinaison</i>	155	0,97	4,07
<i>5e déclinaison</i>	123	0,77	3,23
<i>Anomaux</i>	26	0,16	0,68
<i>Déclinaison grecque</i>	178	1,12	4,67
TOTAUX	3.805	23,95	99,97

NOMBRE DE MOTS DIFFERENTS

	Valeurs absolues	% du vocabulaire total	% du vocabulaire des substantifs
<i>1ère déclinaison</i>	218	7,98	18,30
<i>2e déclinaison</i>	371	13,58	31,15
<i>3e déclinaison</i>	419	15,34	35,18
<i>4e déclinaison</i>	57	2,08	4,78
<i>5e déclinaison</i>	11	0,40	0,92
<i>Anomaux</i>	7	0,25	0,58
<i>Déclinaison grecque</i>	108	3,95	9,06
TOTAUX	1.191	43,62	99,99

ANNEXE 6

a bis.- Distribution des substantifs et de leurs occurrences selon les déclinaisons, les cas et les nombres.

Valeurs absolues

Cas	Nombre	1ère	2e	3e	4e	5e	Anomaux D.grecque		Totaux
Nom.	sg	120	187	300	25	16	5	74	727
	pl	35	108	70	9	2		5	229
	tot	155	295	370	34	18	5	79	956
Voc.	sg	7	21	3				1	32
	pl		2						2
	tot	7	23	3				1	34
Acc.	sg	156	189	309	30	29	7	34	754
	pl	51	127	93	21	7	1	4	304
	tot	207	316	402	51	36	8	38	1058
Gen.	sg	85	102	103	8	5		24	327
	pl	16	111	58	1	27		2	215
	tot	101	213	161	9	32		26	542
Dat.	sg	14	31	38		2	2	6	93
	pl	6	27	22	1	5			61
	tot	20	58	60	1	7	2	6	154
Abi.	sg	149	199	338	51	20	2	27	786
	pl	34	126	82	9	10			261
	tot	183	325	420	60	30	2	27	1047
Loc.	sg	1	3				3	1	8
	pl								
	tot	1	3				3	1	8
Ind.	sg						6		6
	pl								
	tot						6		6
Totaux	sg	532	732	1091	114	72	25	167	2733
	pl	142	501	325	41	51	1	11	1072
	tot	674	1233	1416	155	123	26	178	3805

ANNEXE 7

POURCENTAGES

1) Déclinaisons et cas

Cas	Déclinaisons							Totaux
	1ère	2e	3e	4e	5e	Ano- maux	Décl. grecque	
Nominatif	4,07	7,75	9,72	0,89	0,47	0,13	2,07	25,12
Vocatif	0,18	0,60	0,07	-	-	-	0,02	0,89
Accusatif	5,44	8,30	10,56	1,34	0,94	0,21	0,99	27,80
Génitif	2,65	5,59	4,23	0,23	0,84	-	0,68	14,24
Datif	0,52	1,52	1,57	0,02	0,18	0,05	0,15	4,04
Ablatif	4,80	8,54	11,03	1,57	0,78	0,05	0,70	27,51
Locatif	0,02	0,07	-	-	-	0,07	0,02	0,21
Indéclinables	-	-	-	-	-	0,15	-	0,15
TOTAUX	17,71	32,40	37,21	4,07	3,23	0,68	4,67	99,99

2) Cas et nombres

Cas	Nombre		Totaux
	Singulier	Pluriel	
Nominatif	19,10	6,01	25,12
Vocatif	0,84	0,05	0,89
Accusatif	19,81	7,98	27,80
Génitif	8,59	5,65	14,24
Datif	2,44	1,60	4,04
Ablatif	20,65	6,85	27,51
Locatif	0,21	-	0,21
Indéclinables	0,15	-	0,15
TOTAUX	71,82	28,17	99,99

ANNEXE 8

d.- Distribution des verbes et de leurs occurrences selon les conjugaisons et les voix

NOMBRE D'OCCURRENCES

Valeurs absolues

	Actif	Passif	Totaux	Déponents	Semi-déponents	Totaux généraux
<i>1e conjugaison</i>	376	180	556	66	-	622
<i>2e conjugaison</i>	297	106	403	30	20	453
<i>3e conjugaison</i>	614	289	903	112	7	1.022
<i>4e conjugaison</i>	106	16	122	21	-	143
<i>3e conjugaison mixte</i>	134	58	192	16	-	208
TOTAUX	1.527	649	2.176	245	27	2.448
<i>Verbes anomaux</i>	1.044	23	1.067	-	-	1.067
TOTAUX GENERAUX	2.571	672	3.243	245	27	3.515

% du nombre total d'occurrences

	Actif	Passif	Totaux	Déponents	Semi-déponents	Totaux généraux
<i>1e conjugaison</i>	2,37	1,13	3,50	0,41	-	3,92
<i>2e conjugaison</i>	1,87	0,66	2,54	0,18	0,12	2,85
<i>3e conjugaison</i>	3,87	1,82	5,69	0,70	0,04	6,44
<i>4e conjugaison</i>	0,66	0,10	0,76	0,13	-	0,90
<i>3e conjugaison mixte</i>	0,84	0,36	1,21	0,10	-	1,31
TOTAUX	9,62	4,09	13,72	1,54	0,17	15,43
<i>Verbes anomaux</i>	6,58	0,14	6,72	-	-	6,72
TOTAUX GENERAUX	16,21	4,23	20,44	1,54	0,17	22,16

% des occurrences de verbes

	Actif	Passif	Totaux	Déponents	Semi-déponents	Totaux généraux
<i>1e conjugaison</i>	10,69	5,12	15,81	1,87	-	17,69
<i>2e conjugaison</i>	8,44	3,01	11,46	0,85	0,56	12,88
<i>3e conjugaison</i>	17,46	8,22	25,68	3,18	0,19	29,07
<i>4e conjugaison</i>	3,01	0,45	3,47	0,59	-	4,06
<i>3e conjugaison mixte</i>	3,81	1,65	5,46	0,45	-	5,91
TOTAUX	43,44	18,46	61,90	6,97	0,76	69,64
<i>Verbes anomaux</i>	29,70	0,65	30,35	-	-	30,35
TOTAUX GENERAUX	73,14	19,11	92,26	6,97	0,76	100,00

ANNEXE 9

b. Tableau de distribution des occurrences selon les modes et les fonctions

Valeurs absolues

	Verbes principaux	Verbes de prop. subordonnées	Autres	Totaux
<i>Indicatif</i>	968	658	-	1.626
<i>Impératif</i>	22	-	-	22
<i>Subjonctif</i>	85	654	-	739
<i>Infinitif</i>	-	367	347	714
<i>Participe</i>	-	29	234	263
<i>Adjectif verbal</i>	-	-	94	94
<i>Gérondif</i>	-	-	53	53
<i>Supin I</i>	-	-	1	1
<i>Supin II</i>	-	-	3	3
TOTAUX	1.075	1.708	732	3.515

% du nombre total d'occurrences

	Verbes principaux	Verbes de prop. subordonnées	Autres	Totaux
<i>Indicatif</i>	6,10	4,14	-	10,25
<i>Impératif</i>	0,13	-	-	0,13
<i>Subjonctif</i>	0,53	4,12	-	4,65
<i>Infinitif</i>	-	2,31	2,18	4,50
<i>Participe</i>	-	0,18	1,47	1,65
<i>Adjectif verbal</i>	-	-	0,59	0,59
<i>Gérondif</i>	-	-	0,33	0,33
<i>Supin I</i>	-	-	-	0,006
<i>Supin II</i>	-	-	0,01	0,01
TOTAUX	6,76	10,75	4,58	22,11

% des occurrences de verbes

	Verbes principaux	Verbes de prop. subordonnées	Autres	Totaux
<i>Indicatif</i>	27,55	18,70	-	46,26
<i>Impératif</i>	0,62	-	-	0,62
<i>Subjonctif</i>	2,41	18,59	-	21,01
<i>Infinitif</i>	-	10,43	9,86	20,30
<i>Participe</i>	-	0,82	6,65	7,47
<i>Adjectif verbal</i>	-	-	2,70	2,70
<i>Gérondif</i>	-	-	1,50	1,50
<i>Supin I</i>	-	-	0,02	0,02
<i>Supin II</i>	-	-	0,08	0,08
TOTAUX	30,58	48,54	20,81	99,96

