

# HYPERBASE<sup>©</sup>

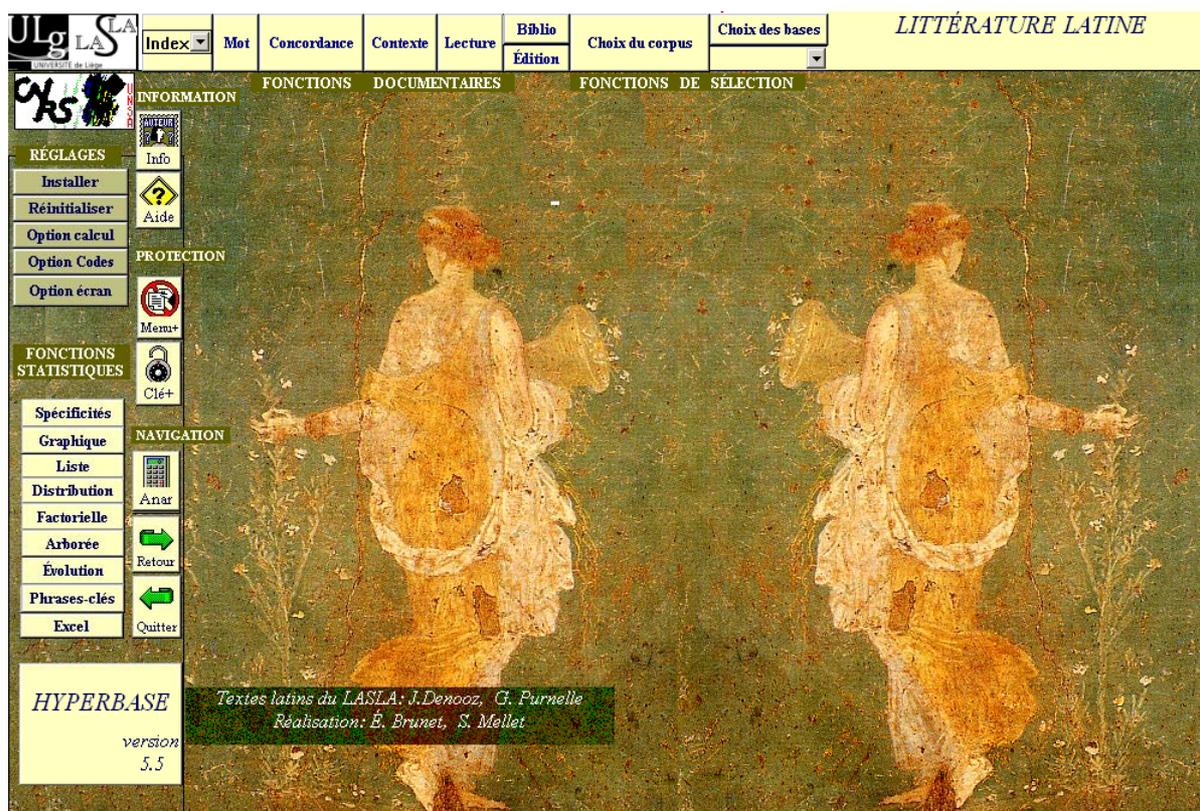
Logiciel hypertexte  
pour le traitement documentaire et statistique des corpus textuels  
conçu et développé par Étienne BRUNET

## MANUEL DE RÉFÉRENCE

pour la base de

## LITTÉRATURE LATINE

Sylvie MELLET, Étienne BRUNET\*



\* Nous remercions tout particulièrement Dominique Longrée pour ses nombreux contrôles sur le logiciel, ses suggestions stimulantes et sa relecture attentive du manuel. Il a grandement contribué à faire aboutir la réalisation de ce projet.

## PRÉSENTATION

---

Le CD-Rom « Littérature latine », réalisé dans le cadre du laboratoire « Bases, Corpus et Langage » (UMR 6039, CNRS/UNSA) et en collaboration étroite avec le Laboratoire d'Analyse Statistique des Langues Anciennes (LASLA, Université de Liège), comporte neuf bases, dont la principale – LATIN.EXE – rassemble trente-six recueils de textes couvrant une large période de la littérature latine allant de Plaute à Tacite ; ces textes ont été numérisés et lemmatisés au LASLA, selon des principes exposés par ailleurs dans diverses publications<sup>1</sup>. Ils sont associés, dans ce cédérom, au logiciel HYPERBASE, outil très performant d'exploration des textes et de traitement statistique des données numériques développé par Étienne Brunet pour l'analyse des textes français et adapté aux textes latins par lui-même et Sylvie Mellet.

Le maniement de la base générale étant assez lourd, le CD-Rom offre aussi des bases partielles, regroupant les œuvres d'un auteur particulièrement prolifique (Plaute, Cicéron, Sénèque, Tacite) ou bien des textes relevant d'un même genre littéraire (poésie, histoire, discours, traités) ; sur ce dernier point, nous nous en sommes tenus aux regroupements les plus pertinents et avons évité, par exemple, d'associer les comédies de Plaute et les tragédies de Sénèque sous une même étiquette « Théâtre » : la différence de ton et la distance chronologique entre ces textes sont trop importantes pour permettre leur confrontation. Les bases partielles permettent de travailler sur des sous-corpus restreints et parfois mieux adaptés à telle ou telle recherche. Leur taille réduite a aussi permis de donner à chaque livre d'une œuvre son autonomie, là où l'ampleur de la base générale nous avait obligés à des regroupements sévères. Ainsi, dans la base LATIN.EXE, les *Histoires* de Tacite forment un seul texte alors que la base TACITE.EXE donne accès à chacun des cinq livres séparément.

On trouvera donc, offerts au traitement par HYPERBASE, huit comédies de Plaute, les œuvres de Caton, Lucrèce, Catulle, la *Guerre des Gaules* (y compris son livre 8), la *Guerre Civile* et les trois commentaires dus à des épigones de César, tous les discours de Cicéron et trois de ses traités (*de Amicitia*, *de Officiis*, *de Senectute*), l'œuvre de Salluste, les *Géorgiques*, les *Bucoliques* et l'*Enéide* de Virgile, les poésies d'Horace, de Tibulle, Properce et Ovide, les quelques livres de Quinte-Curce qui nous sont parvenus, l'ensemble de l'œuvre de Sénèque (à l'exception des *Questions Naturelles*), le *Satiricon* de Pétrone, les *Satires* de Juvénal, le *Panegyrique de Trajan* de Pline le Jeune et, enfin, les *Histoires*, les *Annales* et les œuvres mineures (*Germania*, *Agricola*, *de Oratoribus*) de Tacite. La taille et les références exactes des éditions de référence sont accessibles *via* un menu BIBLIO qui donne accès à un fichier préparé par Gérald Purnelle.

Il est à noter enfin que les œuvres de la base ont été classées, autant que faire se pouvait, par ordre chronologique ; bien sûr, des incertitudes subsistent sur la date de plusieurs textes et les cas de superposition ont été tranchés de manière parfaitement

---

<sup>1</sup> Voir par exemple J. Denooz (s.d.) : *L'ordinateur et le latin*, « Techniques et méthodes », Université de Liège : LASLA ; S. Mellet (1996) : « Les atouts de la lemmatisation », in G. Moracchini (éd.) *Actes du Colloque international « Bases de données linguistiques : conceptions, réalisations, exploitations »* (Corte 11-13 octobre 1995), Univ. de Corse / Univ. de Nice - Sophia Antipolis, pp. 309-316 ; G. Purnelle (1996) : « Utilisation d'une banque de données de textes latins lemmatisés et analysés », in G. Moracchini (éd.) *Actes du Colloque international « Bases de données linguistiques : conceptions, réalisations, exploitations »* (Corte 11-13 octobre 1995), Univ. de Corse / Univ. de Nice - Sophia Antipolis, pp. 295-307.

arbitraire. On verra néanmoins ultérieurement l'intérêt d'un tel classement, en dépit de ses imperfections indéniables (cf. en particulier p. 27).

## INSTALLATION

La version Windows du présent logiciel a été réalisée avec TOOLBOOK (version 4). C'est celle qui sera présentée en détail ici, appliquée aux bases de textes latins lemmatisés et grammaticalement étiquetés, ce qui permet de travailler aussi bien sur les catégories grammaticales que sur le lexique.

L'installation doit pouvoir se faire sur un disque C (ce qui pose parfois des problèmes quand ce disque est réservé à l'administrateur d'un réseau) et dans un dossier HYPERBAS (sans -E à la fin) qui sera créé automatiquement par le programme ; s'il existe déjà un dossier HYPERBAS sur le disque, il convient donc de le débaptiser avant la nouvelle installation, en modifiant au moins une des huit lettres de son nom (un HYPERBAS\_OLD, par exemple, n'a aucune efficacité car la lecture du nom du dossier se fait seulement sur les huit premières lettres).

Pour lancer l'installation, solliciter l'un des deux programmes disponibles sur le CD-Rom : SETUPMIN.EXE qui réalisera l'installation minimale pour laquelle l'espace requis sur le disque C est de 21Mo seulement, car seuls les programmes – et pas les bases – sont installés sur le disque ; ou SETUPMAX.EXE qui installera et les programmes et les bases, configuration optimale qui nécessite un espace libre de 346 Mo au moins. Dans ce dernier cas, le CD-Rom n'est plus nécessaire et la rapidité de l'exploitation est grandement améliorée. Dans le cas d'une installation minimale au contraire, le CD-Rom reste indispensable et c'est à partir de lui qu'il faut ouvrir la base de son choix.

### **Outils complémentaires**

Le programme fait appel à des outils complémentaires standards tels que Word ou Excel. A la première utilisation, il convient donc de préciser l'emplacement de ces outils en sollicitant le bouton INSTALLER. Un texte s'affiche alors donnant l'adresse par défaut des programmes complémentaires ; si ces paramètres ne correspondent pas à la réalité de la configuration dont on dispose, répondre qu'ils ne sont pas exacts et suivre pas à pas le déroulement du questionnaire qui permettra d'enregistrer une fois pour toutes les bonnes adresses, cataloguées dans le dossier C:\HYPERBAS\REGLAGES.THF. A noter que Acrobat est fourni sur le CD-Rom et que son emplacement par défaut est donc correct.

### **Instructions si l'on dispose d'un système non français (Windows).**

Les paramètres régionaux qui accompagnent le système peuvent avoir une influence sur la gestion des caractères accentués, qu'il s'agisse du clavier, de l'affichage ou du tri. Il n'y a pas lieu de faire une réinstallation complète de Windows, si, par exemple, on utilise un système anglais et qu'on veuille traiter des données françaises. Windows a prévu qu'on puisse adapter aux besoins les réglages internationaux. La procédure est la suivante :

- cliquer sur POSTE DE TRAVAIL (ou l'équivalent anglais de ce terme)
- puis sur PANNEAU DE CONFIGURATION
- puis sur PARAMETRES REGIONAUX. On voit alors apparaître la carte du monde et au dessus de celle-ci un menu déroulant où l'on choisit FRANCAIS (STANDARD)
- puis cliquer sur OK et redémarrer.

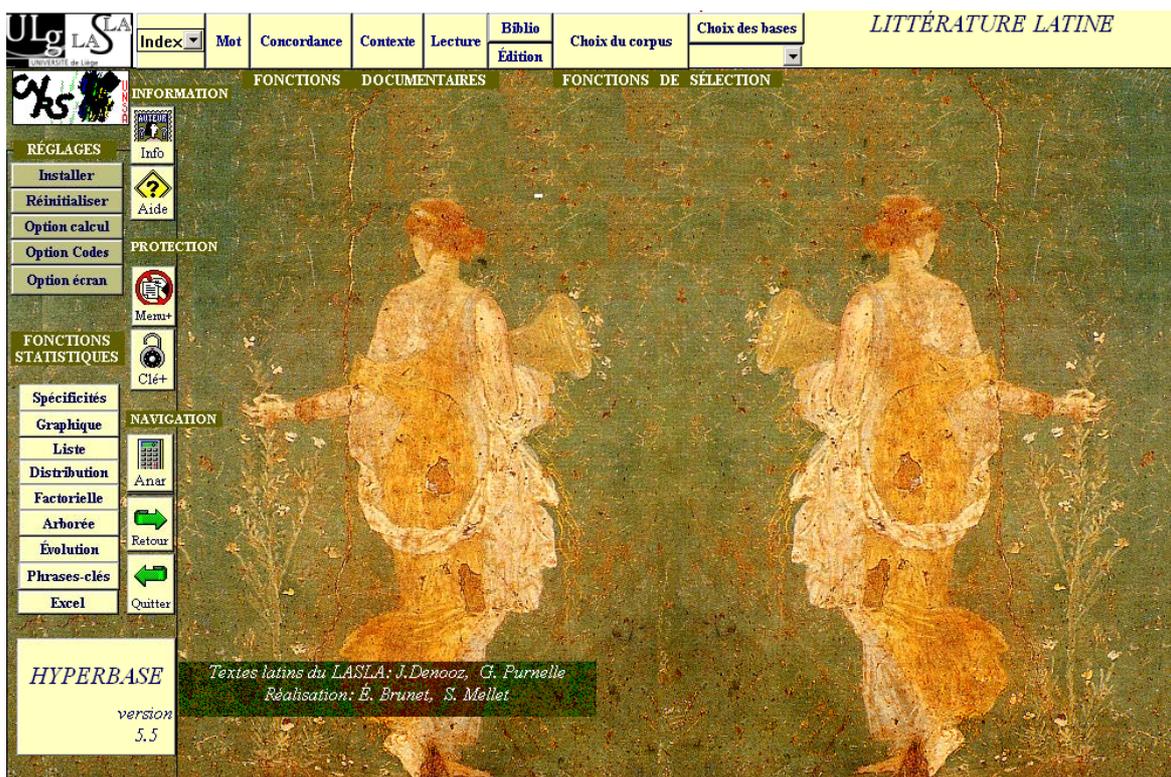
⚠ Veiller à ne modifier ni le nom, ni l'emplacement des fichiers ou dossiers essentiels à l'indexation ou à l'exploitation, en particulier MODELE.TXT qui sert de référence pour le calcul des spécificités, et toute la bibliothèque propre à l'application TOOLBOOK, sur laquelle la présente base est établie. Le « Runtime » essentiel porte le nom de TB40RUN.EXE et tous les programmes accessoires dont il a besoin ont un nom qui commencent par les mêmes initiales. Ne rien changer à cette bibliothèque.

## Aide générale

Dès que le curseur est dans le voisinage d'un bouton, une aide succincte apparaît au bas de l'écran pour indiquer la fonction de ce bouton. L'aide est beaucoup plus détaillée si l'on sollicite le bouton spécialisé AIDE. Elle correspond à l'essentiel du présent mode d'emploi. Un sommaire général est présenté d'abord, qui propose un choix de rubriques explicatives et qui permet de s'informer sur une action particulière. Si l'on choisit par un clic un des mots-clés du sommaire, on est renvoyé au paragraphe qui traite de la question. On peut aussi obtenir une aide appropriée à certains boutons du menu principal en pressant la touche MAJUSCULE pendant qu'on clique sur le bouton qui fait problème. Enfin cette aide en ligne peut être dissociée de la base : on la consulte de façon indépendante avec le lecteur d'Acrobat (qui est fourni).

## LE MENU PRINCIPAL

Le menu principal est le carrefour où l'on revient généralement après une opération. C'est là qu'on est introduit quand on ouvre la base. C'est là qu'on est conduit quand on sollicite le bouton SOMMAIRE, qui apparaît sur tous les écrans (sous la forme d'une flèche verte coudée). C'est là que mène l'ordre spécial (appui simultané sur les touches ALT et flèche haut) que le langage a prévu pour faire apparaître la première page de la base.