ANNEXE 10

a. Liste ordonnée selon le mode, le type de subordination et le temps

INDICATIF

AC, ATQVE

Indicatif parfait				
fuit	3, 3,	3, 3,	3, 6	
est merita	5, 6,	3, 1		

CVM

Indicatif présent				
confirmat	3, 3,	7, 8		
agitur	3, 5,	13, 7		
negas	3, 21,	2, 3		
potestis	3, 25,	2, 4		
conuertuntur	3, 31,	5, 7		
uertit	3, 31,	6, 8		
oritur	3, 31,	7, 2		
commeant	3, 31,	8, 3		
inest	3, 36,	11, 1		
considero	3, 39,	4, 1		
dicimus	3, 41,	4, 2		
circumimus	3, 47,	3, 3		
facitis	3, 63,	6, 1		
dicitis	3, 90,	5, 4		
uidemus	3, 91,	7, 5		
est	5, 16,	18, 4		
uolunt	5, 18,	6, 2		
dicit	5, 24,	10, 7		
dicimus	5, 29,	2, 6		
fatentur	5, 50,	8, 6		
praeteritur	5, 54,	7, 7		
uidet	5, 70,	8, 4		
Indicatif imparfait				
adferēbas	3, 18,	7, 3		
censebat	5, 57,	11, 9		
Indicatif futur				
disputabimus	3, 17,	6, 5		
Indicatif parfait				
amplificata sunt	3, 34,	4, 7		
dedit	3, 70,	14, 4		
imitatus est	3, 74,	7, 1		
sustulit	3, 76,	11, 5		
perdidit	3, 76,	12, 5		
communicauit	5, 56,	3, 5		
respondit	5, 56,	6, 2		
Indicatif plus-que-parf.				
clauserat	5, 59,	7, 5		
Indicatif futur antérieur				
perorāris	3, 4,	12, 5		
cognouero	3, 61,	10, 2		
traduxeris	5, 14,	3, 1		
perspexerit	5, 69,	4, 8		
uiderit	5, 69,	5, 3		

DVM

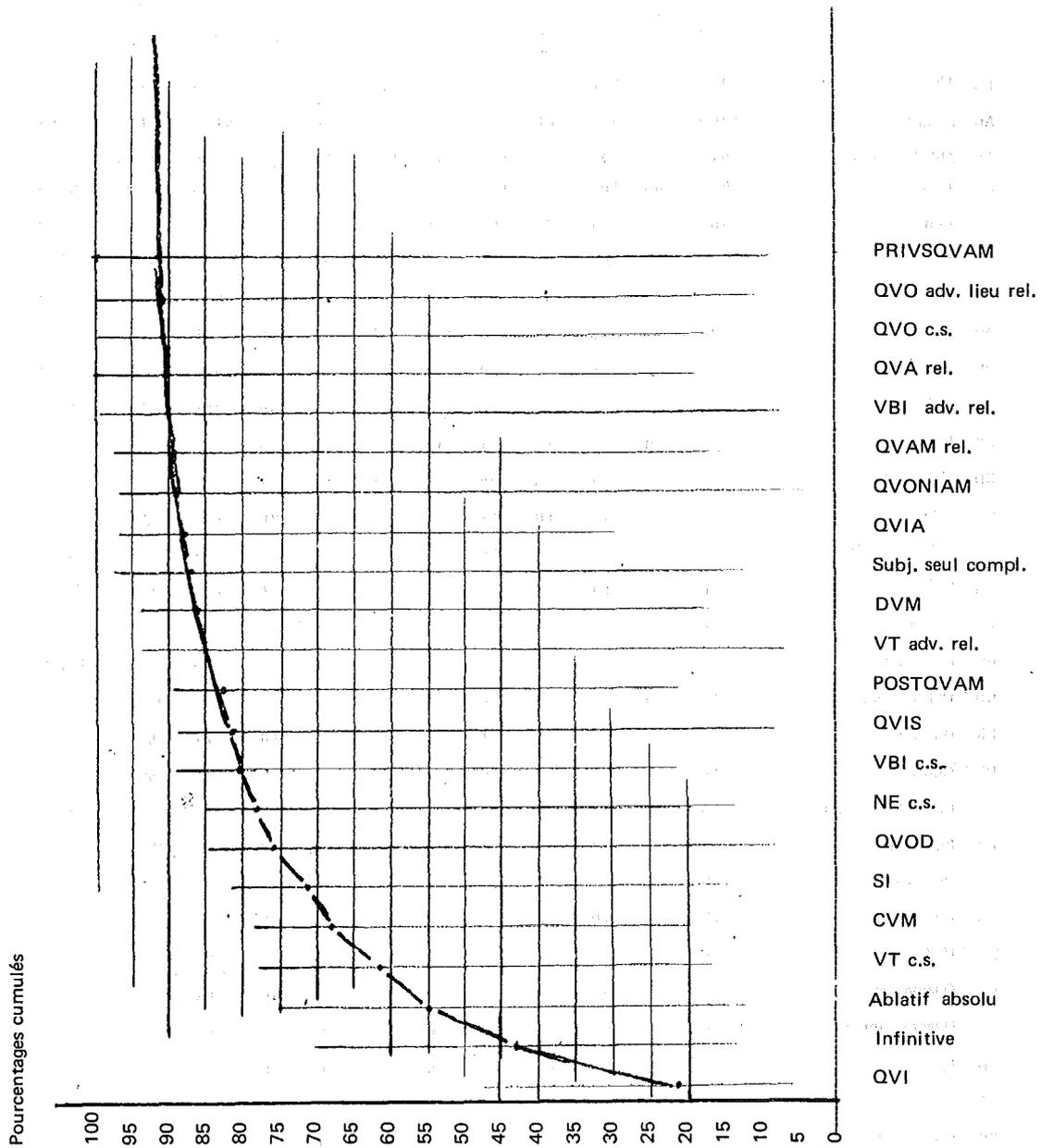
Indicatif futur				
potero	3, 94,	11, 3		
ETSI				
Indicatif présent				
nescio	3, 15,	14, 4		
est	5, 1,	7, 5		
est	5, 15,	2, 1		
debent	5, 38,	13, 4		
Indicatif parfait				
disputata sunt	5, 18,	11, 2		
adepta est	5, 54,	1, 5		
tradidit	5, 58,	3, 1		
NI				
Indicatif futur				
tenebimus	5, 46,	1, 4		
NISI				
Indicatif présent				
putamus	3, 27,	4, 8		
anteponuntur	3, 45,	15, 2		
est	5, 50,	4, 10		
fit	5, 51,	10, 6		
sunt	5, 52,	10, 6		
Indicatif futur				
censebit	5, 42,	10, 6		
POSTQVAM, POSTEAQVAM				
Indicatif présent				
adpropinquat	3, 67,	4, 1		
parat	3, 67,	4, 7		
QVALIS rel.				
Indicatif présent				
uis	3, 64,	11, 4		
uolumus	5, 42,	9, 6		
Indicatif parfait				
fuerunt	3, 12,	3, 5		
fuit	5, 54,	7, 5		
QVAM rel.				
Indicatif présent				
est	3, 95,	6, 7		
Indicatif imparfait				
putabatis	3, 93,	11, 7		
QVAMQVAM				
Indicatif présent				
uidentur	5, 23,	3, 1		
QVAMVIS				
Indicatif présent				
licet	3, 88,	9, 2		
QVANDO c-s				
Indicatif parfait				
deduxit	3, 43,	1, 7		
QVANTO rel.				