**Le menu propose deux sortes de fonctions principales :**

- les fonctions documentaires proposées horizontalement, au haut de l'écran
- les fonctions statistiques placées verticalement, à gauche de l'écran

À cela s'ajoutent quelques commandes de contrôle :

1 - les fonctions de **navigation** (verticalement en bas à gauche), relatives aux opérations de routine, qu'on trouve dans toute application : sortie (bouton QUITTER), retour à la page qu'on vient de quitter (bouton RETOUR). Une dérivation ouvre une base vide (ANAR.EXE, dans laquelle des calculs statistiques indépendants peuvent être réalisés).

2 - les fonctions de **réglage** (verticalement en haut à gauche). Outre le bouton INSTALLER (on ne le sollicite généralement qu'une seule fois), on peut parfois solliciter le bouton RÉINITIALISER, lorsqu'on désire modifier les titres qu'on a donnés aux textes du corpus. Attention ! Cette action modifie les titres dans toute la base et déclenche la sauvegarde des modifications. D'autres réglages, provisoires, intéressent soit le mode de certains calculs, selon la loi normale ou le modèle hypergéométrique (bouton OPTION CALCUL), soit la taille qu'on souhaite donner aux écrans (bouton OPTION ÉCRAN), soit enfin le mode d'affichage des textes (les formes graphiques pouvant être accompagnées de leur lemme - entrée du dictionnaire correspondante – ou du code grammatical qui leur est associé).

3 - les fonctions d'**information** : propriété intellectuelle (bouton en forme de timbre-poste) et AIDE (renvoyant au manuel en format pdf).

4 - les fonctions de **sélection** (horizontalement, en haut de l'écran après les fonctions documentaires), qui permettent de choisir une sous-base ou un sous-corpus.

La flèche qui se trouve sous le bouton CHOIX DES BASES permet de sélectionner l'une des neuf bases disponibles sur le CD-Rom.

 C'est bien la flèche orientée vers le bas qui ouvre ce menu. En cas d'erreur (clic sur le bouton rectangulaire CHOIX DES BASES qui donne la possibilité de supprimer ou d'ajouter une base), cliquer d'abord sur n'importe laquelle des réponses proposées par le programme, puis sur CANCEL.

 Il est à noter que si l'installation sur disque dur n'a pas été faite et qu'on travaille directement sur le CD-Rom, ce menu peut ne pas être opérationnel immédiatement. Il convient alors d'afficher le contenu du CD-Rom et de double-cliquer sur le fichier .EXE commandant l'ouverture de la base désirée.

Le bouton CHOIX DU CORPUS, quant à lui, permet de sélectionner quelques textes seulement parmi ceux qui constituent une base, en ignorant provisoirement les autres. La liste des textes se présente et il suffit de cliquer sur ceux sur lesquels on désire travailler (au clic de la souris les textes retenus passent dans le champ de droite) ; quand l'utilisateur estime son panier suffisamment rempli il sollicite le bouton TERMINER. Pour chaque texte sélectionné on est averti du nombre de mots qu'il contient (zone bleu marine en bas, à côté du bouton TERMINER).

ULg LA S UNIVERSITÉ de Liège		Index	Mot	Concordance	Contexte	Lecture	Biblio Édition	Choix du corpus	Choix des bases
CLIQUER SUR LE(S) TEXTE(S) SOUHAITÉ(S)			Sélection du corpus de travail (par défaut TOUS les textes)				FONCTIONS DE SÉLECTION		
N° 1 Quinctio 3-107 N° 2 S Roscio 107-264 N° 3 Q Roscio 264-322 N° 4 Tullio 322-383 N° 5 Caecilium 383-458 N° 6 Verrem 458-1478 N° 7 Fonteio 1478-1540 N° 8 Caecina 1540-1645 N° 9 Imperio 1645-1717 N° 10 Chentio 1717-1926 N° 11 Agraria 1926-2079 N° 12 Rabirio 2079-2122 N° 13 Catilinam 2122-2243 N° 14 Murena 2243-2345 N° 15 Sulla 2345-2441 N° 16 Archia 2441-2475 N° 17 Flacco 2475-2592 N° 18 Senatu 2592-2634 N° 19 Quirites 2634-2661 N° 20 Domo 2661-2812 N° 21 Hauspicum 2812-2881 N° 22 Sestio 2881-3039 N° 23 Vatinius 3039-3084 N° 24 Caelio 3084-3170 N° 25 Provinciis 3170-3221 N° 26 Balbo 3221-3290 N° 27 Pisonem 3290-3412 .....			1 Quinctio 4 Tullio 7 Fonteio 10 Chentio				 CICERON		

*Choix du corpus de travail*

## CHAPITRE 1

### *L'exploration*

#### EXPLORATION LIBRE

---

L'exploration a recours aux méthodes de l'hypertexte, de trois façons différentes et complémentaires :

##### **1 – en sollicitant le bouton MOT : navigation dans le dictionnaire**

Dans une première démarche, en sollicitant le bouton MOT, on est renvoyé au dictionnaire des fréquences, à l'endroit où se trouve le mot recherché. On peut aussi utiliser le menu déroulant INDEX pour choisir la lettre de l'alphabet où le dictionnaire doit être ouvert<sup>2</sup>. Le dictionnaire se présente en plusieurs colonnes : l'une est consacrée aux *formes* du texte, une autre aux *lemmes*, c'est-à-dire aux entrées de dictionnaire permettant ainsi de rassembler sous une même entrée toutes les formes fléchies d'un mot (genre, nombre, temps verbaux, etc.) ; les autres colonnes (codes, structures) seront examinées plus loin.

Une fois que le dictionnaire est sous les yeux, on peut parcourir la liste alphabétique en s'aidant de l'ascenseur et des flèches noires gauche et droite qui font défiler les pages. En cliquant sur une forme jugée intéressante, on voit apparaître tous les passages où cette forme est employée. Elle y apparaît en vidéo inverse dans le texte et l'écran reste figé, en attendant un clic de la souris pour passer à l'occurrence suivante. On peut à tout moment interrompre la recherche en appuyant sur la touche ALT, qui arrête le défilement à la page courante ou, au choix, restitue à l'écran la page du dictionnaire où la forme avait d'abord été observée.

Chaque ligne d'une colonne « formes » ou « lemmes » précise dans l'ordre :

a - la fréquence du mot ou du lemme dans le corpus

b - la forme elle-même

c - enfin, séparée du reste par une virgule, la liste des sous-fréquences de la forme ou du lemme dans les différentes parties du corpus. Noter que seules sont mentionnées les sous-fréquences non nulles, la lecture procédant par paires de chiffres dont le premier indique le numéro d'ordre du texte concerné et le second la fréquence du mot dans ce texte. Ainsi, dans l'image-écran ci-dessous, la suite 1 23 4 1 5 2 6 5 9 4... signifie que la forme *miles*, dans le texte n°1 se rencontre 23 fois, dans le texte n°4 se rencontre 1 fois, dans le texte n°5 se rencontre 2 fois, dans le texte n°6 5 fois, dans le texte n°9 4 fois, etc. Ce mode

---

<sup>2</sup> Cette méthode de navigation est utile si l'on n'est pas certain de la présence d'un mot dans un texte.

d'enregistrement fait gagner de la place, même si le déchiffrement n'en est pas très aisé. Au reste un tableau plus explicite, donnant notamment le nom des textes, apparaît au premier plan dès qu'on clique sur une forme particulière. Un nouveau clic sur l'un des textes affichés permet de circonscrire la recherche des contextes à ce seul texte. En cliquant sur la ligne TOUS LES TEXTES, on déclenche la visualisation exhaustive (qu'on peut toutefois interrompre avec la touche ALT). Un clic dans toute autre partie de la fenêtre blanche et bleue la fait disparaître et annule la sélection, qui peut reprendre avec un autre mot de la liste ou être transférée à n'importe quelle page du dictionnaire (grâce au bouton CHERCHER UN MOT).

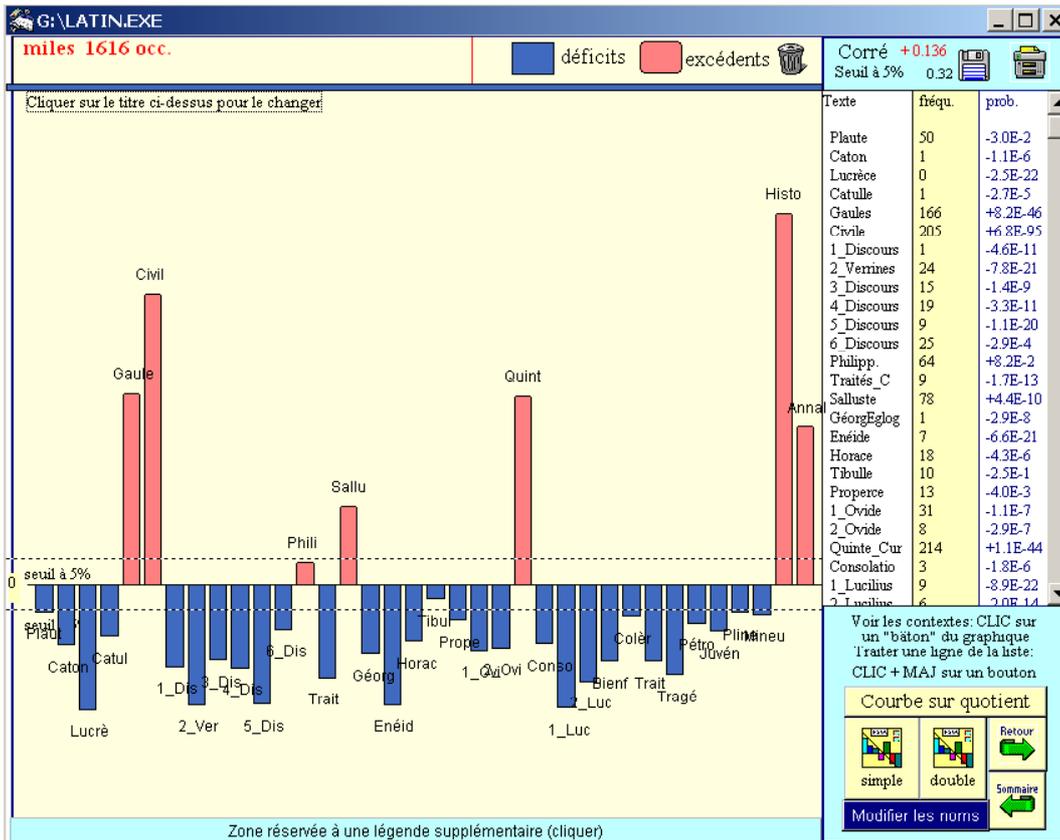
The screenshot shows the 'G:\LATIN.EXE' application window. It features a sidebar on the left with several buttons: 'Index hiérarchique', 'Chercher un mot', 'Graphique', 'Éditer', 'Quitter', 'Retour', and 'Sommaire'. The main area is divided into three columns: 'Formes', 'Lemmes', and 'Codes'. The 'Formes' column lists various grammatical forms of the word 'miles' with their respective frequencies. The 'Lemmes' column lists the source texts for these forms, such as 'N° 1 Plaute 23' and 'N° 4 Catulle 1'. The 'Codes' column contains alphanumeric codes. A pop-up window is overlaid on the 'Lemmes' column, displaying the word 'miles' and its total frequency of 299. It also provides instructions: 'CLIQUER SUR UN TEXTE (ou sur TOUS) pour y repérer les contextes du mot "miles"' and 'Cliquer AILLEURS dans cette fenêtre pour l'effacer'.

Une page du dictionnaire (ou index)

Outre la navigation d'un mot à l'autre, d'une page à l'autre du dictionnaire et de celui-ci au texte, le menu de la colonne de gauche offre la possibilité de quelques autres actions :

- l'impression de la page portée à l'écran (bouton EDITER),
- la représentation graphique de la distribution du mot sélectionné dans l'ensemble des textes (bouton GRAPHIQUE),
- enfin, une présentation hiérarchique du dictionnaire, par fréquences décroissantes (bouton INDEX HIÉRARCHIQUE).

Nous reviendrons sur l'analyse détaillée des deux copies d'écran ci-dessous dans la partie « Exploitation statistique ».



Graphique de la distribution du lemme miles dans le corpus général

G:\LATIN.EXE

**Les hautes fréquences**

Formes			Lemmes			Codes		
rang	frq	mot	rang	frq	mot	rang	frq	mot
1	93565	.	1	88604	.	1	122962	_s
2	41305	et	2	40004	et2	2	88603	.
3	27809	in	3	39989	svm1	3	87265	_m__1
4	25799	que	4	38546	qvi1	4	86771	_r
5	20417	non	5	27800	in	5	48930	_t
6	18680	est	6	25801	qve	6	34219	_a311
7	13815	cum	7	20415	non	7	33533	_a361
8	13531	ut	8	19092	hic1	8	32173	_a331
9	11009	ad	9	16929	is	9	25739	_a231
10	10735	qui	10	13729	ille	10	24640	_p
11	10597	quod	11	13637	ego	11	24302	_a232
12	10343	si	12	12589	tv	12	23967	_a261
13	9090	sed	13	12234	svm2	13	22411	_c1111
14	8196	quae	14	11156	qviv1	14	21607	_a211
15	8124	quam	15	10995	ad2	15	17867	_a341
16	7364	esse	16	10388	si2	16	17684	_i_11
17	6764	quid	17	10249	vt4	17	17489	_a332
18	6521	atque	18	9856	omnis	18	16432	_a131
19	6424	nec	19	9238	ab	19	15536	_a241
20	6411	aut	20	9130	sed	20	15294	_a111
21	6197	ex	21	9002	svi1	21	15233	_i_31
22	5951	se	22	8533	possvm1	22	15182	_b6_1_111300
23	5878	hoc	23	8299	cvm3	23	14724	_n
24	5790	de	24	7318	ipse	24	14135	_a161
25	5616	ne	25	7236	ex	25	13780	_c1321
26	5607	me	26	7149	facio	26	13030	_c1611
27	5571	te	27	7093	dico2	27	12822	_a262
28	5547	a	28	6714	svvs	28	12656	_j_11
29	4993	ac	29	6671	res	29	12555	_c1311
30	4459	per	30	6565	video	30	12254	_#
31	4329	enim	31	6441	atqve1	31	10262	_a212
32	4205	sunt	32	6410	avt	32	9994	_l_11
33	4055	mihi	33	6399	nec2	33	9335	_a312
34	3983	iam	34	5790	de	34	9216	_a362
						35	9144	_a342

Le dictionnaire hiérarchique

## **2 – en sollicitant le bouton LECTURE : navigation dans le texte**

La seconde démarche propre à l'hypertexte et symétrique de la précédente consiste à feuilleter les pages du texte : le bouton LECTURE permet d'ouvrir l'une des œuvres du corpus à la page qu'on veut<sup>3</sup> (avec la possibilité de choisir un affichage sur deux colonnes des formes et des lemmes ou des formes et des codes grammaticaux). Dans cette exploration, on peut suivre la séquence des pages (en suivant les flèches noires droite ou gauche) ou se reporter directement à un endroit particulier du même texte (en sollicitant le bouton PAGE), ou même rejoindre un autre texte du corpus (en sollicitant le bouton TEXTES). Deux boutons dont le symbolisme est clair sont prévus pour l'impression ou l'enregistrement de la page montrée à l'écran. L'enregistrement se fait dans un fichier au format *ascii* qui sert à empiler tout ce que l'utilisateur juge intéressant. Les extraits « empilés » ne peuvent excéder 300 mots l'un. Ce fichier, nommé EXTRAIT.TXT et automatiquement placé dans le dossier C:\HYPERBAS (sur le disque C donc) ; il laisse voir le dernier enregistrement et propose le choix entre l'ajout des données nouvelles aux anciennes ou l'effacement des données précédemment enregistrées.

Un bouton LOUPE grossit et rétrécit alternativement les caractères affichés à l'écran. Mais un hypertexte ne se réduit pas, bien évidemment, à la fonction lecture. Les fonctions proprement hypertextuelles sont celles de la recherche.