ANNEXE 11

		César De bello gallico	Salluste Catilina	Salluste Jugurtha	Titte-Live Extraits	Tacite Extraits	Ovide Métam.	Virgile Enéide	Virgile Bucoliques	Virgile Géorgiques	TOTAUX	Pourcentages	Pourcentages cumulés
LN	Qui	1252	224	460	260	256	191	379	62	139	3223	21,72	
AG	Infinitive	1260	161	416	564	298	69	266	40	90	3164	21,32	43,04
AD	Ablatif absolu	912	67	173	219	178	72	145	11	73	1850	12,46	55,50
XK	Ut c.s.	467	60	108	152	102	42	52	9	21	1013	6,82	62,32
BN	Cum	361	36	41	125	41	34	113	23	107	881	5,93	68,25
SX	Si	181	39	62	72	47	30	124	37	61	653	4,40	72,65
PG	Quod	343	33	38	33	29	17	4	1	3	501	3,37	76,02
DD	Ne c.s.	126	17	47	46	35	18	14	5	16	324	2,18	78,20
WK	Ubi c.s.	46	49	89	23	10	8	54	3	41	323	2,18	80,38
MS	Quis	96	27	15	29	13	7	55	16	30	288	1,94	82,32
ES	Postquam	19	30	75	19	21	13	31	4	3	215	1,44	83,76
XD	Ut adv. rel.	75	7	51	13	24	9	9	4	11	203	1,36	85,12
CD	Dum	30	11	13	10	18	27	47	19	20	195	1,31	86,43
AA	Subj. seul compl.	53	16	27	26	30	7	21	4	3	187	1,26	87,69
MD	Quia	0	17	41	33	33	2	5	0	4	135	0,90	88,59
PX	Quoniam	20	10	33	4	2	6	3	5	3	86	0,57	89,16
GG	Quam rel.	25	9	16	20	4	1	4	1	3	83	0,55	89,71
WD	Ubi adv. int.	18	3	16	8	6	4	15	1	9	80	0,53	90,24
FK	Qua rel.	12	1	5	6	1	9	18	1	18	71	0,47	90,71
NN	Quo c.s.	16	14	11	9	10	3	6	0	0	69	0,46	91,17
ND	Quo adv. lieu rel.	20	0	12	6	2	3	8	4	8	63	0,42	91,59
FA	Priusquam	30	6	7	10	1	0	6	0	2	62	0,41	92,00
MK	Quin c.s.	26	3	3	10	5	1	4	0	3	55		
EG	Nisi	11	6	10	6	9	2	1	1	4	50		
WS	Unde rel.	8	1	3	4	2	6	15	0	10	49		
HK	Quamquam	0	0	24	6	7	1	6	1	0	45		
KD	Quantus,um rel.	14	3	4	7	2	4	7	2	2	45		
ED	Ni	0	4	9	10	1	0	14	0	6	44		
EA	Neve, neu	9	6	7	9	1	7	1	0	2	42		
XG	Ut, adv. int.	3	1	6	4	1	0	9	11	5	40		
MG	Quicumque	12	5	4	5	1	3	7	0	2	39		

ANNEXE 12

-67-



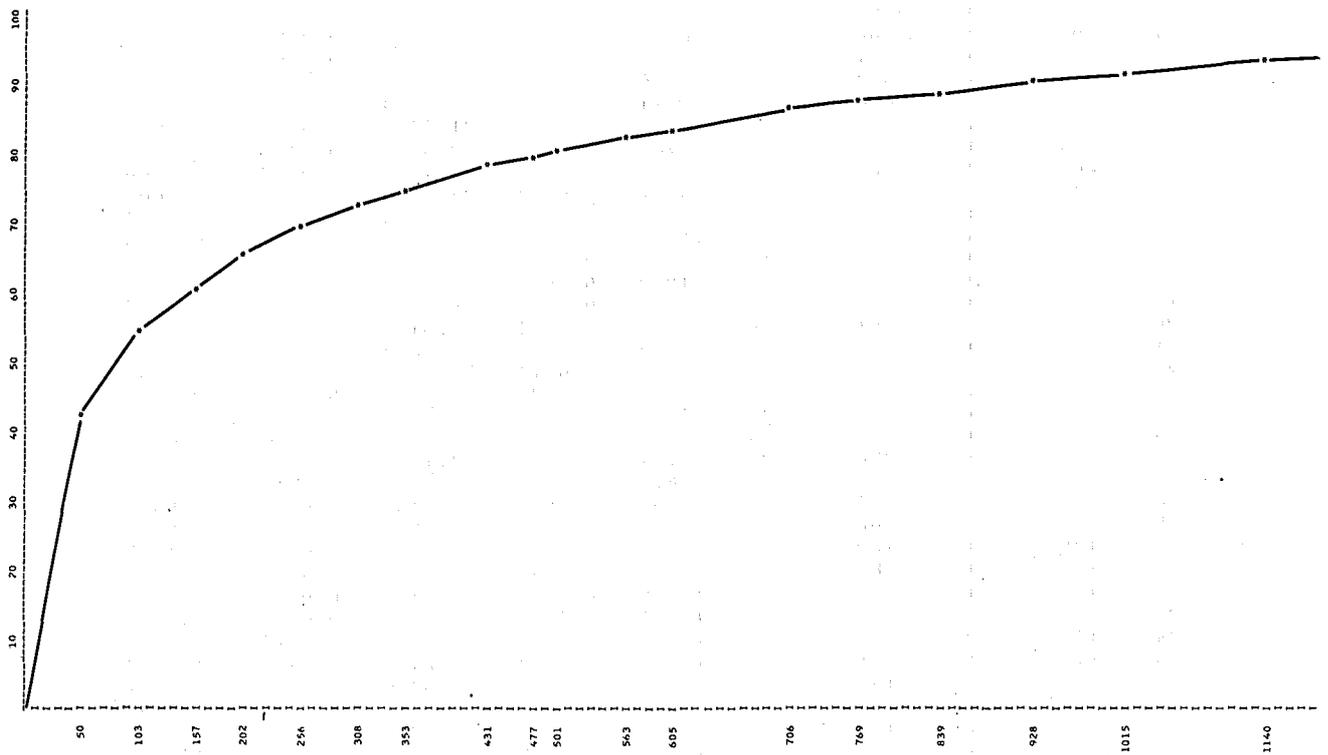
Courbe de distribution des subordinations en fonction du pourcentage

ANNEXE 13

CESAR, Bellum gallicum

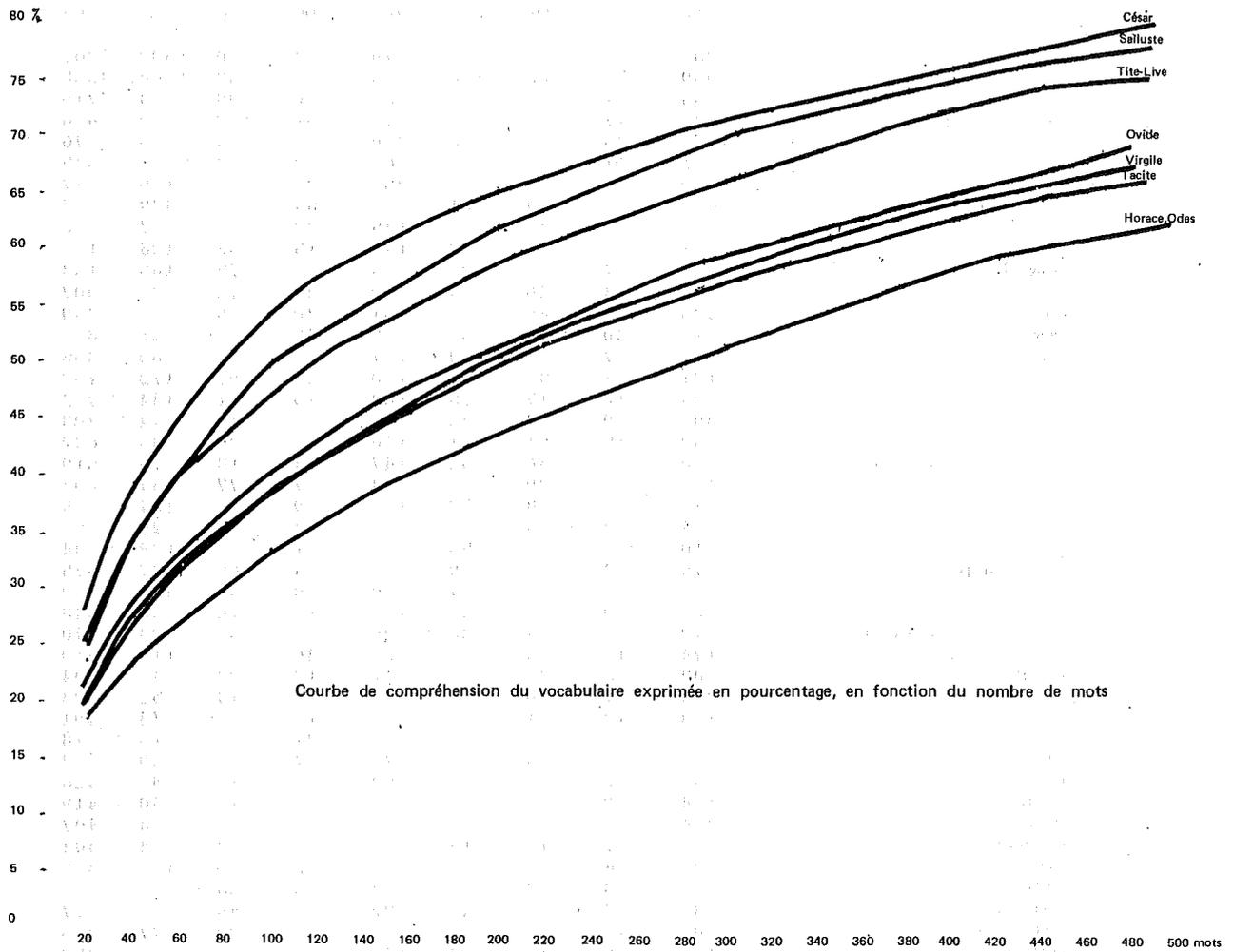
1	2	3	4	5	6	7	8	9	10
1	1172	1	1172	1	1172	2.7366	2.7366	0.0357	0.0357
2	1083	1	1083	2	2255	2.5288	5.2654	0.0357	0.0714
3	991	1	991	3	3246	2.3140	7.5795	0.0357	0.1072
4	983	1	983	4	4229	2.2953	9.8748	0.0357	0.1429
5	909	1	909	5	5138	2.1225	11.9973	0.0357	0.1786
6	907	1	907	6	6045	2.1178	14.1152	0.0357	0.2144
7	706	1	706	7	6751	1.6485	15.7637	0.0357	0.2501
8	661	1	661	8	7412	1.5434	17.3072	0.0357	0.2859
9	622	1	622	9	8034	1.4523	18.7596	0.0357	0.3216
10	610	1	610	10	8644	1.4243	20.1840	0.0357	0.3573
11	542	1	542	11	9186	1.2655	21.4495	0.0357	0.3931
12	523	1	523	12	9709	1.2212	22.6708	0.0357	0.4288
13	467	1	467	13	10176	1.0904	23.7612	0.0357	0.4646
14	435	1	435	14	10611	1.0157	24.7770	0.0357	0.5003
15	392	1	392	15	11003	0.9153	25.6923	0.0357	0.5360
16	352	1	352	16	11355	0.8219	26.5142	0.0357	0.5718
17	313	1	313	17	11668	0.7308	27.2451	0.0357	0.6075
18	302	1	302	18	11970	0.7051	27.9503	0.0357	0.6433
19	293	1	293	19	12263	0.6841	28.6344	0.0357	0.6790
20	291	1	291	20	12554	0.6794	29.3139	0.0357	0.7147
21	286	1	286	21	12840	0.6678	29.9817	0.0357	0.7505
22	279	1	279	22	13119	0.6514	30.6332	0.0357	0.7862
23	278	1	278	23	13397	0.6491	31.2823	0.0357	0.8220
24	273	1	273	24	13670	0.6374	31.9198	0.0357	0.8577
25	267	1	267	25	13937	0.6234	32.5433	0.0357	0.8934
26	250	1	250	26	14187	0.5837	33.1270	0.0357	0.9292
27	244	2	488	28	14675	1.1394	34.2665	0.0714	1.0007
28	234	1	234	29	14909	0.5463	34.8129	0.0357	1.0364
29	206	1	206	30	15115	0.4810	35.2939	0.0357	1.0721
30	189	1	189	31	15304	0.4413	35.7353	0.0357	1.1079
31	188	1	188	32	15492	0.4389	36.1742	0.0357	1.1436
32	187	1	187	33	15679	0.4366	36.6109	0.0357	1.1794
33	182	1	182	34	15861	0.4249	37.0359	0.0357	1.2151
34	174	2	348	36	16209	0.8125	37.8485	0.0714	1.2866
35	171	1	171	37	16380	0.3992	38.2477	0.0357	1.3223
36	167	1	167	38	16547	0.3899	38.6377	0.0357	1.3581
37	166	1	166	39	16713	0.3876	39.0253	0.0357	1.3938
38	158	1	158	40	16871	0.3689	39.3942	0.0357	1.4295
39	155	1	155	41	17026	0.3619	39.7562	0.0357	1.4653
40	152	1	152	42	17178	0.3549	40.1111	0.0357	1.5010
41	148	2	296	44	17474	0.6911	40.8023	0.0714	1.5725
42	146	1	146	45	17620	0.3409	41.1432	0.0357	1.6082
43	144	1	144	46	17764	0.3362	41.4794	0.0357	1.6440
44	135	1	135	47	17899	0.3152	41.7947	0.0357	1.6797
45	133	1	133	48	18032	0.3105	42.1052	0.0357	1.7155
46	132	1	132	49	18164	0.3082	42.4134	0.0357	1.7512
47	131	1	131	50	18295	0.3058	42.7193	0.0357	1.7869
48	125	1	125	51	18420	0.2918	43.0112	0.0357	1.8227
49	124	2	248	53	18668	0.5790	43.5903	0.0714	1.8942
50	121	3	363	56	19031	0.8476	44.4379	0.1072	2.0014

ANNEXE 14



1 Courbe de distribution du vocabulaire de CESAR, *De Bello Gallico*

ANNEXE 15



Courbe de compréhension du vocabulaire exprimée en pourcentage, en fonction du nombre de mots

ANNEXE 16

Echantillon du lexique de base proprement dit: ordre décroissant des fréquences

	CES	PH	OV	CAT	JUG	BUC	GEO	EN	TOT
QVE	907	29	546	124	279	73	650	1599	4207
ET	909	57	268	120	536	155	569	1044	3658
SVM	983	91	262	333	588	96	109	253	2715
QVI adj.-pr.	1172	77	153	222	457	61	132	352	2626
IN	1083	36	149	220	360	64	196	338	2446
IS	991	11	9	212	360	2	9	41	1635
AC, ATQVE	610	6	9	174	252	23	123	180	1377
HIC adj.-pr.	622	44	80	61	74	69	89	329	1368
AD	706	25	38	64	145	20	73	157	1228
SVI	661	20	25	72	173	15	62	128	1156
OMNIS	523	7	34	92	223	31	78	165	1153
AB	542	13	46	66	170	25	42	103	1007
VT adverbe	435	29	49	54	161	30	35	56	849
NON	313	30	73	43	88	45	94	128	814
EX	467	7	19	69	157	5	22	60	806
ILLE	121	43	84	78	150	54	97	172	799
MAGNVS	293	12	17	78	109	19	61	144	733
RES	392	9	8	85	154	2	14	44	708
SVVS	352	15	61	51	140	12	26	35	692
CVM préposition	250	6	36	64	147	24	18	74	619
IPSE	187	6	29	41	83	47	77	134	604
AVT	148	2	8	57	136	29	115	104	599
NEQVE	244	0	8	91	180	12	32	29	596
EGO	10	32	55	55	103	111	29	199	594
CVM conjonction	279	38	25	29	39	23	65	77	575
PER	104	2	57	47	90	9	68	175	552
FACIO	267	10	29	67	118	17	14	26	548
QVIS interrogatif	112	12	33	38	28	47	67	203	540
SI	174	14	28	41	61	40	48	113	519
POSSVM	302	18	34	29	41	12	23	45	504
LOCVS	278	7	20	25	107	1	13	52	503
SED	103	12	31	99	125	13	34	74	491
VIDEO	144	23	52	33	50	27	34	115	478
ALIVS	81	7	3	85	170	7	40	49	442
DE	206	6	20	58	62	7	16	51	426
HABEO	174	9	19	75	86	14	12	30	419
PARS	244	6	22	14	68	3	12	38	407
NOSTER	234	7	19	18	39	26	7	54	404
DO	98	14	45	19	31	6	40	146	399
IAM	71	3	23	7	43	33	68	135	383
MVLTVS	88	8	15	45	92	8	44	77	377
DICO, -ere	102	14	46	31	45	29	22	77	366
QVAM	109	5	11	54	140	6	11	16	352
ANIMVS	66	3	15	57	94	2	20	81	338
DIES	188	2	8	15	65	6	20	33	337
VENIO	135	14	14	11	40	26	28	67	335
VBI adverbe	56	5	10	47	95	4	43	60	320
SUPERVS	148	4	35	14	21	2	25	69	318
FERO	68	10	30	6	12	15	48	128	317
NE conjonction	155	10	18	19	55	9	20	29	315