Le bouton CHERCHE correspond à une recherche locale qui se limite à la page courante et montre en rouge les occurrences du mot proposé, s'il est présent dans la page. Une fonction plus puissante est attachée à chacun des mots de l'écran. Un clic sur l'un d'entre eux suffit à montrer sa fréquence et sa répartition dans le corpus de travail. Un second clic à l'intérieur du tableau de répartition (fenêtre blanche superposée) déclenche la recherche et la visualisation des autres contextes où le même mot se rencontre dans le texte choisi ou dans l'ensemble du corpus, avant et après la page lue.

Ces parcours peuvent être suspendus provisoirement (pression sur le bouton de la SOURIS) ou stoppés définitivement (pression simultanée sur la touche ALT et la SOURIS). A chaque instant il est possible d'imprimer la page visible à l'écran.

Le voyage exploratoire dans les textes ne montre généralement que du texte ou des codes. Les chiffres sont pourtant sous-jacents dans chaque page. Leur présence est révélée dès qu'on clique sur le bouton ECARTS, au haut de la page, qui provoque la mise en relief (en rouge) des mots qui, à l'intérieur de cette page, sont caractéristiques du texte considéré par rapport au corpus d'ensemble (pour plus de détails sur cette notion d'écart, voir pp. 25-26).

## **3 – en sollicitant le bouton BIBLIOGRAPHIE : accès aux informations externes**

Le menu principal de la base propose un bouton BIBLIO qui permet de consulter les informations enregistrées : le fichier BASE.DOC donne la liste des textes constituant la base générale et la façon dont certains d'entre eux ont été regroupés. Le fichier CODESNEW.DOC donne accès à un tableau récapitulatif des codes en usage au LASLA (on y voit par exemple qu'un verbe de la première conjugaison à l'indicatif présent est codé B111) et le fichier CODESUB.DOC donne la liste des codes des propositions subordonnées. FICHIERS.DOC fournit la liste très complète de tous les textes contenus dans les neuf

---

<sup>3</sup> Les références fournies par le programme renvoient à des entités variées (phrases, paragraphes, vers) selon le type de texte ; elles peuvent en outre être approximatives l'unité de positionnement de l'outil informatique étant la page.

bases, avec leur édition de référence, leur taille et le type de structuration du texte (voir note 3). INDICES.DOC est un fichier extrêmement utile qui permet de lever les ambiguïtés attachées à l'homonymie de certains lemmes. Deux lemmes homonymes sont en effet distingués l'un de l'autre, au moment du codage, par un indice (1 et 2) ; bien que cette indexation ait été faite selon des principes fixes et explicites, le résultat n'en est pas facile à interpréter ni à mémoriser ; INDICES.DOC n'est autre chose que le dictionnaire des homonymes permettant de retrouver rapidement que *absque1* est adverbe tandis que *absque2* est préposition, que *accido1* est le dérivé de *cado* tandis que *accido2* est celui de *caedo*, etc. Enfin STRUCTUR.DOC est un fichier plus technique donnant la structure des enregistrements préparés par G. Purnelle pour l'exploitation HYPERBASE.



On rappelle que l'accès à ces fichiers n'est possible que si l'adresse de l'éditeur (généralement WinWord) a été correctement paramétrée grâce au bouton INSTALLER.



Notons aussi que si l'on travaille à partir du CD-Rom dans le cadre d'une installation minimale, le fait de sortir du CD-Rom pour solliciter WinWord oblige ensuite à revenir dans le CD-Rom et à ouvrir de nouveau la base de textes par un double-clic sur son icône. Avec une installation complète sur disque dur, ces allées et venues d'un programme à l'autre se font automatiquement.

## CHAPITRE 2

### *L'exploitation documentaire*

#### LA FONCTION CONTEXTE

---

En dehors de la circulation libre à travers le texte et le dictionnaire, le logiciel propose dans le menu principal les outils propres à assurer une exploitation méthodique de la documentation. Les deux programmes essentiels CONCORDANCE et CONTEXTE obéissent aux mêmes principes et ne se distinguent que par la présentation des résultats :

1 - Si l'on met en œuvre le bouton CONTEXTE, chaque occurrence de ce qu'on cherche est située et montrée dans le contexte naturel du paragraphe. Pour permettre la reconnaissance aisée du mot dans le contexte, ce mot est converti en capitales dans le paragraphe où il est rencontré.

On peut refuser l'équivalence entre contexte et paragraphe. Si cette option par défaut ne convient pas, cliquer sur le bouton LONGUEUR et choisir la longueur désirée pour chacun des extraits, exprimée en nombre de caractères (de 50 à 1 000). Noter que dans tous les cas on n'outrepasse pas la limite de cinq paragraphes consécutifs. Si l'extrait est raccourci, la contrainte de voisinage sera plus étroite et les résultats d'une recherche thématique (voir p. 30) perdront de leur pertinence.

Quand le mot cherché a une fréquence élevée, les résultats deviennent vite encombrants. Ils sont alors répartis sur plusieurs pages auxquelles on accède par les flèches droite et gauche (en haut et à droite de l'écran). Si ces flèches sont neutralisées, c'est que le terminus est atteint dans un sens ou dans l'autre. Dans tous les cas, les résultats affichés sont tronqués au bout de la sixième page, quand leur volume dépasse 200 000 caractères. Mais la recherche ne s'arrête pas pour autant et entrepose les résultats complets dans le fichier CONTEXTE.TXT (situé automatiquement dans le dossier C:\HYPERBAS). Là comme partout les deux boutons d'enregistrement et d'impression sont disponibles.

Le contexte restitué est généralement suffisamment explicite, d'autant que les références du passage sont livrées en clair, avec indication du texte, du chapitre et du paragraphe, mais aussi de la page au sein du corpus et de la zone dans la page (grâce à un code alphabétique qui commence à la lettre *a*, pour le début de page, et s'arrête à la lettre *f*, *g* ou *h*, pour la fin de page). Mais un lien est établi pour chaque extrait avec la page originale, où l'on est conduit instantanément lorsqu'on clique sur l'extrait en question.

Quo usque tandem abutere CATILINA patientia nostra .  
 \_\_\_\_\_ Cicéron In Catilinam I,1,1-11 Page: 2124 a \_\_\_\_\_

Ad mortem te CATILINA duci iussu consulis iam pridem oportebat in te conferri pestem quam tu in nos omnis iam diu machinaris .  
 \_\_\_\_\_ Cicéron In Catilinam I,2,1-8 Page: 2125 b \_\_\_\_\_

Habemus senatus consultum in te CATILINA uehemens et graue non deest rei publicae consilium neque auctoritas huius ordinis nos nos dico aperte consules desumus .  
 \_\_\_\_\_ Cicéron In Catilinam I,3,1-12 Page: 2126 b \_\_\_\_\_

Habemus enim eius modi senatus consultum uerum inclusum in tabulis tamquam in uagina reconditum quo ex senatus consulto confestim te interfectum esse CATILINA conuenit .  
 \_\_\_\_\_ Cicéron In Catilinam I,4,1-18 Page: 2127 c \_\_\_\_\_

Si te iam CATILINA comprehendi si interfici iussero credo erit uerendum mihi ne non hoc potius omnes boni serius a me quam quisquam crudelius factum esse dicat .  
 \_\_\_\_\_ Cicéron In Catilinam I,5,1-13 Page: 2128 a \_\_\_\_\_

Etenim quid est CATILINA quod iam amplius expectes si neque nox tenebris obscurare coetus nefarios nec priuata domus parietibus continere uoces coniurationis tuae potest si inlustrantur si erumpunt omnia .  
 \_\_\_\_\_ Cicéron In Catilinam I,6,1-13 Page: 2129 a \_\_\_\_\_

Num me fefellit CATILINA non modo res tanta tam atrox tam que incredibilis uerum id quod multo magis est admirandum dies .  
 \_\_\_\_\_ Cicéron In Catilinam I,7,1-15 Page: 2130 a \_\_\_\_\_

Fuisti igitur apud Laecam illa nocte CATILINA distribuisti partis Italiae statuisti quo quemque proficisci placeret delegisti quos Romae relinqueres quos te cum educeres discripsisti urbis partis ad incendia confirmasti te insum iam esse exiturum dixisti nauum tibi esse etiam nunc morae quod ego uiuerem

Exemple de résultats du programme CONTEXTE

## LA FONCTION CONCORDANCE

2 - Si l'on fait appel à la fonction CONCORDANCE du menu principal, on obtient un contexte étroit qui tient en une ligne et qui montre la forme cherchée, en position centrale, avec une demi-douzaine de mots à gauche et à droite. À côté des boutons habituels de navigation, d'enregistrement et d'impression, on remarquera une fonction de tri (bouton TRIER), qui est destinée à fournir une autre présentation de ces lignes. Au lieu de suivre l'ordre normal qui respecte la suite des textes, les contextes sont groupés selon l'ordre alphabétique du mot situé immédiatement à gauche ou à droite du mot-pôle. Cela souligne la résurgence de syntagmes répétitifs (par exemple *senatus auctoritas* ou *senatus consultum* dans l'image ci-dessous) qui soit ressortissent aux contraintes syntaxiques de la langue soit révèlent une tendance à la lexicalisation de certaines expressions ou encore les tendances phraséologiques spécifiques d'un auteur.

Si l'on estime trop étroite la fenêtre de concordance, un simple clic sur une ligne permet de retrouver la page concernée, qui reste exposée (avec mise en relief du mot) jusqu'au moment où un second clic la fait disparaître.

L'écran, même allongé par le défilement de l'ascenseur, ne peut guère montrer que 200 lignes à la fois. Si la concordance obtenue dépasse cette limite, un bouton apparaît qui propose les 200 lignes suivantes, et, de proche en proche, tout le contenu du fichier C:\HYPERBAS\CONCORDANCE.TXT où les résultats complets sont enregistrés.

The screenshot shows a window titled 'G:\LATIN.EXE' with a menu bar containing 'Forme', 'Lemme', 'Code', 'Syntaxe', 'Expr.', 'Initial', 'Final', 'Chain', 'Liste', 'Tout', and 'Nb'. The 'Nb' field displays '857' and the title bar says 'CONCORDANCE'. Below the menu bar, there are icons for 'Trier', 'Notes', a printer, a trash can, 'Retour', and 'Sommaire'. The main area contains a list of text fragments with their search results. The list is as follows:

Forme	Lemme	Code	Syntaxe	Expr.	Initial	Final	Chain	Liste	Tout	Nb	CONCORDANCE
D2	3131c			frequentis esse in curia se iussu	senatus	a nobis	petere ut	ueniremus .			
D2	2575b			titutis superiorum ab auctoritate	senatus	ab iure omnium	Sicilorum ne	re			
Ga	1616b			uem magistratum magnam que partem	senatus	ad eum conuenisse	legatos ad V				
D2	3078b			em non oportere . Tum frequens	senatus	ad istum uenit	pollicetur sign				
D2	2947b			si homines aut benigni cum pluris	senatus	aestimasset quam	quanti esset				
D2	2947b			asset quam quanti esset annonae ex	senatus	aestimatione non	ex annonae ra				
D4	4086c			ius Flaccus consules post cunctus	senatus	atque ille senatus	quem etiam				
D5	4603c			s me cum iura magistratum me cum	senatus	auctoritas me cum	libertas me				
D2	2932c			tibi hoc concessit quae lex quae	senatus	auctoritas quae	porro aequitas				
D2	3291c			s ubi forum plenum iudiciorum ubi	senatus	auctoritas ubi	existimatio pop				
D4	4307b			ta sunt in quibus tamen Murena ab	senatus	auctoritate	defenditur . Qu				
Ci	1715b			em legati domum referunt atque ex	senatus	auctoritate haec	Caesari renun				
D2	2752a			od tua sponte iniussu populi sine	senatus	auctoritate iura	prouinciae Si				
D4	4366a			hoc iudices nisi recenti memoria	senatus	auctoritatem huius	indici monu				
D4	4390b			t improbis largitionibus restitit	senatus	auctoritatem numquam	impediuit				
D5	4584b			nitatem ut rei publicae statum ut	senatus	auctoritatem ut	ciuis optime m				
Ci	1668b			se sibi consilium capturum neque	senatus	auctoritati obtemperatum	hab				
D2	2859b			iciliae uenire potuerunt si id te	senatus	aut populus Romanus	facere uol				
D2	3091b			es in eum dantur . Rem cunctus	senatus	Catinensium	legibus iudicabat				
D2	2947c			sumpsisses id nummorum quod tibi	senatus	cellae nomine	concesserat .				
D4	4182c			um ex memoria uobis Quirites quid	senatus	censuerit	exponam .				
Ga	1250c			m M. Messala M. Pisone consulibus	senatus	censuisset uti	quicumque Galli				
D4	3982a			Quid . Ista omnia de quibus	senatus	censuit nominatim	in lege pers				
D4	4370b			recentem putarem memoriam cuncti	senatus	commentario meo	posse superari				
D4	4408c			qui praetores fuerunt et uniuersi	senatus	communis est	laus ut constet p				
D5	4609c			cio deorum immortalium testimonio	senatus	consensu Italiae	confessione i				
D4	4034a			t litterae Quirites publicae sunt	senatus	consulta complura .	Statuer				
D2	2832a			teras adfert ut sibi contra omnia	senatus	consulta contra	omnia iura con				
D4	3981b			uendere omnia de quibus uendendis	senatus	consulta facta	sunt M. Tullio				
D2	2333a			uli neque suas leges neque nostra	senatus	consulta neque	communia iura t				
Ci	1675c			nus plebis . De reliquis rebus	senatus	consulta perscribuntur .	Pr				
Ga	1263d			is cum Haeduis intercederent quae	senatus	consulta quotiens	quam que hon				
D4	4037a			qui malis contionibus turbulentis	senatus	consultis iniquis	imperiis rem				
D2	2820b			itate defendimus unum hoc aliquot	senatus	consultis reprehensum	decretum				
D2	2414a			num relatarum . Recita . Ex	senatus	consulto .	Ut hoc pacto rat				
D4	4110b			M. Fuluius consularis . Simili	senatus	consulto C. Mario	et L. Valeri				
D2	3138a			it ex ordine . Id adeo ex ipso	senatus	consulto cognoscite	nam princ				

At the bottom right of the window, there is a red button labeled 'Extrait 1 Suite ->'.

### Exemple de résultats du programme CONCORDANCE

⚠ Il est à noter qu'à la différence de ce qui se passe pour le fichier EXTRAIT.TXT, le programme ne propose pas, pour les fichiers CONTEXTE.TXT et CONCORDANCE.TXT, un ajout des nouvelles données aux résultats déjà enregistrés : toute nouvelle requête écrase les données précédentes ; il est donc indispensable, si l'on veut sauver une série de contextes ou une concordance, de l'enregistrer immédiatement sous un autre nom au moyen des fonctions de Word.

### OBJET DE LA RECHERCHE

Dans les deux procédures, CONTEXTE et CONCORDANCE, des options sont offertes à l'utilisateur pour qu'il puisse préciser l'objet de sa recherche (qui est limitée au corpus de travail préalablement sélectionné). Cette recherche peut porter non seulement sur une forme, mais aussi :

a - sur une EXPRESSION (une suite de mots ou de signes), par ex. *senatus populus que*. Bien noter à propos de cet exemple que les enclitiques sont séparés du mot sur lequel ils s'appuient par un blanc.

b - sur un LEMME, par exemple toutes les formes du verbe *dico*. L'option LEMME permet le regroupement des formes derrière la forme canonique telle qu'elle apparaît dans le Gaffiot (avec pratique systématique de l'assimilation consonantique du préfixe). L'affichage des résultats peut se faire, au choix, sous la forme d'un texte réduit à la succession des lemmes (peu lisible) ou en rétablissant les formes fléchies. Par ailleurs il est possible d'introduire dans la requête une contrainte supplémentaire concernant soit le code grammatical affecté à l'occurrence du lemme recherché (bouton CODE), soit son insertion dans un type de syntagme (bouton STRUCTURE).

**Désirez-vous imposer un code grammatical ou une structure syntaxique?**

aucune contrainte

code

structure

La première option permet, par exemple, de rechercher toutes les occurrences du verbe *fero* affectées du code du subjonctif ; la seconde option permet de rechercher toutes les occurrences du lemme *amicus* intégré dans une structure syntagmatique /adjectif + substantif + verbe/ dans laquelle, bien sûr, il occupera la première place ou du lemme *tandem* intégré en deuxième position dans le syntagme /adverbe interrogatif + adverbe + verbe/.

Sommaire
Notes
Cherche
Forme
Lemme
Code
Syntax
LONG
parag
CONTEXTES page 1
autres résultats

Expr
Coocc
Initial
Final
Liste
134
Cliquer sur un extrait pour voir la page

Quem ego si FERREM iudicem refugere non deberet cum testem producam reprehendere audebit .  
 \_\_\_\_\_ Cicéron Pro Q. Roscio Comoedo oratio,45,1-11 Page: 310 b \_\_\_\_\_

Cognitor si fuisset tuus quod uicisset iudicio FERRES tuum cum suo nomine petiit quod abstulit tibi non sibi exegit .  
 \_\_\_\_\_ Cicéron Pro Q. Roscio Comoedo oratio,53,1-9 Page: 318 b \_\_\_\_\_

Nunc ne nouo querimoniae genere uti possit Hortensius et ea dicere opprimi reum de quo nihil dicat accusator nihil esse tam periculosum fortunis innocentium quam tacere aduersarios et ne aliter quam ego uelim meum laudet ingenium cum dicat me si multa dixissem subleuaturum fuisse eum quem contra dicerem quia non dixerim perdidisse morem illi geram utar oratione perpetua non quo iam hoc sit necesse uerum ut experiar utrum ille FERAT molestius me tunc tacuisse an nunc dicere .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda I,24,1-9 Page: 539 b \_\_\_\_\_

Quo modo ex deciens HS sescenta sint facta quo modo DC eodem modo quadrarint ut illa de Cn. Carbonis pecunia reliqua HS sescenta facta sint quo modo Chrysogono expensa LATA sint cur id nomen infimum in litura que sit uos existimabitis .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda I,92,1-16 Page: 609 c \_\_\_\_\_

Quod minus Dolabella Verri acceptum rettulit quam Verres illi expensum TULERIT HS quingenta triginta quinque milia et quod plus fecit Dolabella Verrem accepisse quam iste in suis tabulis habuit HS ducenta triginta duo milia et quod plus frumenti fecit accepisse istum HS deciens et octingenta milia quod tu homo castissimus aliud in tabulis habebas .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda I,100,1-14 Page: 617 a \_\_\_\_\_

Homo stultissime et amentissime tabulas cum conficeres et cum extraordinariae pecuniae crimen subterfugere uelles satis te elapsurum omni suspitione arbitrabare si quibus pecuniam credebas iis expensum non FERRES neque in tuas tabulas ullum nomen referres cum tot tibi nominibus acceptum Curtii referrent .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda I,102,1-9 Page: 619 a \_\_\_\_\_

*Contextes du lemme fero affecté du code subjonctif*

G:\CICERON.EXE

Forme Lemme Code Syntax LONG paragr CONTEXTES page 1

Expri Coocc Initial Final Liste 89 Cliquer sur un extrait pour voir la page autres résultats

Fac omnis AMICISSIMOS Fonteio fuisse tantum hominum numerum ignotissimorum atque alienissimorum pepercisse huius capiti consuluisse famae res ipsa tamen ac ratio litterarum confectio que tabularum habet hanc vim ut ex acceptis et datis quidquid fingatur aut surripiatur aut non constet appareat .  
 \_\_\_\_\_ Cicéron Pro M. Fonteio oratio,3,1-20 Page: 1491 b \_\_\_\_\_

Atque ut omnes intellegant me L. Lucullo tantum impertire laudis quantum forti viro et sapienti homini et magno imperatori debeatur dico eius aduentu maximas Mithridati copias omnibus rebus ornatas atque instructas fuisse urbem que Asiae clarissimam nobis que AMICISSIMAM Cyzicenorum oppressam esse ab ipso rege maxima multitudine et oppugnatam uehementissime quam L. Lucullus uirtute adsiduitate consilio summis obsidionis periculis liberauit .  
 \_\_\_\_\_ Cicéron De Imperio Cn. Pompei ad Quirites oratio = Pro lege Manlia,20,1-14 Page: 1665 b \_\_\_\_\_

Erant illi compti capilli et madentes cincinnorum fimbriae et fluentes purpurissatae que buccae dignae Capua sed illa uetere nam haec quidem quae nunc est splendidissimorum hominum fortissimorum uirorum optimorum ciuium mihi que AMICISSIMORUM multitudine redundat .  
 \_\_\_\_\_ Cicéron In L. Calpurnium Pisonem oratio,25,1-15 Page: 3335 a \_\_\_\_\_

Quaerunt quidem C. Laeli multi ut est a Fannio dictum sed ego id respondeo quod animum aduertit te dolorem quem acceperis cum summi uiri tum AMICISSIMI morte ferre moderate nec potuisse non commoueri nec fuisse id humanitatis tuae quod autem Nonis in collegio nostro non adfuisse ualetudinem respondeo causam non maestitiam fuisse .  
 \_\_\_\_\_ Cicéron De Amicitia,7-9 Page: 4754 b \_\_\_\_\_

*Contextes du lemme amicus intégré dans un syntagme /adjectif + substantif + verbe/*

Forme Lemme Code Syntax LONG paragr CONTEXTES page 1

Expri Coocc Initial Final Liste 162 Cliquer sur un extrait pour voir la page autres résultats

Quod si ei potissimum censes permitti oportere accusandi potestatem cui maximam C. Verres iniuriam fecerit utrum TANDEM censes hos iudices grauius ferre oportere te ab illo esse laesum an prouinciam Siciliam esse uextatam ac perditam .  
 \_\_\_\_\_ Cicéron Divinatio in Q. Caecilium,54,1-10 Page: 438 a \_\_\_\_\_

A Phylarcho Centuripino homine locupleti ac nobili phaleras pulcherrime factas quae regis Hieronis fuisse dicuntur utrum TANDEM abstulisti an emisti .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda IV,29,1-19 Page: 1151 a \_\_\_\_\_

Exspectabant omnes quo TANDEM progressurus aut quidnam acturus esset cum repente hominem propi atque in foro medio nudari ac deligari et uirgas expediri iubet .  
 \_\_\_\_\_ Cicéron In C. Verrem actio secunda V,161,1-14 Page: 1448 a \_\_\_\_\_

In fundum Caecina utrum TANDEM noluit an non potuit accedere .  
 \_\_\_\_\_ Cicéron Pro A. Caecina oratio,48,1-9 Page: 1589 a \_\_\_\_\_

Quo usque TANDEM abutere Catilina patientia nostra .  
 \_\_\_\_\_ Cicéron In Catilinam I,1,1-11 Page: 2124 a \_\_\_\_\_

Sed tamen ne me cum his principibus ciuitatis qui adsunt P. Sestio sed ut te cum cum homine uno non solum impudentissimo sed etiam sordidissimo atque infimo conferam de te ipso homine et adrogantissimo et mihi inimicissimo quaero Vatini utrum TANDEM putes huic ciuitati huic rei publicae huic urbi his templis aerario curiae uiris his quos uides horum bonis fortunis liberis ciuibus ceteris denique deorum immortalium delubris auspiciis religionibus melius fuisse et praestabilis me ciuem in hac ciuitate nasci an te .  
 \_\_\_\_\_ Cicéron In Vatinium,10,1-19 Page: 3050 b \_\_\_\_\_

Addo illud etiam qui TANDEM conuenit aut tam graui onere tributorum ad tantam pecuniam cogendam nulle talentum accessionem esse factam aut in tanta mercede hominis ut uis auarissimi mille talentum decessionem esse concessam .  
 \_\_\_\_\_ Cicéron Pro Rabirio Postumo 31,1-11 Page: 3556 a \_\_\_\_\_

*Contextes du lemme tandem intégré dans un syntagme /adv. interrogatif + adverbe + verbe/*

c - sur les CODES grammaticaux qui sont présentés dans un tableau récapitulatif explicite (avec le nom des catégories auxquelles ils renvoient et non pas sous leur seule forme codée).

The screenshot shows a software window titled "G:\LATIN.EXE" with a grid of grammatical categories and their codes. The categories are arranged in columns and rows, with a "Code choisi" field at the bottom left and a list of codes at the bottom right.

Catégorie 1	Sous-catégorie 2	Mode 6	Temps 7	Voix 8	Personne 9
Verbe B	1re conjugaison 1	Indicatif 1	Présent 1	Actif 1	1re personne 1
	2e conjugaison 2	Impératif 2	Imparfait 2	Passif 2	2e personne 2
	3e conjugaison 3	Subjonctif 3	Futur 3	Déponent 3	3e personne 3
	4e conjugaison 4	Participe 4	Parfait 4	Semi-dépon. 4	
	conjugaison mixte 5	Adj.verbal 5	Plus-que-parfait 5		
	conjugaison anormale 6	Gérondif 6	Futur antérieur 6		
		Infinitif 7	Périphrase au parfait 7		
		Supin en -UM 8	Periph.pl-q-parfait 8		
		Supin en -U 9	Périph.futur ant. 9		

Adjectif C	Degré 5	Cas 3	Personnel E
1re classe 1	Positif 1	Nominatif 1	Personnel E
2e classe consonantique 2	Comparatif 2	Vocatif 2	Possessif F
2e classe en -er 3	Superlatif 3	Accusatif 3	Réfléchi G
2e classe en -is 4		Génitif 4	Poss. réfl. H
2e cl.imparisyllabique 5		Datif 5	Démonstr. I
Anomal 6		Ablatif 6	Relatif J
Déclinaison grecque 7		Locatif 7	Interrogatif K
		Indéclinable Z	Indéfini L
		Tous les cas _	Adverbe M
			Adv. relat. N
			Adv. négat. O
			Adv.int.nég Q
			Préposit. R
			Coordinat. S
			Subordinat. T
			Interjection U

Substantif A	Nombre 4	Mise à zéro	Continuer
1re déclinaison 1	Singulier 1		
2e déclinaison 2	Pluriel 2		
3e déclinaison 3	Sing + Plur 3		
4e déclinaison 4			
5e déclinaison 5			
Anomal 6			
Déclinaison grecque 7			

Numéral D	Code choisi
Cardinal 1	1 2 3 4 5 6 7 8 9 10 11
Ordinal 2	
Distributif 3	
Multiplicatif 4	
Adverbe ordinal 5	
Adverbe multiplicatif 6	

Choisir la combinaison de codes souhaitée. Un clic sur une option sert alternativement à activer ou désactiver la sélection. Les colonnes 6 à 11 sont réservées aux verbes (zone bleue) et la colonne 5 (degré) aux adjectifs et aux adverbes. Les colonnes 3 (cas) et 4 (nombre) concernent toutes les parties du discours, sauf les invariables. Veiller à éviter les choix incohérents. Une fois réalisée la sélection, cliquer sur CONTINUER pour la transmettre au traitement en cours. Les boutons bleus donnent accès à la série entière.

Tableau de sélection des catégories grammaticales

Il suffit de cliquer sur les catégories désirées, et avec le niveau de détail souhaité, puis sur le bouton CONTINUER, pour en obtenir les contextes ou la concordance. Une erreur de frappe se rectifie au moyen du bouton MISE A ZERO.

d - sur les STRUCTURES SYNTAXIQUES : à partir de la liste de catégories reproduite ci-dessous, on peut obtenir, par exemple, la concordance de toutes les séquences où un pronom relatif est suivi d'un pronom réfléchi, puis d'un verbe.

1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

Structure choisie **jgb**

Pronom\_relatif Pronom\_réfléchi Verbe

Effacer

Continuer

Substantif a

Pronom\_indéfini l

Verbe b

Adverbe m

Adjectif c

Adverbe\_relatif n

Numéral d

Adverbe\_interrog o

Pronom\_personnel e

Adverbe\_négatif p

Pronom\_possessif f

Adv\_interr\_négatif q

Pronom\_réfléchi g

Préposition r

Pron\_poss\_réfléchi h

Coordination s

Pronom\_démonstr. i

Subordination t

Pronom\_relatif j

Interjection u

Pronom\_interrogatif k

Verbe\_auxiliaire v

## Menu de sélection d'une structure syntaxique

G:\CICERON.EXE

Forme Lemme Code Syntaxe Expr. Initial Final Chain Liste Tout Nb 34 CONCORDANCE  
Clic sur une ligne pour voir la page

SR 225a| n fugiet cum per eius fidem laeditur cui se commiserit . Atque ea sunt animad  
Am 4787a| o quisquam amicus esse poterit eius cui se putabit inimicum esse posse .  
Ve 783a| Nihil dicimus tabulae sunt in medio quae se corruptas atque interlitas esse cla  
P2 4212a| emper ipse laetatus quod ea faceret quae sibi obici ne ab inimico quidem possen  
Fl 2515a| drachmarum CC CCVI milia dedisse . Quantum sibi ablatum homo impudentissimus d  
P1 3970a| Antonius a Brundisio cum eis copiis quas se habiturum putabat Romam ut minabatu  
Fl 2494a| quod uereor ne Laelius ex his rebus quas sibi suscepit gloriae causa putet aliq  
Tu 337a| o M. Tullius paternum recuperatores quem se habere usque eo non moleste tulit d  
Of 5045a| unt digni calamitate . In iis tamen qui se adiuuari uolent non ne adfligantur s  
Pl 3479a| satis facere non possunt non in eis qui se afluere ab istis studiis confitentur  
Mi 3589a| Appia factam esse constaret non eum qui se defendisset contra rem publicam feci  
Ve 634a| ferentur . Quid ait . Se ei daturum qui se dicat heredem esse . Quid ergo  
Of 5198a| s Calliphontem et Dinomachum iudico qui se dirempturos controuersiam putauerunt  
Fl 2527a| m credo significari isto loco illum qui se erigit Heraclidam . Utrum enim es  
Ve 1024a| ab se omnem suspicionem propulsare qui se eripere ex infamia cuperet . Addu  
P2 4216a| uentus est nemo . Hominem ridiculum qui se exserere aere alieno putet posse cum  
Ve 1112a| a uobis postulentur . Absoluite eum qui se fateatur maximas pecunias cum summa  
Va 3078a| ilem tribunum plebis posse reperiri qui se interponat quo minus reus mea lege f  
Of 5006a| in re publica consequemur . Etenim qui se metui uolent a quibus metuentur eosd  
Ca 1591a| putamus eorum esse causa compositum qui se praecipitatos ex locis superioribus  
Ve 529a| auissimo atque ornatissimo M. Annio qui se praesente ciuem Romanum securi perc  
Ca 1632a| nec precario possederit . Itaque is qui se restituisse dixit magna uoce saepe c  
St 2963a| si obnuntiasset Fabricio is praetor qui se seruasse de caelo dixerat accepisset  
Ve 750a| eri quoque impetrant ne retineantur qui se uelle dicebant alterutri eorum qui t  
St 2977a| s focus ferrum flammam que depellit qui sibi licere uult tuto esse in foro in t  
Of 5143a| moliri iussus arbitrum illum adegit quidquid sibi dare facere oporteret ex fide  
Ve 1006a| lem futurum . At quid est tandem in quo se elaborasse dicit . Recita . Ut  
Am 4784a| t eum diligat quem metuat aut eum a quo se metui putet . Coluntur tamen simu  
Se 4741a| olonis quidem sapientis elogium est quo se negat uelle suam mortem dolore amico  
Fr 4526a| hoc animo erga uos illa plebs erit quo se ostendit esse cum legem Aureliam cum  
QR 317a| tegram reliquisset Fanni concedo sed quod sibi exegit id commune societatis fact  
QR 317a| quem ad modum non sibi exegit . Nam quod sibi petitur certe alteri non exigitur  
Qu 27a| Ait se auctionatum esse in Gallia quod sibi uideretur se uendidisse curasse n  
QR 321a| tes . Perstat in sententia Saturius quodcumque sibi petat socius id societatis

## Concordance de la structure /relatif + réfléchi + verbe/

e - sur les DÉBUTS DE MOT, ce qui permet par exemple de trouver tous les mots commençant par un même préfixe, ou de trouver une famille de mots.

f - sur les FINS DE MOT. Noter qu'il s'agit ici nécessairement d'une recherche sur les formes fléchies, ce qui ne permet pas de retrouver tous les lemmes munis d'un même suffixe en raison des variations flexionnelles.

g - sur les COOCCURRENCES (disponible uniquement dans le programme CONTEXTE). Seuls sont restitués les contextes qui contiennent la présence simultanée de deux formes choisies. Par contexte il faut entendre la page originale. La cooccurrence n'est pas toujours vérifiable dans les extraits affichés à l'écran, parce que les extraits respectent la limite du paragraphe et non de la page ; mais elle est contrôlable si l'on clique sur l'extrait : on est alors renvoyé à la page originale, où les deux mots cooccurrents sont affichés en rouge.

h - sur une CHAÎNE DE CARACTÈRES où qu'elle se trouve dans le mot (disponible uniquement dans le programme CONCORDANCE). Ainsi en cherchant la chaîne *ibil* on a un moyen détourné pour trouver tous les adjectifs suffixés en *-ibilis* ; mais les résultats sont légèrement bruités, comme on le voit sur la copie d'écran ci-dessous, avec les deux occurrences du verbe *exsibilo* :



Concordance des formes contenant la chaîne de caractères *ibil*

i - sur une LISTE DE MOTS, *préalablement constituée* à l'aide du bouton LISTE du menu SOMMAIRE.