ANNEXE 17

Echantillon du lexique de base proprement dit: ordre alphabétique

	CES	PH	OV	CAT	JUG	BUC	GEO	EN	TOT
AB	542	13	46	66	170	25	42	103	1007
ABVM	49	0	8	4	12	1	5	7	86
AC, ATQVE	610	6	9	174	252	23	123	180	1377
ACCIPIO	59	5	6	7	40	4	9	36	166
AD	706	25	38	64	145	20	73	157	1228
AGO	38	3	16	18	53	7	25	42	202
ALIVS	81	7	3	85	170	7	40	49	442
ALTER	56	6	8	11	22	8	12	16	139
AN	10	3	8	7	24	7	6	10	75
ANIMVS	66	3	15	57	94	2	20	81	338
ANNVS	46	1	12	6	11	3	16	18	113
ANTE, ANTEA	45	0	4	15	39	9	7	13	132
ANTE préposition	26	2	8	4	12	3	12	41	108
ARMA	81	0	3	24	56	1	12	91	268
AT	37	5	16	18	36	11	42	60	225
AVDIO	30	0	3	7	15	6	10	28	99
AVT	148	2	8	57	136	29	115	104	599
BONVS	15	4	10	51	59	4	13	21	177
CAPIO	64	5	7	17	51	4	14	27	189
CAVSA	152	5	6	18	15	2	5	18	221
CERTVS	47	1	10	5	7	1	10	15	96
CORPVS	12	6	25	18	16	1	27	54	159
CVM préposition	250	6	36	64	147	24	18	74	619
CVM conjonction	279	38	25	29	39	23	65	77	575
DE	206	6	20	58	62	7	16	51	426
DEVS	10	4	40	15	12	16	20	95	212
DICO, -ere	102	14	46	31	45	29	22	77	366
DIES	188	2	8	15	65	6	20	33	337
DO	98	14	45	19	31	6	40	146	399
DOMVS	33	4	13	28	24	15	17	47	181
DVM	26	14	23	11	12	18	21	38	163
DVO	79	5	13	7	22	6	6	9	147
EGO	10	32	55	55	103	111	29	199	594
EO, ire	27	3	16	5	29	11	16	68	175
EQVVS	28	1	14	5	11	1	17	40	117
ET	909	57	268	120	536	155	569	1044	3658
ETIAM	68	3	6	10	23	12	30	18	170
EX	467	7	19	69	157	5	22	60	806
FACIO	267	10	29	67	118	17	14	26	548
FAMA	12	1	5	11	17	2	5	31	84
FERO	68	10	30	6	12	15	48	128	317
FORTIS	11	2	4	8	11	0	9	17	62
GENVS	33	5	10	11	22	1	17	40	139
GRAVIS	22	4	5	5	9	5	15	19	84

Echantillon du lexique de base proprement dit: ordre grammatical

ADJECTIFS	CES	PH	OV	CAT	JUG	BUC	GEO	EN	TOT
PREMIERE CLASSE									
<i>Adjectifs en -VS</i>									
BONVS	15	4	10	51	59	4	13	21	177
CERTVS	47	1	10	5	7	1	10	15	96
MAGNVS	293	12	17	78	109	19	61	144	733
MEDIVS	29	1	20	3	7	6	30	83	179
MVLTVS	88	8	15	45	92	8	44	77	377
NOVVS	34	1	9	15	24	9	19	28	139
PARVVS	26	4	18	18	10	6	10	20	112
SVPERSVS	148	4	35	14	21	2	25	69	318
DEUXIEME CLASSE									
<i>Adjectifs consonantiques</i>									
VETVS	12	1	5	5	9	3	4	16	55
<i>Adjectifs en -IS</i>									
FORTIS	11	2	4	8	11	0	9	17	62
GRAVIS	22	4	5	5	9	5	15	19	84
NUMERAUX									
CARDINAUX									
DVO	79	5	13	7	22	6	6	9	147
VNVS	112	5	21	3	21	4	8	49	223
ORDINAUX									
PRIMVS	72	3	8	6	19	6	38	93	245
ADVERBES ORDINAUX									
PRIMVM	24	1	6	9	19	4	15	33	111
ADJECTIFS - PRONOMS									
PERSONNELS									
EGO	10	32	55	55	103	111	29	199	594
NOS	20	3	4	27	24	29	10	38	155

Taux de compréhension exprimé en pourcentages pour chaque auteur du programme en fonction du nombre de mots mémorisés

	Nombre de mots	César	Phèdre	Ovide	Catilina	Jugurtha	Bucoliques	Géorgiques	Énéide
BASE	144	48,83	36,54	37,73	46,11	46,75	36,34	34,65	36,58
CESAR	562	37,46	16,71	13,02	21,98	23,55	11,48	11,79	13,41
PHEDRE	111		15,03	4,76	1,10	3,17	7,83	3,33	4,40
OVIDE	343			22,54	4,75	5,04	12,61	14,15	16,17
CATILINA	272				14,29	8,17	2,85	2,89	3,23
JUGURTHA	123					4,51	0,79	1,23	1,16
BUCOLIQUES	139						13,58	5,64	2,56
GEORGIQUES								9,67	4,70
ENEIDE									5,40
TOTAUX	2077	86,29	68,28	78,05	90,23	91,19	85,48	83,35	87,61

ANNEXE 18

Microlexique de César, de B.G.

	CES	PH	OV	CAT	JUG	BUC	GEO	EN	TOT
QVOD	291	4	15	34	39	3	2	5	393
HOSTIS	286	4	2	30	97	1	6	23	449
CASTRÀ, -orum	273	0	0	9	40	1	3	9	335
LEGIO	189	0	0	5	9	0	1	0	204
CIVITAS	182	1	0	18	15	0	0	0	216
BELLVM	171	0	1	52	95	2	11	54	386
RELIQVVS	167	1	0	12	14	0	0	0	194
MILES	166	1	1	14	58	1	0	4	245
COPIA	146	0	3	20	31	1	2	3	206
OPPIDVM	133	0	2	0	47	1	4	0	187
EXERCITVS, -us	132	1	0	34	59	0	1	2	229
COGNOSCO	131	0	6	20	47	4	6	4	218
FINIS	125	0	4	3	20	3	2	28	185
LEGATVS	124	0	0	17	55	0	0	0	196
MILLE	124	0	3	1	8	1	1	13	151
CONSILIVM	121	2	1	35	29	0	0	8	196
PROELIVM	121	1	0	9	40	1	7	9	188
EQVES	120	1	0	4	33	0	1	3	162
ITER	113	0	7	13	33	0	2	20	188
EQVITATVS, la cavalerie	111	0	0	0	13	0	0	0	124
NAVIS	107	0	2	0	6	0	0	40	155
FLVMEN	103	1	9	0	12	11	30	13	179
INTERFICIO	83	2	0	5	14	0	1	0	105
MVLTITVDO	83	0	0	12	16	0	0	0	111
PROFICISCOR	80	0	0	10	25	0	0	4	119
CONSTITVO	79	0	0	5	10	0	1	6	101
EXISTIMO	79	1	0	12	19	0	0	0	111
PRO préposition	79	1	7	38	55	5	3	19	207
CONTENDO	76	1	1	3	3	2	2	7	95
RECIPIO	75	2	6	1	7	0	1	11	103
MVNITIO	74	0	0	0	2	0	0	0	76
OBSSES	73	0	0	0	1	0	0	0	74
VIRTVS	72	4	6	28	45	2	0	12	169
COGO	70	5	5	0	21	5	7	10	123
FVGA	70	0	1	5	23	0	8	26	133
GERO	70	1	7	12	37	0	3	13	143
AVXILIVM	68	3	2	7	20	0	1	14	115
CELERITER	65	1	0	0	0	0	0	0	66
EO adverbe	65	0	0	15	26	0	0	0	108
AGER	64	0	5	12	21	7	20	4	133
PRAESIDIVM	64	1	0	13	27	0	0	0	105
TRANSEO	64	1	6	2	3	0	2	3	81
TOTVS	63	3	17	2	7	7	22	38	159
PASSVS	62	0	3	0	2	0	1	2	70
COEPIO	61	11	4	11	19	4	4	6	120
CONVENIO	61	0	2	6	8	1	2	7	87
IMPERO	59	0	2	3	17	0	1	1	83
PVGNO	59	0	2	10	13	0	0	3	87
IMPERIVM	58	0	0	34	46	0	1	26	165
PERVENIO	58	0	4	2	15	1	1	2	83

ANNEXE 19

STYLISTIQUE

par

L. DELATTE

L.A.S.L.A.

Université de Liège

UNE MICROANALYSE STYLISTIQUE (Tacite, Annales, XIII, 15)

Les documents qui ont fait l'objet d'une description dans la communication précédente peuvent être utilement exploités au niveau du style. Je voudrais en faire la démonstration, en me fondant sur les listes de fréquence et les relevés grammaticaux.

J'ai choisi de procéder à une microanalyse d'un chapitre des Annales de Tacite (XIII, 15). Cet essai entrepris sans idée préconçue, sinon celle que le style de Tacite évoque le style poétique, montrera l'intérêt d'une analyse quantitative du fait littéraire.

Je propose de mettre en exergue à cet exposé, une réflexion d'un linguiste russe, Iouri Lotman, dans un livre intitulé "*La structure du texte poétique*", p. 234.

"L'étude de la fonction artistique des catégories grammaticales est équivalente, dans des relations déterminées, au jeu des structures géométriques dans les aspects spaciaux de l'art. Certains linguistes

pensent que l'élément esthétique, dans le texte, est exclusivement réservé aux "images". Cette idée implique que seule une couche peu importante de l'oeuvre est organisée artistiquement. Mais, la découverte de la fonction esthétique de la structure grammaticale permet de voir toute l'épaisseur du texte comme esthétiquement active."

Cette réflexion, qui met en évidence la valeur stylistique des éléments grammaticaux, justifie l'étude qui suit.

Par ailleurs, dans une certaine mesure, je m'inspirerai de la Nouvelle Critique et, dans ce cadre, de la stylistique structurale, telle qu'elle a été élaborée, entre autres, par Michael Riffaterre, un des plus brillants critiques de notre temps. Il s'agit d'une approche du fait littéraire très différente de la critique littéraire traditionnelle, dont le subjectivisme est si évident que les résultats d'une telle analyse éclairent plus la personnalité du critique que la personnalité de l'auteur critiqué.

La stylistique structurale est, en fait, une description linguistique objective du texte littéraire, le style étant compris comme "un soulignement (emphasis) expressif, affectif ou esthétique ajouté à l'information transmise par la structure linguistique, sans altération de sens" (Riff. p. 30). Cette mise en relief qui impose certains éléments de la séquence verbale à l'attention du lecteur n'apparaît que lorsque ces éléments sont un écart par rapport à une norme. Qu'est-ce qui constitue la norme ? Est-ce la langue en général ? Est-ce un genre littéraire ? Est-ce la langue poétique, lorsqu'on étudie un poète ? Toute définition prête à discussion. M. Riffaterre croit éviter tout reproche

de subjectivité en prenant comme norme, le contexte même du texte qu'il étudie.

Tout écart, par rapport à ce contexte, sera ressenti comme stylistiquement actif. Cette position n'est pas sans présenter quelque inconvénient mais elle évite à coup sûr le subjectivisme et l'arbitraire. En partant de ce principe, M. Riffaterre fonde la stylistique textuelle, c'est-à-dire finalement l'explication de texte sur deux critères : le contraste et la convergence. La convergence est l'accumulation d'une série de procédés stylistiques indépendants en un seul endroit du texte. Nous n'aurons pas à nous préoccuper ici de ce critère qui concerne plus les images que le texte lui-même. Par contre, le second critère est mieux adapté à notre étude : l'effet de contraste résulte de l'apparition dans le discours d'un terme inattendu.

Tel est l'essentiel de la doctrine de M. Riffaterre. J'emploierai un peu la même technique. Je comparerai le chapitre 15 à l'ensemble des Annales, après avoir montré les différences et les ressemblances avec d'autres oeuvres latines, historiques et poétiques. La comparaison, au niveau des catégories grammaticales mettra en relief le parti qu'on peut tirer de simples dénombrements pour faire apparaître certaines caractéristiques stylistiques d'un texte.

Voici le texte du chapitre 15 du livre XIII des Annales :

Turbatus his Nero et propinquo die quo quartum decimum aetatis annum Britannicus explebat, uolutare secum modo matris uiolentiam, modo ipsius indolem, leui quidem experimento nuper cognitam, quo tamen fauorem late quaesiuisset. Festis Saturno diebus inter alia aequalium ludicra regnum lusu sortientium euenerat ea sors Neroni. Igitur ceteris diuersa nec ruborem adlatura; ubi Britannico iussit exsurgeret progressusque in medium cantum aliquem inciperet, inrisum ex eo sperans pueri sobrios quoque conuictus, nedum temulentos ignorantis ille constanter exorsus est carmen, quo euolutum eum sede patria rebusque summis significabatur. Unde orta miseratio manifestior, quia dissimulationem nox et lasciuia exemerat. Nero intellecta inuidia odium intendit; urgentibusque Agrippinae minis, quia nullum crimen neque iubere caedem fratris palam audebat, occulta molitur pararique uenenum iubet, ministro Pollione lulio praetoriae cohortis tribuno, cuius cura attinebatur damnata ueneficii nomine Locusta, multa scelerum fama. Nam ut proximus quisque Britannico neque fas neque fidem pensi haberet olim prouisum erat. Primum uenenum ab ipsis educatoribus accepit tramisitque exoluta aluo parum ualidum, siue temperamentum inerat ne statim saeuiret. Sed Nero lenti sceleris impatiens minitari tribuno, iubere supplicium ueneficae, quod, dum rumore respiciunt, dum

parant defensiones, securitatem morarentur. Promittentibus
 dein tam praecipitem necem quam si ferro urgeretur cubi-
 culum Caesaris iuxta decoquitur uirus cognitis antea uenenis
 rapidum.

On trouvera, dans le tableau ci-dessous, la répartition des diverses catégories grammaticales.

Substantifs	74	36,6 %
Adjectifs	21	10 %
Adjectifs pronoms	16	8 %
Verbes	45	22,5 %
Adverbes	20	10 %
Prépositions	6	3 %
Conjonctions de Coordination	9	4,5 %
Conjonctions de Subordination	11	5,5 %
	<hr/>	<hr/>
	202	100 %

Une telle distribution ne présente guère de signification en elle-même. On constate que la catégorie des substantifs est prédominante; cependant, une telle constatation ne prendra toute sa valeur que si on peut, grâce à une comparaison avec un état de langue donné, établir qu'il s'agit d'une caractéristique de l'auteur étudié.

Mais le problème n'est pas si simple qu'il y paraît à première vue, car les études entreprises jusqu'à présent dans ce domaine font ressortir que la répartition des parties du discours est liée au genre littéraire. On sait que les Anciens maintenaient la distinction des genres avec une grande rigueur.

Quintilien, dans ses *Institutions Oratoires*, insiste souvent sur la nécessité de ne pas confondre les genres, même s'il y a –surtout s'il y a– entre certains d'entre eux de grandes affinités. Pline le Jeune, dans une lettre adressée à Titinius Capito (v, 8), qui lui suggérait d'écrire de l'histoire, se montre réticent, car il y a, dit-il, de grandes différences entre l'éloquence du barreau et le style de l'histoire :

“Sans doute, le discours et l'histoire ont beaucoup d'éléments communs, mais bien des caractères sont opposés même dans ces éléments que l'on croirait communs : l'un et l'autre racontent mais chacun d'une manière différente ... L'un recherche, par-dessus tout, la force, l'âpreté, la fougue, l'autre veut de l'étendue, de l'agrément et même de la grâce. Enfin, autres sont les mots, autres, les sonorités, autre, l'arrangement des mots :

postremo alia verba, alius sonus, alia constructio.

Cette dernière phrase illustre bien la différence entre ces deux genres.