### 1. Création de listes

Solliciter d'abord le bouton LISTE du menu principal (page d'accueil, parmi les fonctions statistiques), qui adresse l'utilisateur à une page spéciale où divers types de paramètres sont proposés pour créer la liste : DEBUT DE MOT, FIN DE MOT, CHAÎNE DE CARACTÈRES dont l'effet a été décrit ci-dessus (liste des formes commençant ou se terminant par telle chaîne de caractères, ou contenant cette chaîne n'importe où dans le mot). Le contenu de la liste peut aussi être importé depuis un fichier extérieur (bouton FICHER), pourvu qu'il ait été enregistré en format « texte seul » dans le répertoire C:\HYPERBAS. Le remplissage de la liste peut encore être libre et saisi manuellement, forme après forme ou lemme après lemme (boutons FORME et LEMME). L'utilisateur fournit alors autant de mots qu'il veut, en signifiant à la machine par un mot vide (OK sur formulaire blanc) que la liste est close. Le bouton LETTRES, quant à lui, dresse la liste des lettres constitutives de l'alphabet latin et compte leur nombre d'occurrences dans chacun des textes du corpus au moyen d'un parcours du dictionnaire de la base qui prend un peu plus de temps que les autres opérations.

On peut également créer des listes, non pas de mots ou de lettres, mais de catégories grammaticales : le bouton CATÉGORIES permet d'avoir accès à des listes préétablies de termes répondant à un même critère de définition qui nous a paru utile : la liste des coordonnants par exemple, ou celle des négations, des prépositions, etc. Le choix des termes retenus pour figurer dans ces listes peut évidemment prêter à discussion, mais, comme on le verra au § 2 ci-dessous, ceux qui ne conviennent pas peuvent aisément être éliminés ; et ceux qui manquent peuvent être ajoutés grâce au bouton FORME et à l'opération AJOUTER à la liste existante. SYNTAXE ouvre la liste présentée au § d ci-dessus et permet de constituer la liste de tous les syntagmes répondant aux critères de sélection retenus (par ex. la liste de tous les syntagmes formés d'une conjonction de coordination suivie d'une préposition) ; en général ce genre de requête donne des résultats très volumineux dont le détail ne peut pas être affiché ; il convient alors de choisir, au moment de l'affichage, l'option CUMUL qui donnera l'effectif global de ladite structure dans chaque texte, sans se soucier des formes précises qui l'incarnent. De même le bouton CODE donne accès au tableau présenté ci-dessus § c et permet de dénombrer, par exemple, tous les subjonctifs présents, auxquels on peut ensuite ajouter le décompte des subjonctifs imparfaits, etc., créant ainsi la liste de tous les temps du subjonctif. Le bouton CONSTRUCTION affiche la liste de tous les types de propositions (propositions principales, subjonctifs paratactiques, ablatifs absolus, infinitives et diverses subordonnées conjonctives).

Enfin quelques-uns des modules de ce menu permettent de créer des listes à partir de critères de sélection numériques dont la gestion est entièrement laissée au programme informatique : ainsi le bouton FRÉQUENCE dresse la liste des formes dont la fréquence dans le corpus dépasse un seuil donné (là encore l'option CUMUL à l'affichage est recommandée). Le bouton GROUPE effectue à peu près le même travail, mais en regroupant toutes les formes du corpus par classes de fréquence (11 classes ont été arbitrairement retenues, le groupe 1 rassemblant les hapax et le groupe 11 les mots les plus fréquents). Le bouton LONGUEUR, pour sa part, classe les formes en fonction du nombre de leurs caractères (lg1 = formes d'une seule lettre, lg2 = formes de deux lettres, ..., lg11 = formes de 11 lettres et plus).

## 2. Modification d'une liste

Toute nouvelle liste créée peut être ajoutée à la précédente ou, au contraire, la remplacer.

Pour supprimer un élément d'une liste il suffit de cliquer simultanément sur la touche MAJUSCULE du clavier et sur la souris pointée vers cet élément. Pour supprimer toute la liste, il suffit de cliquer sur l'icône représentant une poubelle.

On a aussi la ressource d'utiliser le bouton MODIF qui permet divers regroupements de lignes ou de colonnes. Quand on le sollicite, le programme demande d'abord si l'on souhaite modifier les éléments de la liste (en lignes) ou les textes (en colonnes). Puis il propose deux opérations : l'addition ou la soustraction. Pour les lignes, la première opération consiste à regrouper les effectifs de plusieurs lignes *successives* dans la liste : on peut obtenir le total depuis la première à la dernière ligne ou la somme de quelques lignes seulement. Ainsi, en répondant pas à pas aux demandes du programme MODIF, on peut regrouper dans la liste ci-dessous *antea* et *antehac* d'une part, *etiamtum* et *etiamtunc* d'autre part. Pour les colonnes, le programme est plus souple puisqu'il autorise le regroupement de colonnes non consécutives ; dans l'exemple choisi, on a ainsi regroupé les effectifs de Plaute et des Tragédies de Sénèque dans une colonne supplémentaire THEATRE située à la fin. Veiller toutefois à ne pas englober tous les textes dans le regroupement, car c'est précisément le total des colonnes qui sert de référence pour le calcul des écarts (cf. pp. 25-26).

The screenshot shows the 'Liste de mots' window in the LATIN.EXE application. The window title is 'G:\LATIN.EXE'. The interface includes a menu bar with options like 'Finale', 'Initiale', 'Chaine', 'Fichier', 'Fréq.', 'Long.', 'Groupe', and 'Catég.'. Below the menu bar, there are several icons and a toolbar. The main area displays a list of Latin adverbs and their corresponding frequency counts across various categories. The categories are: Plau, Cato, Lucr, Catu, Gaul, Civi, 1\_Di, 2\_Ve, 3\_Di, 4\_Di, 5\_Di, 6\_Di, Phil, Trai, Sall, Géor, Enéi, Hora, Tibu, Prop, 1\_Ov, 2\_Ov, Quin, Cons, 1\_Lu, 2\_Lu, Bien, Colè, Trai, Trag, Pétr, Juvé, Plin, Mine, and Hist. The adverbs listed include ABHINC, ADHVC, ALIAS, ALIQVAMDI, ALIQVANDO, ANTE1, ANTEA, ANTEHAC, DEINDE, DENIQUE, DIV, ETIAMNVNC, ETIAMTVM, ETIAMTVNC, FERRE, HERI, HODIE, IAM, IAMIAM, IAMDIV, IAMPRIDEM, INTEREA, INTERDIV, INTERDVM, INTERIM, MODO1, NOCTE, NOCTV, NONDVM, NVNC, and NVMQVAM. The frequency counts are displayed in a grid format.

	Plau	Cato	Lucr	Catu	Gaul	Civi	1_Di	2_Ve	3_Di	4_Di	5_Di	6_Di	Phil	Trai	Sall	Géor	Enéi
ABHINC	2	0	1	0	0	0	0	2	3	0	0	1	0	1	0	0	0
ADHVC	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ALIAS	0	0	0	0	0	11	ABHINC										
ALIQVAMDI	8	0	4	0	1	3	7	21	14	13	5	6	12	4	7		
ALIQVANDO	1	11	11	0	0	18	14	62	8	53	31	6	8	17	34	30	
ANTE1	13	13	18	36	47	536	ADHVC										
ANTEA	3	0	2	0	8	3	0	2	0	0	1	1	0	6	1		
ANTEHAC	1	0	4	1	0	0	1	24	0	2	6	4	4	4	0	0	
DEINDE	0	0	1	2	12	93	ALIAS										
DENIQUE	0	0	0	0	0	2	2	0	1	1	1	0	0	1	1	1	
DIV	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	
ETIAMNVNC	0	0	0	0	0	12	ALIQVAMDI										
ETIAMTVM	0	0	0	0	2	1	11	32	16	22	39	23	21	18	6		
ETIAMTVNC	0	2	0	0	0	1	1	2	6	61	18	26	17	34	5	8	
FERRE	3	10	6	4	17	412	ALIQVANDO										
HERI	9	3	71	7	36	22	13	41	23	26	35	21	37	37	25		
HODIE	17	33	7	14	13	51	30	54	6	12	8	9	3	2	13	17	
IAM	9	20	11	10	34	779	ANTE1										
IAMIAM	0	0	1	0	1	9	5	20	66	37	24	28	11	12	6	35	
IAMDIV	0	0	1	0	0	0	0	23	0	0	0	0	0	0	0	1	
IAMPRIDEM	0	8	0	5	13	306	ANTEA										
INTEREA	7	0	2	0	0	0	0	0	1	0	0	0	0	0	1	1	
INTERDIV	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
INTERDVM	0	0	0	0	3	16	ANTEHAC										
INTERIM	9	22	18	7	7	3	23	74	44	46	49	21	20	28	96		
MODO1	7	31	5	1	8	5	1	205	19	75	70	62	37	32	6	38	
NOCTE	13	10	10	57	119	1278	DEINDE										
NOCTV	6	6	1	96	4	5	5	29	78	46	90	96	45	25	8	29	
NONDVM	6	4	15	1	3	18	5	3	8	9	19	13	6	4	2	1	
NVNC	0	11	7	10	17	725	DENIQUE										
NVMQVAM	10	9	7	7	19	6	13	14	17	10	21	9	12	16	8		

Liste des adverbes de temps dans la base LATIN.EXE

Finale	Initiale	Chaîne	Fichier	Fréq.	Long.	Groupe	Catég.									
Forme	Lemme	Syntaxe	Constr.	Codes	Bicode	Tricode	Lettres									
Plau Cato Lucr Catu Gaul Civi 1_Di 2_Ve 3_Di 4_Di 5_Di 6_Di Phil Trai Sall Géor Enéi																
Hora Tibu Prop 1_Ov 2_Ov Quin Cons 1_Lu 2_Lu Bien Colè Trai Trag Pétr Juvé Plin Mine																
Hist Anna THEA																
ABHINC	2	0	1	0	0	0	2	3	0	0	1	0	1	0	0	
ADHVC	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
ALIAS	0	0	0	0	0	2	,	11	ABHINC							
ALIQVAMDI	8	0	4	0	1	3	7	21	14	13	5	6	12	4	7	
ALIQVANDO	1	11	11	0	0	18	14	62	8	53	31	6	8	17	34	30
ANTE1	13	13	18	36	47	42	,	536	ADHVC							
antea/hac	3	0	2	0	8	3	0	2	0	0	1	1	0	6	1	
DEINDE	1	0	4	1	0	0	1	24	0	2	6	4	4	4	0	0
DENIQVE	0	0	1	2	12	3	,	93	ALIAS							
DIV	0	0	0	0	0	2	2	0	1	1	1	0	0	1	1	1
ETIAMNVNC	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
etiamtu-	0	0	0	0	0	0	,	12	ALIQVAMDI							
FERE	0	0	0	0	0	2	1	11	32	16	22	39	23	21	18	6
HERI	0	2	0	0	0	1	1	2	6	61	18	26	17	34	5	8
HODIE	3	10	6	4	17	5	,	412	ALIQVANDO							
IAM	9	3	71	7	36	22	13	41	23	26	35	21	37	37	25	
IAMIAM	17	33	7	14	13	51	30	54	6	12	8	9	3	2	13	17
IAMDIV	9	20	11	10	34	22	,	779	ANTE1							
IAMPRIDEM	7	1	2	1	9	5	20	67	37	24	28	11	12	7	36	
INTEREA	0	0	2	0	0	0	0	23	0	0	0	0	0	0	1	
INTERDIV	0	8	0	5	16	7	,	322	antea/hac							
INTERDVM	9	22	18	7	7	3	23	74	44	46	49	21	20	28	96	
INTERIM	7	31	5	1	8	5	1	205	19	75	70	62	37	32	6	38
MODO1	13	10	10	57	119	15	,	1278	DEINDE							
NOCTE	6	1	96	4	5	5	29	78	46	90	96	45	25	8	29	
NOCTV	6	4	15	1	3	18	5	3	8	9	19	13	6	4	2	1
NONDVM	0	11	7	10	17	8	,	725	DENIQVE							
NVNC	10	9	7	7	19	6	13	14	17	10	21	9	12	16	8	
NVMQVAM	1	10	12	1	6	19	13	44	22	51	30	19	9	33	25	37
NVPER	6	3	8	22	48	35	,	597	DIV							
OLIM	7	0	3	0	2	0	6	5	1	8	3	2	2	1	3	

Liste précédente modifiée par regroupement de lignes et colonnes

Le même bouton MODIF permet également l'opération de soustraction. On peut ainsi envisager de constituer, par exemple au moyen du bouton CODES, la liste cumulée de toutes les occurrences de substantifs (ligne 1), la liste de toutes les formes de nominatif (ligne 2) et celle de toutes les formes de vocatif (ligne 3) :

Finale	Initiale	Chaîne	Fichier	Fréq.	Long.	Groupe	Catég.									
Forme	Lemme	Syntaxe	Constr.	Codes	Bicode	Tricode	Lettres									
Plau Cato Lucr Catu Gaul Civi 1_Di 2_Ve 3_Di 4_Di 5_Di 6_Di Phil Trai Sall Géor Enéi																
Hora Tibu Prop 1_Ov 2_Ov Quin Cons 1_Lu 2_Lu Bien Colè Trai Trag Pétr Juvé Plin Mine																
Hist Anna																
-_A_	11317	4846	13507	3645	14355	10856	6250	26801	13030	17246	19792	11665	14503	12282	11292	
-_A_1_	694123372	134623712	7942	19498	10934	22293	4961	15007	10455	9224	5304	10241	20230	8666		
-_A_2_	80354789	68872015833308	,	456806	-_A_											
	2515	617	3334	826	2106	1794	1304	4925	2548	3246	3752	2360	2939	3056	2480	
	187056123583	1105	2346	6214	3336	4590	1141	4337	3087	2300	1440	2546	5756	2070		
	225211641584	4400	6451	,	104986	-_A_1_										
	670	23	27	368	4	4	235	382	226	393	299	303	265	77	72	
	222460	323	136	302												
	87	65	21	18	38											
											41	138				

Repérer le numéro du mot ou code dont vous voulez réduire l'effectif:  
 1 -\_A\_-, 2 -\_A\_1\_-, 3 -\_A\_2\_-.