Un autre témoignage est précieux dans ce domaine : Suétone, parlant du style d'Auguste, remarque que :

"son premier soin était toujours d'exprimer clairement sa pensée et, pour y mieux parvenir, pour ne jamais arrêter ni embarrasser l'esprit de ses lecteurs ou de ses auditeurs, il n'épargnait ni les prépositions qui déterminent le sens des mots ni les conjonctions qui lient les phrases et dont la suppression n'ajoute à la grâce du style qu'aux dépens de la clarté".

J'ai donc comparé ce chapitre à l'ensemble de l'oeuvre de Tacite, à la prose (Sénèque, *Epîtres à Lucilius*), à la prose historique (Salluste, *Catilina*, Tite-Live, Quinte-Curce), à la poésie épique (Virgile, *Enéide*) et à la poésie élégiaque (Tibulle) et aux Tragédies de Sénèque.

Cette comparaison est illustrée par un tableau que l'on trouvera en annexe.

La lecture des chiffres met en lumière diverses particularités du chapitre 15.

Tacite, plus que quiconque, utilise le substantif. Par contre, il emploie peu d'adjectifs-pronoms et peu de prépositions. Sur ces points, son style se rapproche davantage du style poétique que de la prose : ceci est particulièrement vrai de l'adjectif-pronom qui, par sa fonction même de suppléant du nom, favorise la négligence, le laisser-aller du style; rien n'est moins poétique que l'adjectif-pronom démonstratif, relatif, indéfini.

Pour s'en convaincre, il suffit de lire ce texte de Sénèque : (Cons. à Polybe, ch. 24) : *In hoc vero, cuius tam explorata pietas, pro certo habendum est nihil esse illi posse acerbius quam si tibi hic casus ejus acerbus est, si te ullo modo torquet ...*

En ce qui concerne l'adjectif, on remarquera que Tacite en fait un usage plus abondant que les autres historiens, sans cependant atteindre le niveau de la poésie.

Mais ce qui reste le plus frappant dans cette comparaison est l'abondance anormale des substantifs. Une analyse plus détaillée nous montrera le but que Tacite désire atteindre, consciemment ou inconsciemment : ralentir la lecture de son oeuvre.

Ce chapitre comporte 74 substantifs dont 40 sont des mots de caractérisation, c'est-à-dire des mots de fréquence relativement faible, mais d'une grande restriction de sens et très précis.

On remarquera qu'ainsi, Tacite obtient l'effet de contraste dont parle Riffaterre : effet qui résulte de l'apparition dans le discours, d'un terme inattendu. A cet égard, il faut remarquer qu'il est vrai que le contraste dépend moins de la valeur du terme contrastant dans la langue que de son opposition avec un terme neutre dans le discours. Il me paraît évident cependant que l'effet sera plus frappant encore si le terme contrastant est un mot de fréquence peu élevée, ce qui est le cas dans ce chapitre.

Violentia, indoles, experimentum, lusus, rubor, cantum, irrisum, convictus, dissimulatio, lascivia, etc. sont des mots rares dans le *corpus* taciteen. Leur faible probabilité d'apparition dans les contextes où nous les rencontrons et l'absence complète de redondance imposent au lecteur un décodage total du message, c'est-à-dire une attention exceptionnelle concentrée sur la forme.

C'est ce que veut Tacite, me semble-t-il. Toute son œuvre tend à demander au lecteur une réflexion, des retours en arrière, bref un ralentissement de la lecture.

Ceci est plus vrai encore si l'on prend en considération la nature des substantifs employés par Tacite : concrets ou abstraits.

Dans ce chapitre, sur les 40 mots de faible fréquence, 20 sont des abstraits. Cette proportion est étonnante : dans un texte qui raconte une histoire, on attendrait plutôt un vocabulaire extrêmement concret. Mais Tacite n'emploie l'abstrait que dans un sens concret : chez lui, l'abstraction est une figure qui sublime et par conséquent poétise les choses : *violentia* ne signifie pas la violence, mais un caractère violent, *convictus*, dont le sens premier est société, ne se comprend pas immédiatement dans son sens de banquet. Il y a une autre caractéristique stylistique très importante qui relève de la définition même de l'abstrait, c'est ce que j'appellerai son universalité. L'abstraction, dans la mesure où elle représente un concept, où elle est une propriété des sujets, confère au contexte où elle apparaît, une valeur universelle, spatiale et temporelle.

Considérons la phrase : "*Unde orta miseratio manifestior quia dissimulationem nox et lascivia exemerat*". L'emploi de *lascivia*, *dissimulatio*, *miseratio* suggère que la phrase est une vérité de tous les temps et chez tous les peuples.

De là, le texte prend une coloration poétique et une ampleur tragique que n'auraient pu lui accorder des noms concrets : ceux-ci auraient banalisé l'expression. La structure même de la phrase conduit à des conclusions identiques : il s'agit en effet, d'une phrase nominale, sans verbe personnel exprimé. Et Benveniste, dans un des articles qui constituent son volume intitulé : *Problèmes de linguistique générale*, note avec raison que "une assertion nominale pose l'énoncé en dehors de toute localisation temporelle, modale et hors de la subjectivité du locuteur" (p. 160).

Voilà ce qu'on peut dire de la catégorie des substantifs. On pourrait naturellement pousser plus loin l'analyse.

Considérons la dernière phrase de ce chapitre :
... *Cubiculum Caesaris juxta decoquitur virus, cognitio antea venenis rapidum.*

Tacite y joint, de façon frappante, en forme de chiasme, deux hapax, *virus* et *decoquitur* à deux mots fréquents, *cubiculum* et *venenum*. L'intention stylistique est évidente. Les mots qui appartiennent à un vocabulaire de la vie de tous les jours (et l'on peut dire que *cubiculum* et *venenum* sont dans ce cas) et qui, hors des stéréotypes conventionnels, jurent avec le contexte,

sont donc mis en relief au même titre que les hapax et se chargent ainsi d'affectivité.

Voilà comment Tacite intervient subtilement dans son récit et marque, dans les mots, son émotion devant les préparatifs sordides de la mort de Britannicus.

J'étudierai, à présent, l'emploi des modes et des temps dans ce chapitre et l'on verra combien la grammaire est utile pour discerner les nuances de la pensée.

On remarque que tout ce qui concerne Néron agissant est à l'infinitif narratif ou à l'indicatif présent : *volutare - intendit - molitur - jubet - minitari - jubere - decoquitur*. Le reste se trouve aux temps passés : *evenerat - exorsus est - provisum erat - accepit - tramisit*.

Il faut aussi remarquer que les deux séries de verbes sont entrelacées de telle sorte que les verbes au passé mettent ceux qui sont au présent de l'indicatif et à l'infinitif narratif, d'autant plus en relief. Et l'on comprend bien cette insistance si l'on se souvient de la valeur affective de ces modes.

D'abord l'infinitif. Marouzeau, dans son *Traité de stylistique latine*, remarque, à la suite de Perrochat, *L'infinitif de narration en latin*, que l'infinitif de narration se présente moins comme une expression que comme une notation.

Notation sommaire, dont on se contente, entre autres, lorsque, pour réaliser une espèce de coup de théâtre, on énonce brutalement et sans préparation le fait nu ou bien lorsque l'action à rapporter est par elle-même si notable, si propre à faire un effet, qu'il suffit de l'énoncer sans détour. Et Perrochat ajoute, (pp. 77, 78) que les historiens

''dans leur désir de se constituer un style personnel, éloigné des procédés de la rhétorique banale, ont largement utilisé un type d'énoncé qui s'accordait avec leur goût de l'expression brève, mâle, nerveuse, en même temps qu'avec leur conception dramatique et pittoresque de l'histoire''.

L'application au texte du chapitre 15 est évidente. *Volutare* relève de la seconde explication : c'est la préparation psychologique du crime, la bonne conscience que se donne le criminel, exprimée d'ailleurs par un intensif dont l'expressivité prend ici toute sa valeur. Par contre, *minitari* et *jubere* sont justifiés par la première nuance relevée par Marouzeau : lorsque, pour réaliser un coup de théâtre, on énonce brutalement le fait nu. Ici, en effet, c'est la décision finale, l'endurcissement de la volonté criminelle qui s'exprime. Et l'on remarque que Tacite, comme précédemment, emploie un intensif *minitari* dont la connotation de fébrilité est évidente.