OK

Puis on va soustraire aux effectifs de la ligne 1 ceux de la ligne 2 et ceux de la ligne 3. On obtiendra alors en ligne 1 le décompte des seuls cas régimes :

	Finale	Initiale	Chaîne	Fichier	Fréq.	Long.	Groupe	Catég.												
	Forme	Lemme	Syntaxe	Constr.	Codes	Bicode	Tricode	Lettres												
Plau Cato Lucr Catu Gaul Civi 1_Di 2_Ve 3_Di 4_Di 5_Di 6_Di Phil Trai Sall Géor Enéi																				
Hora Tibu Prop 1_Ov 2_Ov Quin Cons 1_Lu 2_Lu Bien Colè Trai Trag Pétr Juvé Plin Mine																				
Hist Anna																				
-_A_-	8132	4206	10146	2451	12245	9058	4711	21494	10256	13607	15741	9002	11299	9149	8740					
-_A_1_-	4849	17300	9556	2471	5294	12797	7273	17651	3785	10586	7331	6891	3854	7641	13733	6458				
-_A_2_-	5696	3560	5282	1574	26819		3448	04												
	2515	617	3334	826	2106	1794	1304	4925	2548	3246	3752	2360	2939	3056	2480					
	1870	5612	3583	1105	2346	6214	3336	4590	1141	4337	3087	2300	1440	2546	5756	2070				
	2252	1164	1584	4400	6451		1049	86												
	670	23	27	368	4	4	235	382	226	393	299	303	265	77	72					
	222	460	323	136	302	487	325	52	35	84	37	33	10	54	741	138				
	87	65	21	18	38		7016													

Liste des substantifs modifiée par soustraction des nominatifs et vocatifs

Enfin le bouton TRIER permet de redisposer les lignes dans l'ordre alphabétique, ou dans l'ordre hiérarchique (par fréquences décroissantes), ou dans l'ordre initial de saisie.

### 3. Exploitation d'une liste

Une fois constituée, la liste peut être soumise d'un coup aux programmes de concordance ou de recherche en contexte, mais aussi à l'analyse factorielle (voir p. 31) ou à la représentation graphique, en sorte que son statut relève davantage des fonctions statistiques que nous allons examiner dans le chapitre suivant.

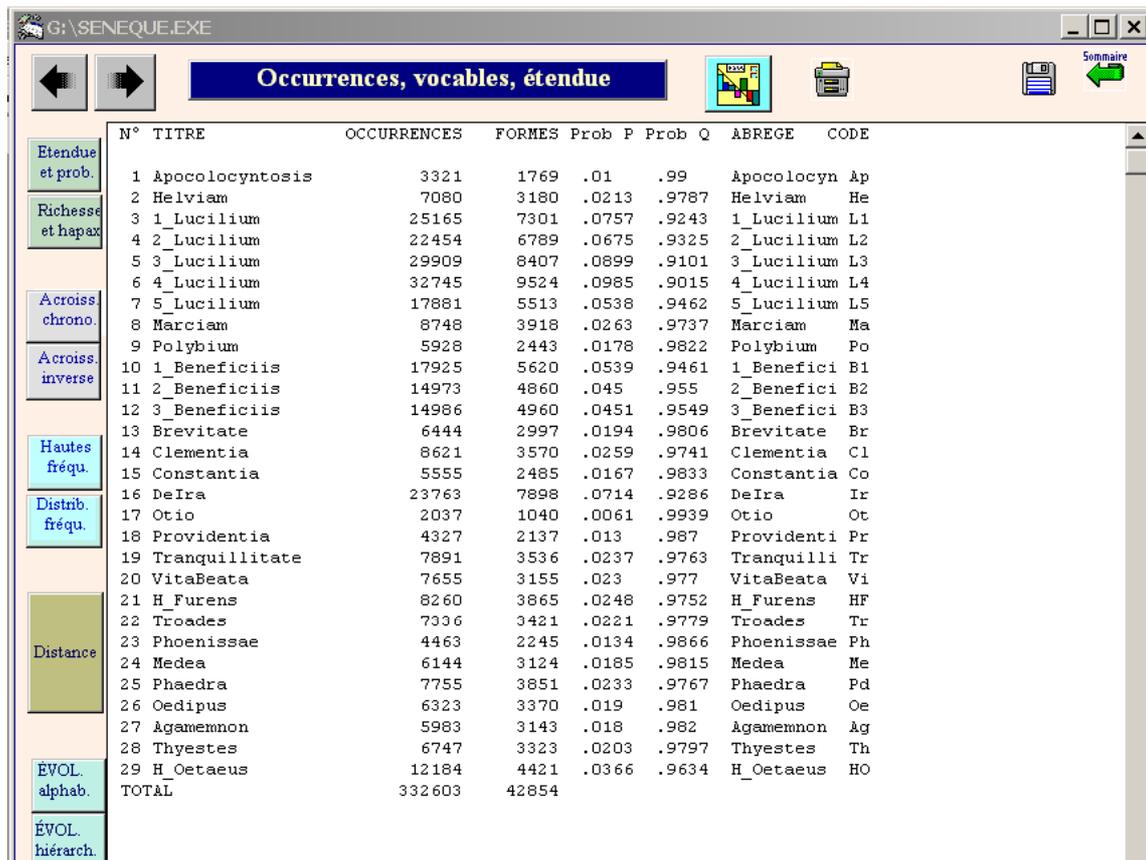
Ces traitements quantitatifs utilisent d'une part le tableau des fréquences observées pour chaque terme de la liste dans les divers textes du corpus, d'autre part les écarts que ces fréquences manifestent par rapport aux fréquences théoriquement attendues dans le cas d'une distribution aléatoire (*i.e.* si la distribution des termes de la liste était tout à fait indifférente aux genres littéraires, aux thèmes des œuvres, aux styles des écrivains, aux époques, etc. ; voir pp. 25-26). Les valeurs de ces écarts apparaissent dans le tableau quand on sollicite le premier bouton à gauche de la barre de menu (ECART↔FRÉQU)

## CHAPITRE 3

### *L'exploitation statistique*

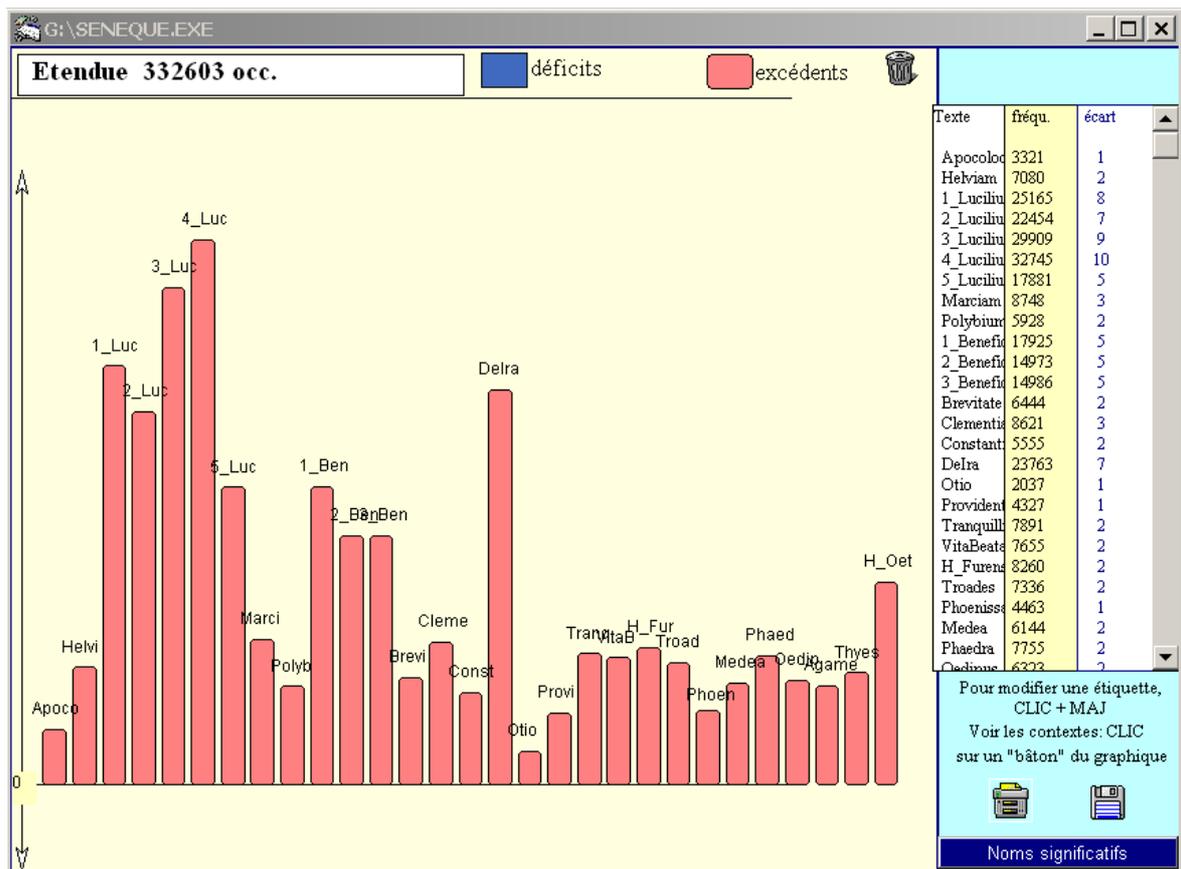
#### PARTITION et STATISTIQUE

Les calculs s'appuient sur les lois classiques de la statistique linguistique, principalement la loi normale et la loi binomiale. Les probabilités p et q qu'on lit dans la distribution ci-dessous servent à tous les calculs de pondération. Si l'on souhaite vérifier ces calculs et contrôler les tests statistiques, l'étendue et les caractéristiques de chaque texte sont montrées quand on sollicite le bouton DISTRIBUTION (écran d'accueil), puis ETENDUE ET PROB, procurant des résultats soit sous forme de liste, soit sous forme de graphique.



N°	TITRE	OCCURRENCES	FORMES	Prob P	Prob Q	ABREGE	CODE
1	Apocolocyntosis	3321	1769	.01	.99	Apocolocyn	Ap
2	Helviam	7080	3180	.0213	.9787	Helviam	He
3	1_Lucilium	25165	7301	.0757	.9243	1_Lucilium	L1
4	2_Lucilium	22454	6789	.0675	.9325	2_Lucilium	L2
5	3_Lucilium	29909	8407	.0899	.9101	3_Lucilium	L3
6	4_Lucilium	32745	9524	.0985	.9015	4_Lucilium	L4
7	5_Lucilium	17881	5513	.0538	.9462	5_Lucilium	L5
8	Marciam	8748	3918	.0263	.9737	Marciam	Ma
9	Polybium	5928	2443	.0178	.9822	Polybium	Po
10	1_Beneficiis	17925	5620	.0539	.9461	1_Benefici	B1
11	2_Beneficiis	14973	4860	.045	.955	2_Benefici	B2
12	3_Beneficiis	14986	4960	.0451	.9549	3_Benefici	B3
13	Brevitate	6444	2997	.0194	.9806	Brevitate	Br
14	Clementia	8621	3570	.0259	.9741	Clementia	C1
15	Constantia	5555	2485	.0167	.9833	Constantia	Co
16	DeIra	23763	7898	.0714	.9286	DeIra	Ir
17	Otio	2037	1040	.0061	.9939	Otio	Ot
18	Providentia	4327	2137	.013	.987	Providenti	Pr
19	Tranquillitate	7891	3536	.0237	.9763	Tranquilli	Tr
20	VitaBeata	7655	3155	.023	.977	VitaBeata	Vi
21	H_Furens	8260	3865	.0248	.9752	H_Furens	HF
22	Troades	7336	3421	.0221	.9779	Troades	Tr
23	Phoenissae	4463	2245	.0134	.9866	Phoenissae	Ph
24	Medea	6144	3124	.0185	.9815	Medea	Me
25	Phaedra	7755	3851	.0233	.9767	Phaedra	Pd
26	Oedipus	6323	3370	.019	.981	Oedipus	Oe
27	Agamemnon	5983	3143	.018	.982	Agamemnon	Ag
28	Thyestes	6747	3323	.0203	.9797	Thyestes	Th
29	H_Oetaeus	12184	4421	.0366	.9634	H_Oetaeus	HO
	TOTAL	332603	42854				

*Étendue relative des textes de la base SENEQUE.EXE*



Graphique symbolisant l'étendue relative des textes de la base

On remarquera que la taille des textes est mesurée d'une part par le nombre d'occurrences graphiques (*i.e.* de mots au sens trivial du terme) qui en constituent la chaîne linéaire, d'autre part par le nombre de formes différentes utilisées dans le texte sans tenir compte de la fréquence de chacune de ces formes (il s'agit bien ici des formes fléchies et non pas des lemmes). Bien sûr, comme de nombreuses formes sont communes à plusieurs textes, le total des formes constitutives de l'ensemble du corpus n'est pas égal à la somme des formes constitutives de chaque texte : il lui est bien inférieur, les formes communes n'étant comptabilisées qu'une fois.