Les indicatifs présents *intendit*, *molitur*, *jubet* et *decoquitur* sont des présents historiques. Riemann, *Syntaxe latine*, donne la valeur de ce mode : il présente les faits comme en voie d'accomplissement au moment même où l'on parle et mis, en quelque sorte, sous les yeux de l'auditeur ou du lecteur. Ainsi, le

lecteur a-t-il le sentiment qu'il participe à l'action. Cette remarque prendra plus de poids si l'on ajoute que l'indicatif présent est le temps de la poésie lyrique : c'est que, comme le fait remarquer P. Guiraud dans son ouvrage sur la Stylistique, la poésie est faite pour faire éprouver au lecteur les émotions et les passions à travers une forme qui, non seulement, les signifie mais encore les manifeste concrètement.

Je voudrais, pour terminer cette étude du verbe, dire quelques mots du participe, présent ou passé.

On sait qu'il s'agit d'un tour fréquent en latin : le participe permet de remplacer une proposition temporelle, causale, conditionnelle ou concessive.

On trouve chez Salluste 12 % de formes participiales, 17 % chez César, 20 % chez Tite-Live, 21 % chez Quinte-Curce, 23 % chez Tacite. Mais dans notre chapitre, on trouve 15 participes sur 46 verbes, soit plus de 32 %.

Je vois plusieurs raisons à cette prédilection de Tacite pour les participes.

Tout d'abord, cet emploi est une contribution au souci général d'économie qui est le propre de la phrase tacitéenne : le participe confère au style une densité, une concision, que n'apporte pas une proposition subordonnée circonstancielle, qui délaie la pensée en l'explicitant. Ceci est d'autant plus vrai que Tacite, dans ce chapitre, n'emploie aucune des particules dont souvent on accompagne le participe pour en préciser le sens : *vixdum, quippe, quasi, ut, nisi, etsi, etc.* Concision donc, mais aussi volonté délibérée de rendre plus difficile le décodage en obligeant le lecteur à rétablir lui-même la

nuance de la pensée de l'auteur.

Il y a, me semble-t-il, une seconde raison à cet emploi : G. Guillaume note, dans ses *Leçons de linguistique*, p. 170, que le participe passé est la forme morte du verbe. Cela est vrai. Le participe passé s'efface, s'estompe devant le verbe à un mode personnel : il exprime une idée importante pour le récit sans doute, mais secondaire par rapport aux verbes à un mode personnel qui lui sont proches dans la microchaîne de la phrase.

Cet emploi permet à Tacite d'hierarchiser sa pensée, même et surtout à un niveau subsyntaxique. Ainsi, *Turbatus*, cède le pas à *explebat* et, a fortiori, à *volutare*, qui, parce qu'il transgresse la norme de la régularité obligatoire d'un texte non artistique, est porteur d'une information capitale.

Ainsi, l'emploi du participe a permis, dans un chapitre consacré à l'un des grands événements du règne de Néron, de ramener tout le poids de l'attention sur les choses essentielles, exprimées elles-mêmes, soit à l'indicatif présent, soit à l'infinitif, et, dans un cas dont je voudrais dire un mot, au plus-que-parfait passif : *provisum erat*.

G. Guillaume ajoute, à la note à laquelle je viens de faire allusion, une réflexion qui nous aide à mieux cerner la pensée de Tacite. Voici ce qu'écrit Guillaume :

"Cette forme morte du verbe fait l'objet d'un traitement qui a cet effet de ressusciter le verbe, après sa mort, et de le conduire au-delà

de lui-même. Ce traitement anastatique, c'est l'addition d'un auxiliaire au participe passé, qui en fait un verbe second qui transcende et outrepassé le premier...

Le passé obtenu en se servant de l'aspect composé est toujours un passé résolu en une situation référée au présent".

Cette conception éclaire le sens de *provisum erat*, au § 6. Tacite aurait pu écrire et cela aurait correspondu à sa manière habituelle *provisum*. Dans cette hypothèse, le participe n'eût pas eu grande importance. Au contraire, avec l'auxiliaire exprimé et malgré le *olim*, Tacite se réfère au présent: Britannicus est, actuellement, dans la situation de quelqu'un pour qui il avait été prévu que sa domesticité n'aurait ni foi ni loi. Et cette décision du passé prend un poids tragique dans les circonstances actuelles puisque c'est de ses précepteurs qu'il reçoit le poison: *Primum venenum ab ipsis educatoribus accepit*.

On conclura, de cet examen général des catégories grammaticales, que, chez Tacite, tout concourt à donner au texte une coloration tragique, et, dans ce sens, on peut parler de style poétique. On s'en étonnera moins lorsqu'on se souviendra que Quintilien, au livre X, 1 des *Institutions oratoires*, constate qu'il y a une grande parenté entre l'histoire et la poésie :

“ L'histoire a beaucoup d'affinité avec la poésie, dont elle ne se distingue guère qu'en ce qu'elle n'est point assujettie à un mètre régulier”.

Par ailleurs, les courbes de fréquence citées dans la communication précédente font apparaître d'une façon frappante comment le style taciteen s'intègre dans la série des poètes.

Qu'indiquent ces courbes ? Que le vocabulaire des poètes est plus riche, plus diversifié que celui des historiens, qu'il utilise moins de mots-outils et de mots thèmes, plus de mots de caractérisation. Ainsi, on constate que pour couvrir 60 % d'un texte historique, il faut 200 mots, mais que pour couvrir 60 % d'un texte poétique, il en faut 350 chez Ovide, Virgile et plus de 450 chez Horace. Or, chez Tacite, il en faut 350 aussi. Ceci montre combien le style de Tacite, par la richesse de son vocabulaire, se rapproche du style poétique.

Il y aurait bien autre chose à remarquer dans ce chapitre et, en particulier, la composition en structures binaires, je veux dire la juxtaposition de deux éléments qui sont en relation de synonymie ou en situation de conflit. Ici encore, il faut citer un texte de Riffaterre (p. 65) qui met bien en relief la valeur de ce procédé : "ce qui fait la structure stylistique d'un texte, c'est une séquence ... de groupes binaires dont les pôles (contexte, contraste par rapport à ce contexte) sont inséparables, inexistants indépendamment l'un de l'autre."

Voici les groupes binaires que l'on trouve dans notre chapitre :

*Turbatus his et propinquo die
modo matris violentiam modo ipsius indolem*

*levi ... cognitam ... quo quaesivisset
ceteris adlatura ... ubi jussit
sobrios convictus ... temulentos
patria rebusque summis
nox et lascivia
nullum crimen neque caedem fratris
parum validum sive temperamentum
dum rumorem respiciunt, dum parant defensiones*

Il y a dans ce procédé répétitif, typique -notons-le-, de la poésie, quelque chose de lancinant qui évoque la marche inexorable du destin. Mais ces considérations sont en dehors de mon propos.

*

*

*

L'analyse que je viens de faire n'a explicité que deux des éléments qui, disait Pline le Jeune, distinguent l'éloquence de l'histoire : *alia verba ... alia constructio*. Il faudrait analyser aussi les sonorités, *alius sonus*, les comparer à d'autres oeuvres, par exemple, aux Lettres à Lucilius de Sénèque. Il s'agirait là d'un travail parfait pour l'ordinateur, dont la puissance s'accommode admirablement de telles recherches.

L. DELATTE

	Tacite Ann.XV, 14	Tacite	Salluste Catilina	Tite-Live Extraits	Quinte- Curce	Sénèque Lettres à Luc.	S. Tra- gédies	Tibulle Corpus	Virgile Enéide
Substantifs	36,6	36	32,5	32,4	31,2	21,5	33,5	29,5	34
Adjectifs	10	8,8	8,8	7,2	7	8,1	13,5	14	12
Adjectifs- Pronoms	8	7,8	12,5	10	11,3	16,5	10,6	12	8,5
Verbes	22,5	20	20,3	21,5	24,5	26,2	25	23	21,5
Adverbes	10	7,7	8	9,5	9	10	6,3	7	6
Prépositions	3	6	6,6	7,3	6	5,5	2,6	3	4
Conjonctions Coordination	4,5	8,8	7,6	7,4	7	7,5	6	8	10,5
Conjonctions Subordination	5,5	2,9	3,1	3,4	2,5	3,2	1,5	2,5	1,3

Répartition des catégories grammaticales exprimée en pourcentages