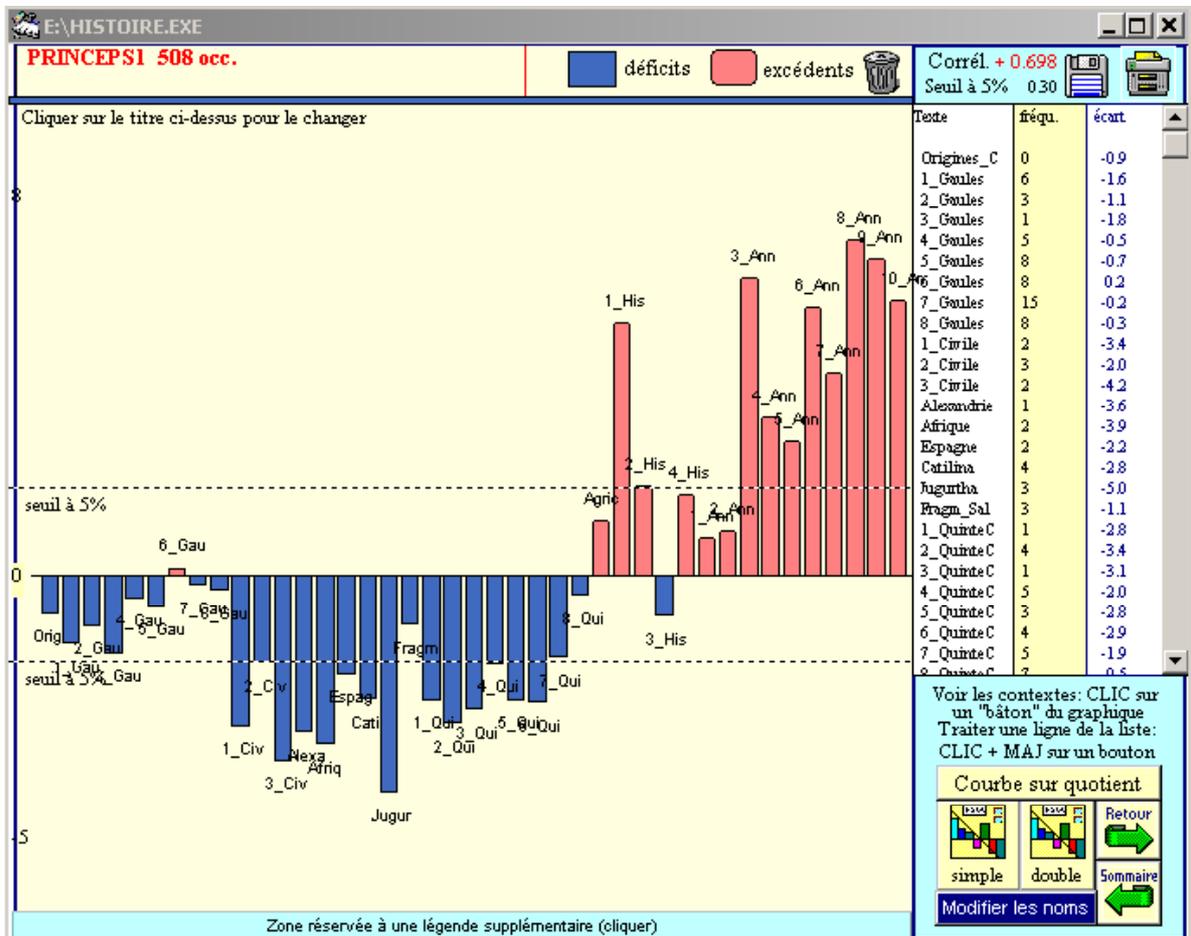
## GRAPHIQUES

Le bouton GRAPHIQUE du menu principal utilise ces probabilités pour établir des excédents ou des déficits d'emploi (calculés sur les occurrences) et les représenter graphiquement. La distribution d'un mot est rarement régulière à travers un corpus et des écarts s'y observent entre la **fréquence observée** d'un mot dans un texte et la **fréquence théorique** qu'on était en droit d'attendre, vu la proportion du texte dans l'ensemble du corpus, et qui s'établit avec une simple règle de trois (fréquence théorique d'un mot dans un texte = fréquence du mot dans le corpus pondérée par la probabilité  $p$  ou part du texte dans le corpus<sup>4</sup>). Cet écart constaté, dit écart absolu, doit ensuite être rapporté à la taille du texte

<sup>4</sup> Par exemple, on voit dans la figure de la page précédente que le texte *Apocolocytosis* représente 1% du corpus de Sénèque.

et du corpus. Dans sa forme la plus simple, le calcul pondère cet écart selon la formule de *l'écart réduit*<sup>5</sup>

Une fois calculés les écarts réduits, le programme présente une illustration graphique de la distribution de l'objet demandé (forme, lemme, code grammatical ou structure syntaxique), sous forme d'histogramme comme ci-dessous pour le lemme *princeps* dans la base des historiens.



*Histogramme du lemme princeps dans HISTOIRE.EXE*

Les « bâtons » de l'histogramme se répartissent de part et d'autre de la ligne médiane qui représente la valeur 0 de l'écart réduit. Chacun de ces « bâtons » est explicité par le titre du texte correspondant. Les effectifs absolus de l'objet dont on a demandé la distribution sont détaillés sur la marge droite, la colonne voisine détaillant les écarts réduits qui servent d'ordonnées à la représentation graphique<sup>6</sup>.

⚠ Veiller à contrôler la valeur de ces écarts réduits, afin de s'assurer qu'on peut ajouter foi aux différences constatées. Si ces valeurs restent à l'intérieur de la plage  $-2$   $+2$ , l'hypothèse nulle ne peut être écartée et les écarts peuvent être considérés comme pouvant s'expliquer par le hasard. Cette précaution s'impose particulièrement lorsque les effectifs sont faibles et qu'on représente la distribution d'un mot peu fréquent.

<sup>5</sup> Pour une explication plus détaillée et illustrée d'exemples simples, voir É. Évrard & S. Mellet (1998) : « Méthodes quantitative en langues anciennes », *LALIES* 18 : 111-155, en part. 128-130.

<sup>6</sup> L'indice de corrélation affiché au-dessus de ces deux colonnes sera commenté plus loin.

⚠ Noter que l'échelle des ordonnées est variable d'un graphique à l'autre (sauf si deux graphiques sont superposés) et que tout l'espace du graphique est occupé, même si l'écart est faible en valeur absolue. Les lignes pointillées symbolisent le seuil au-delà duquel l'écart positif ou négatif est significatif et en deçà duquel il ne l'est pas (seuil à 5% = seuil en-deçà duquel il y a plus de 5 chances sur cent que l'écart soit aléatoire et ne puisse donc pas être commenté de manière fiable).

Quand un grand nombre de textes est représenté dans le même graphique, l'encombrement peut gêner la lisibilité, les zones qui désignent chaque texte se recouvrant. Plusieurs moyens permettent d'y voir clair.

Le premier consiste à effacer ces zones pour ne laisser place qu'aux bâtons de l'histogramme. On activera alors le bouton « Modifier les noms », situé en bas de l'écran à droite. Et pour expliciter la lecture, on remplira à sa guise la légende prévue en bas de l'écran dans la zone turquoise. Un clic à cet endroit suffit à permettre l'écriture.

Le second artifice est d'accepter les autres propositions du même bouton, soit « Rétablir noms longs », soit « Rétablir noms courts », soit surtout « Noms significatifs ». Dans ce dernier cas, les textes situés près de la ligne médiane ne sont pas étiquetés, puisque les écarts de distribution qu'on y observe ne sont pas significatifs. Seuls sont expressément nommés les textes qui présentent un écart marqué relativement à la distribution du mot étudié (déficit ou excédent).

Reste enfin la possibilité de faire apparaître ou disparaître alternativement et individuellement le nom de chaque texte représenté, en cliquant sur la zone qui lui est dévolue, au dessus du bâton correspondant, s'il s'agit d'un excédent, au dessous si l'on a affaire à un déficit.

En haut à droite on voit s'afficher un indice de corrélation. Dans le cas d'un graphique simple, cet indice évalue l'impact de la chronologie sur l'évolution du terme étudié. Une courbe de distribution qui augmente régulièrement et sensiblement au fil du temps<sup>7</sup> est signe d'une corrélation chronologique positive ; une courbe qui diminue de la même façon au fil du temps est signe d'une corrélation chronologique négative. Plutôt que de confier à l'œil l'évaluation de cette corrélation, l'outil informatique calcule un indice de corrélation (coefficient de Bravais-Pearson) qui varie entre -1 et +1 (corrélation négative absolue et corrélation positive absolue). L'absence totale de corrélation est exprimée par l'indice 0. Le seuil de pertinence ou de significativité (dit seuil à 5%) est signalé pour chaque graphique. Dans le cas choisi ici en exemple, la corrélation chronologique positive ne laisse aucun doute (indice de 0,683 alors que le seuil de pertinence est de 0,30).

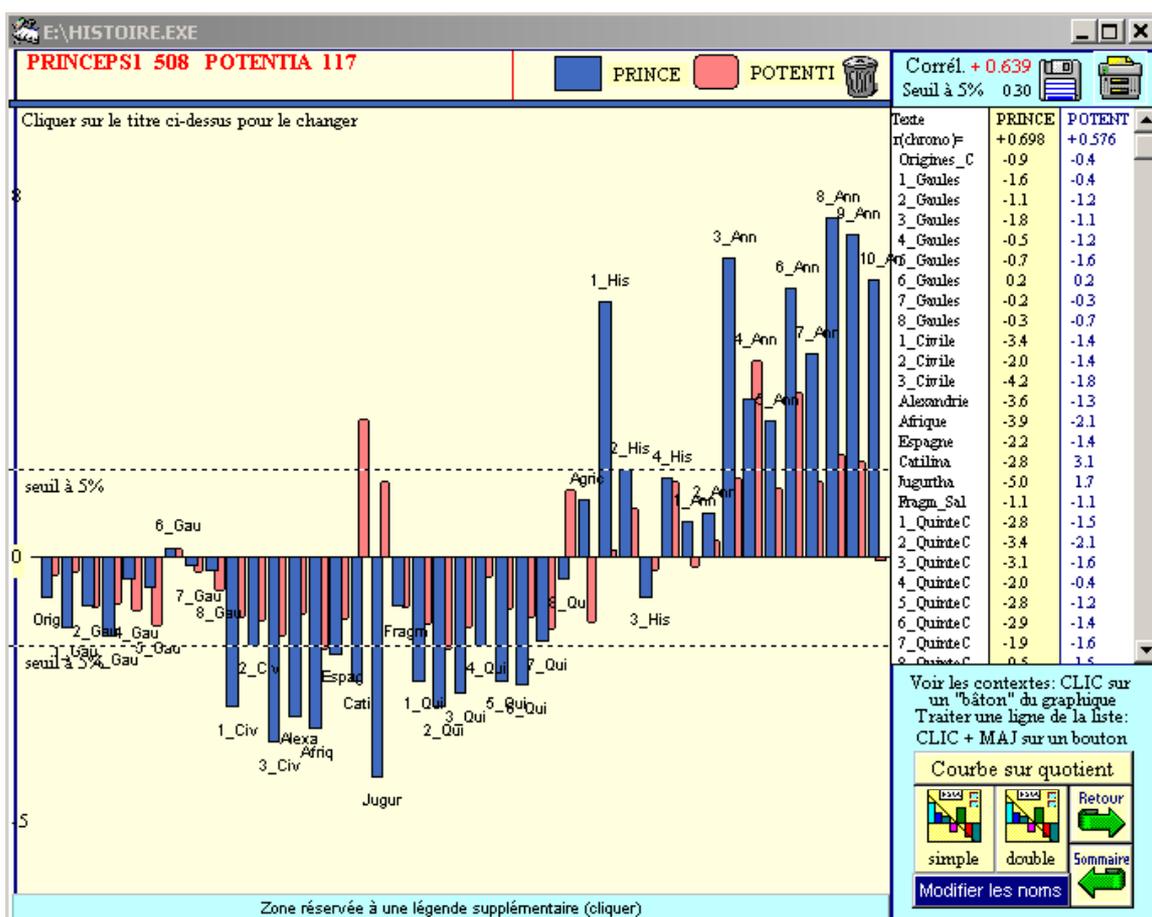
Le bouton DOUBLE est destiné à superposer une seconde distribution à la première et à représenter deux séries sur le même graphique. Dans ce cas, les deux séries d'écarts réduits seront visibles sur la marge droite de l'écran. Le symbolisme des couleurs prend alors une autre signification : le bleu est réservé à la première série (dans l'exemple ci-dessous *princeps*), le rouge à la seconde (ci-dessous *potentia*).

Cette fois-ci l'indice de corrélation évalue le parallélisme de distribution des deux termes étudiés : si l'emploi des deux termes subit la même évolution et que – même sans évolution chronologique notable – ils sont répartis de façon similaire chez les différents auteurs, la corrélation sera positive ; si, au contraire, là où l'un d'eux est en excédent l'autre tend à être en déficit et inversement, alors la corrélation sera négative. Si les deux termes semblent indifférents l'un à l'autre, ne manifestant ni attraction ni répulsion, la corrélation tendra vers zéro. Dans l'exemple ci-dessous, le parallélisme des deux distributions est assez

---

<sup>7</sup> On retrouve là l'importance du classement chronologique des œuvres signalée dès la page 3.

évident, le coefficient de corrélation se situant de nouveau bien au-dessus du seuil requis. Seul Salluste oppose systématiquement et de façon significative l'emploi des deux mots.

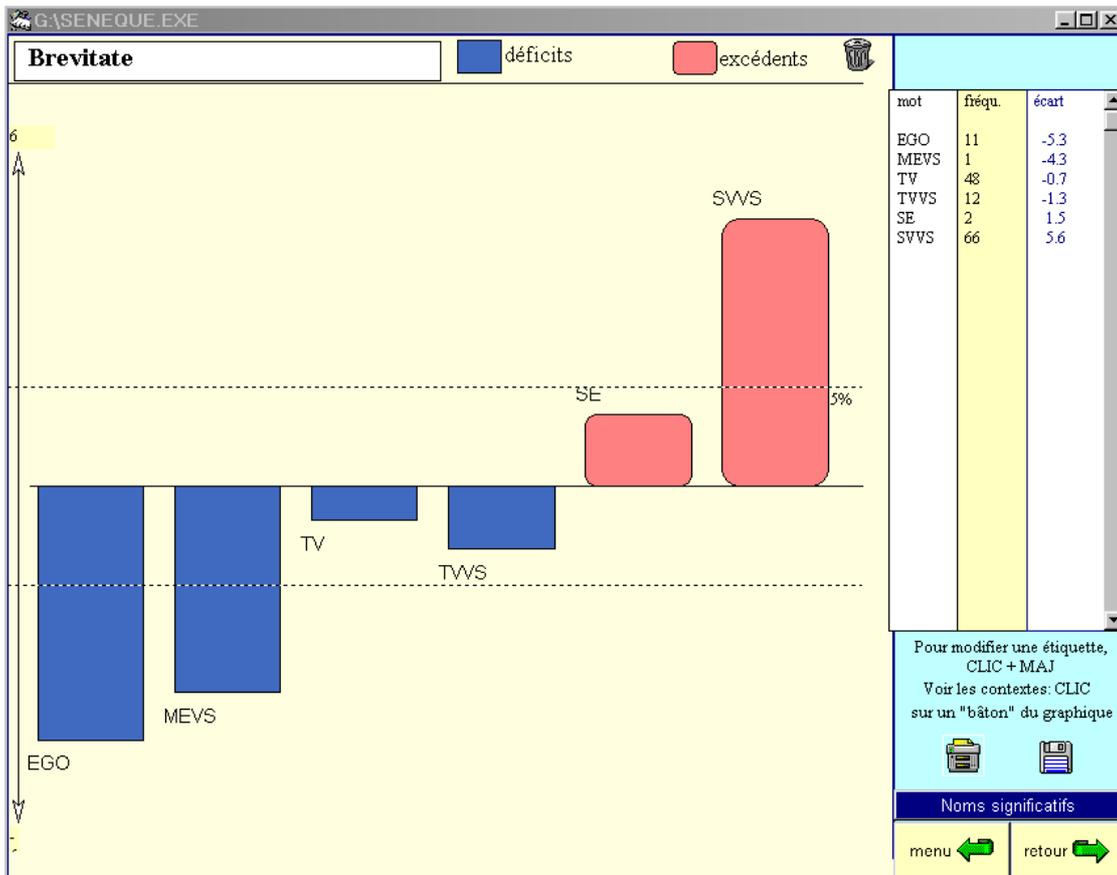


*Distribution comparée des lemmes princeps et potentia dans HISTOIRE.EXE*

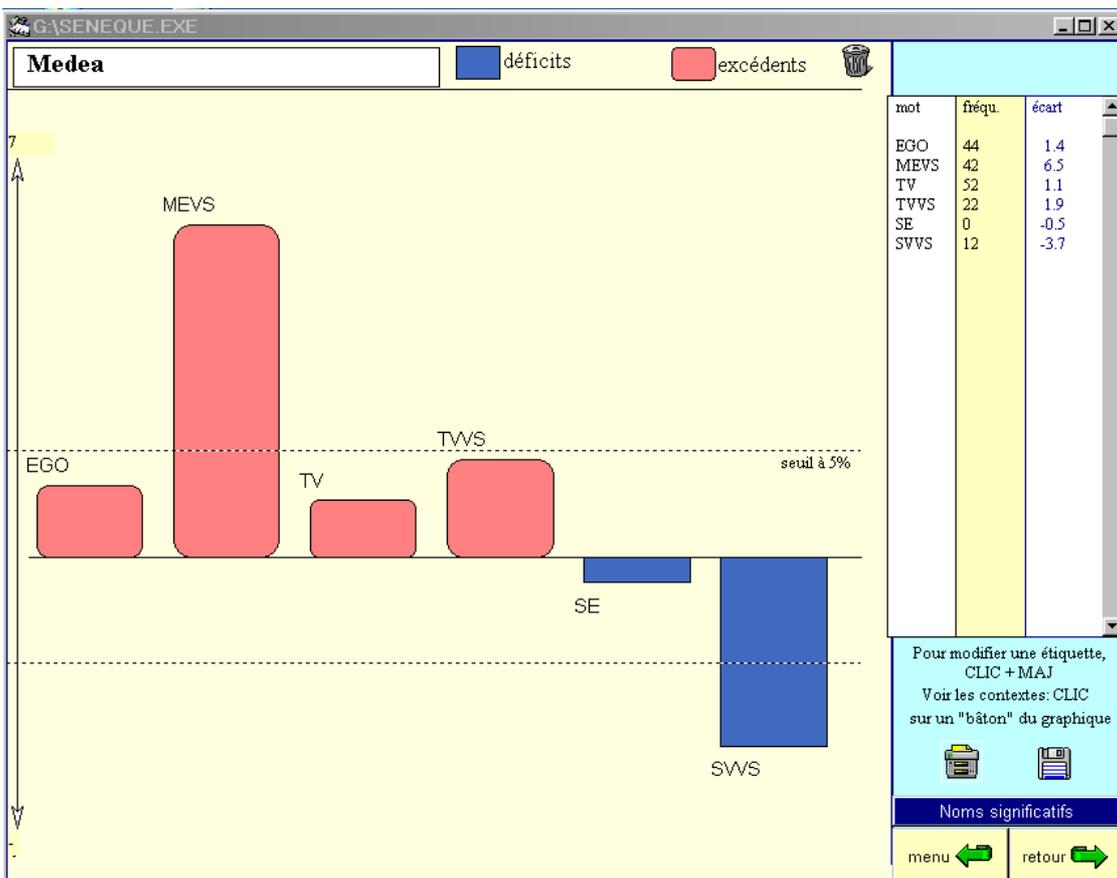
Le bouton GRAPHIQUE est donc accessible directement depuis l'écran d'accueil, laissant le choix du terme dont on veut visualiser la distribution. Mais on le retrouve aussi sur d'autres écrans, notamment dans les pages du dictionnaire (accessible soit par le bouton portant l'icône correspondante, soit en cliquant directement sur l'un des mots affichés) et dans la page consacrée à la constitution et à l'exploitation d'une liste (accessible en cliquant sur une des lignes ou une des colonnes du tableau).

⚠ Pour obtenir un double graphique à partir des termes d'une liste, presser simultanément la touche MAJUSCULE du clavier le clic gauche de la souris en pointant sur le bouton DOUBLE.

Enfin, toujours à l'intérieur de la page consacrée à une liste, le programme d'illustration graphique offre la possibilité de représenter les éléments caractéristiques d'une colonne dans un tableau où généralement, rappelons-le, les colonnes désignent des textes ; il s'agit donc non plus de suivre la distribution d'un mot ou d'une catégorie grammaticale à travers les textes du corpus, mais de dresser le profil d'un texte (ou d'un groupe de textes) à travers les faits linguistiques qu'on a soumis au dénombrement. Comme précédemment, c'est l'écart réduit qui sert d'ordonnée à l'histogramme. On prendra garde cette fois-ci à ignorer le coefficient de corrélation qui, dans cette perspective, n'a guère de signification. Dans le cas le plus fréquent en effet, les éléments des lignes sont rangés de manière arbitraire. Mais si l'on dispose ces éléments dans un ordre logique, comme dans l'exemple ci-dessous où l'on passe de la première à la deuxième puis à la troisième personne, l'organisation du graphique retrouve quelque efficacité.



*Profil du de Brevitate au regard de l'emploi des lemmes ego, meus, tu, tuus, se, suus*



*Profil de Médée au regard de l'emploi des lemmes ego, meus, tu, tuus, se, suus*

## L'ENVIRONNEMENT THÉMATIQUE

Dans la recherche thématique on ne recherche plus une accointance entre un mot et un texte, mais une relation privilégiée entre les mots eux-mêmes. La procédure est appliquée à l'ensemble indéfini de tous les mots qui peuvent se trouver dans l'entourage d'un mot (ou d'un groupe de mots) qu'on définit comme étant le pôle.

Cette procédure se trouve dans la page CONTEXTE. On la déclenche par un clic sur le bouton THEME (bouton rouge en haut à droite de l'écran). Elle prend en considération la liste des contextes recensés pour une recherche donnée, considérés comme un sous-corpus temporaire qui sera comparé au corpus global de la base.

Tous les mots trouvés dans le sous-ensemble des paragraphes constituant le contexte du mot-pôle (sauf ceux qui appartiennent aux références) sont triés et leurs occurrences dans le sous-corpus considéré sont dénombrées. Chaque fréquence observée est alors confrontée à la fréquence théorique attendue si le sous-corpus était un échantillon purement aléatoire de la base ; la même méthode de calcul des écarts réduits permet alors de repérer les mots qui, dans ces paragraphes, sont en excédent ou déficit significatif. Le critère de sélection de ces paragraphes constitués en sous-corpus étant la présence en leur sein du mot-pôle, on en conclut que les mots en excédent sont ceux qui, pour une raison ou une autre, sont attirés par ce dernier, attraction qui s'apparente souvent à une relation thématique (mais pas exclusivement : les phénomènes morpho-syntaxiques jouent aussi un rôle non négligeable).

HIERARCHIQUE				ALPHABETIQUE			
écart	corpus	texte	mot	écart	corpus	texte	mot
118.33	122	122	PATRES	8.60	6	2	ABSENTEM
66.29	46	42	CONSCRIPTI	2.64	44	2	ADDIDIT
14.13	9	4	DECREUERE	12.01	7	3	ADSEUERATION
13.36	196	19	UOS	9.46	5	2	AEDILES
12.01	7	3	ADSEUERATION	2.88	39	2	AETATEM
10.62	4	2	UOCANTUR	2.40	50	2	AGRIPPINA
10.62	4	2	LIBROS	2.83	40	2	ALTORUM
10.62	4	2	INCIPIENTE	6.21	11	2	AMBITUM
10.62	4	2	CENSUISSENT	4.22	74	4	AMICITIA
10.12	26	5	UOSTRA	6.81	618	21	APUD
9.53	29	5	PUBLICA	2.60	45	2	ASTIAE
9.51	19	4	REO	2.47	48	2	AUCTORITATE
9.46	5	2	NULLOS	5.02	57	4	AUGUSTI
9.46	5	2	MERUERANT	6.54	10	2	AUI
9.46	5	2	CENSUERE	2.46	1349	20	AUT
9.46	5	2	CENSUERANT	6.16	24	3	BENEFICIA
9.46	5	2	AEDILES	5.20	32	3	BONOS
8.67	124	10	UOBIS	3.23	33	2	CAMPO
8.66	13	3	EDICTUM	2.00	116	3	CAUSAM
8.60	6	2	SPERNI	9.46	5	2	CENSUERANT
8.60	6	2	ROGITANTES	9.46	5	2	CENSUERE
8.60	6	2	NOLITE	10.62	4	2	CENSUISSENT
8.60	6	2	MEUS	5.66	13	2	CLIENTES
8.60	6	2	EXORSUS	66.29	46	42	CONSCRIPTI
8.60	6	2	DECET	5.66	13	2	CRUCIATUS
8.60	6	2	ABSENTEM	4.42	20	2	CULPA
8.18	25	4	HABERETUR	4.86	17	2	CUNCTATIONE
7.92	7	2	RELIGIONEM	3.76	26	2	CURAE
7.38	8	2	UOCABAT	2.78	41	2	CURAM
7.38	8	2	UELLEM	4.42	20	2	CURIAM
7.29	256	13	ME	6.92	9	2	DATI
7.29	192	11	MIHI	3.98	49	3	DATUM
6.92	9	2	SAEPE NUMERO	2.34	960	15	DE
6.92	9	2	DATI	8.60	6	2	DECET
6.81	618	21	APUD	14.13	9	4	DECREUERE

*Environnement thématique de la forme patres dans HISTOIRE.EXE*

La liste des corrélats est affichée dans l'ordre alphabétique (à droite) – ce qui permet de vérifier aisément la présence d'un mot dont on subodore qu'il pourrait être en relation

étroite avec le mot-pôle – et dans l'ordre hiérarchique (à gauche), en fonction de l'importance de l'écart réduit<sup>8</sup>. La liste hiérarchique donne toujours le mot-pôle en première position : celui-ci est bien évidemment toujours en excédent fortement significatif dans des paragraphes sélectionnés justement du fait de sa présence.

Les résultats sont vérifiables au moyen des trois colonnes de chiffres qui donnent, dans l'ordre, la valeur de l'écart réduit, le nombre d'occurrences de chaque forme dans l'ensemble du corpus et son nombre d'occurrences dans les paragraphes sélectionnés par la fonction CONTEXTE et ici appelés « texte ». Le nombre des formes retenues peut être réduit si on adopte un critère plus sévère de sélection (bouton SEUIL).

## ANALYSE MULTIDIMENSIONNELLE DES LISTES

On a vu que le profil de chaque ligne ou de chaque colonne d'une liste pouvait être représenté par un graphique de distribution. Mais il est possible d'utiliser des procédures statistiques plus synthétiques que de simples histogrammes : les boutons FACTOR et ARBRE permettent de soumettre l'ensemble des données numériques contenues dans le tableau à des analyses multidimensionnelles et de représenter simultanément les similarités ou les distances entre les lignes et les colonnes (distances ou proximités entre les textes au regard de leur usage des diverses formes qui constituent la liste ; similarités ou dissemblances entre les profils de distribution des termes de la liste à travers le corpus ; conjonctions ou répulsions entre textes et termes étudiés).

### **1. L'analyse factorielle**

Le bouton FACTOR déclenche le programme d'analyse factorielle fourni par l'association ADDAD, qui distribue un logiciel complet pour l'analyse des données. Le module ici mis en œuvre est celui de l'analyse de correspondances, qui suit l'algorithme proposé par Jean-Paul Benzécri et dont l'adaptation à Windows a été réalisée par André Salem. L'analyse factorielle est donc réalisée par un programme extérieur, ANCORR.EXE, qu'il faut lancer quand le logiciel le demande<sup>9</sup>.

La représentation graphique des résultats de cette analyse se présente en quelque sorte comme une cartographie des textes et des paramètres d'étude : chacun d'eux est figuré par un point situé sur un plan muni de deux axes orthogonaux ; l'interprétation du graphe se fait en prenant en considération d'une part les distances entre les points, d'autre part leur distribution par rapport aux deux axes. En effet l'axe horizontal des abscisses symbolise le premier facteur d'analyse, celui qui explique le plus largement les données de la matrice initiale ; c'est donc l'opposition entre la droite et la gauche du graphe qu'il convient d'interpréter en premier lieu et qui fournit une bonne part de l'information. L'axe vertical des ordonnées symbolise le deuxième facteur d'analyse : l'opposition entre haut et bas du

---

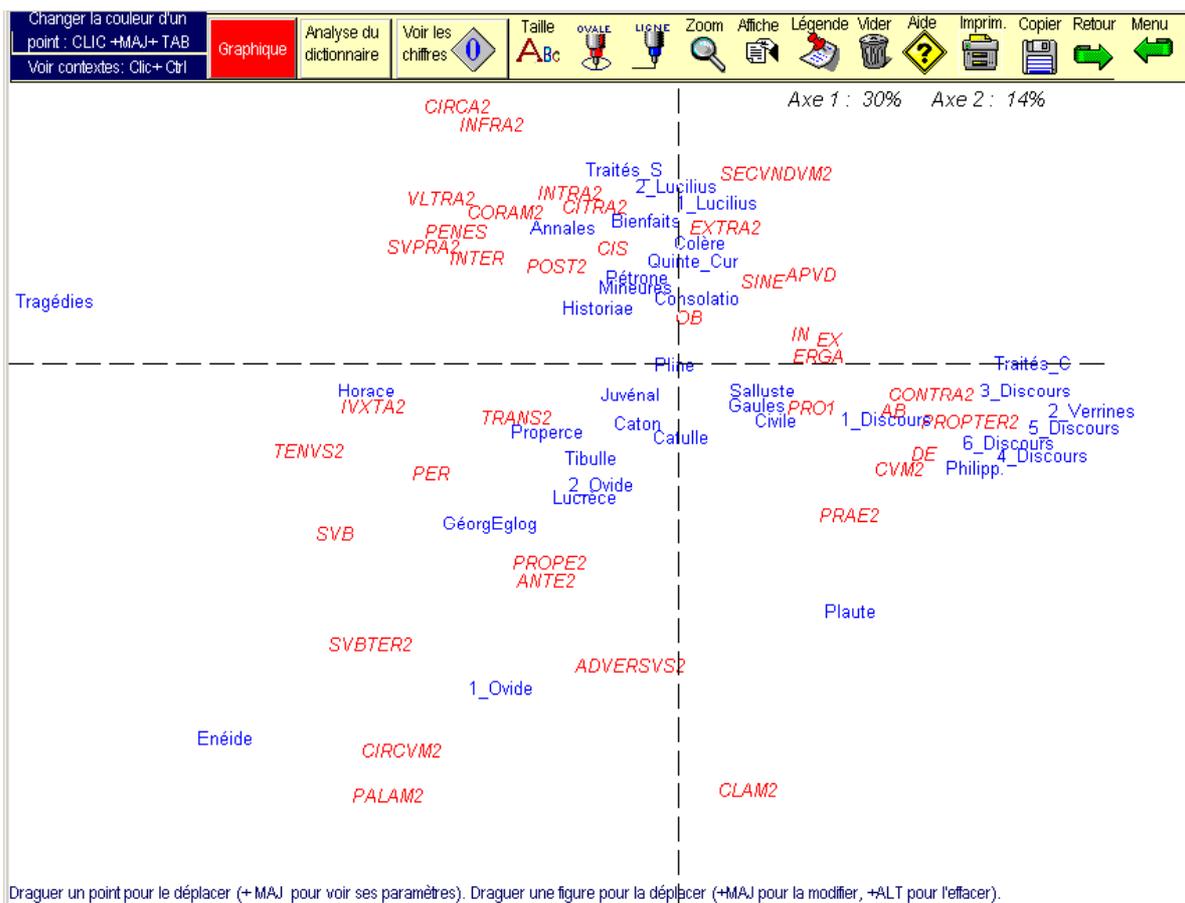
<sup>8</sup> A noter un petit problème d'affichage pour les mots trop longs tels qu'ici *adseueratione*.

<sup>9</sup> On a prévu la distorsion que peut amener dans les données linguistiques l'effet de taille, c'est-à-dire une trop grande disproportion entre les lignes (les mots peuvent avoir des fréquences très inégales dont le rapport peut être de 1 à 1 000), ou entre les colonnes (les textes – ou parties de texte – peuvent avoir des étendues fort déséquilibrées). Afin d'atténuer ces inégalités, le programme calcule les écarts réduits (ou ce qui en tient lieu, quand le modèle hypergéométrique est appliqué), puis les translate dans la zone positive, le plus grand nombre négatif s'alignant sur zéro et les autres éléments gardant leurs distances respectives (car l'analyse n'accepte pas les données négatives).

graphe ajoute donc des éléments plus subtils à l'interprétation (le taux d'information porté par ces deux axes est donné sous forme de pourcentage en haut à droite du graphe). Le logiciel propose aussi la représentation du troisième facteur d'analyse, combiné soit avec le premier, soit avec le deuxième<sup>10</sup>.

Noter que la désignation de chaque point occupe quatre lettres au maximum dans le fichier des résultats (ANALYSE.AFC). On peut rendre le graphique plus lisible en complétant les noms (sans modifier l'emplacement) ou en explicitant les symboles.

Pour illustrer l'analyse factorielle on a choisi comme premier exemple de représenter les proximités et distances entre les textes de la base générale LATIN.EXE en fonction de la distribution des prépositions dans ce corpus. Voici le graphe obtenu pour les deux premiers facteurs d'analyse :



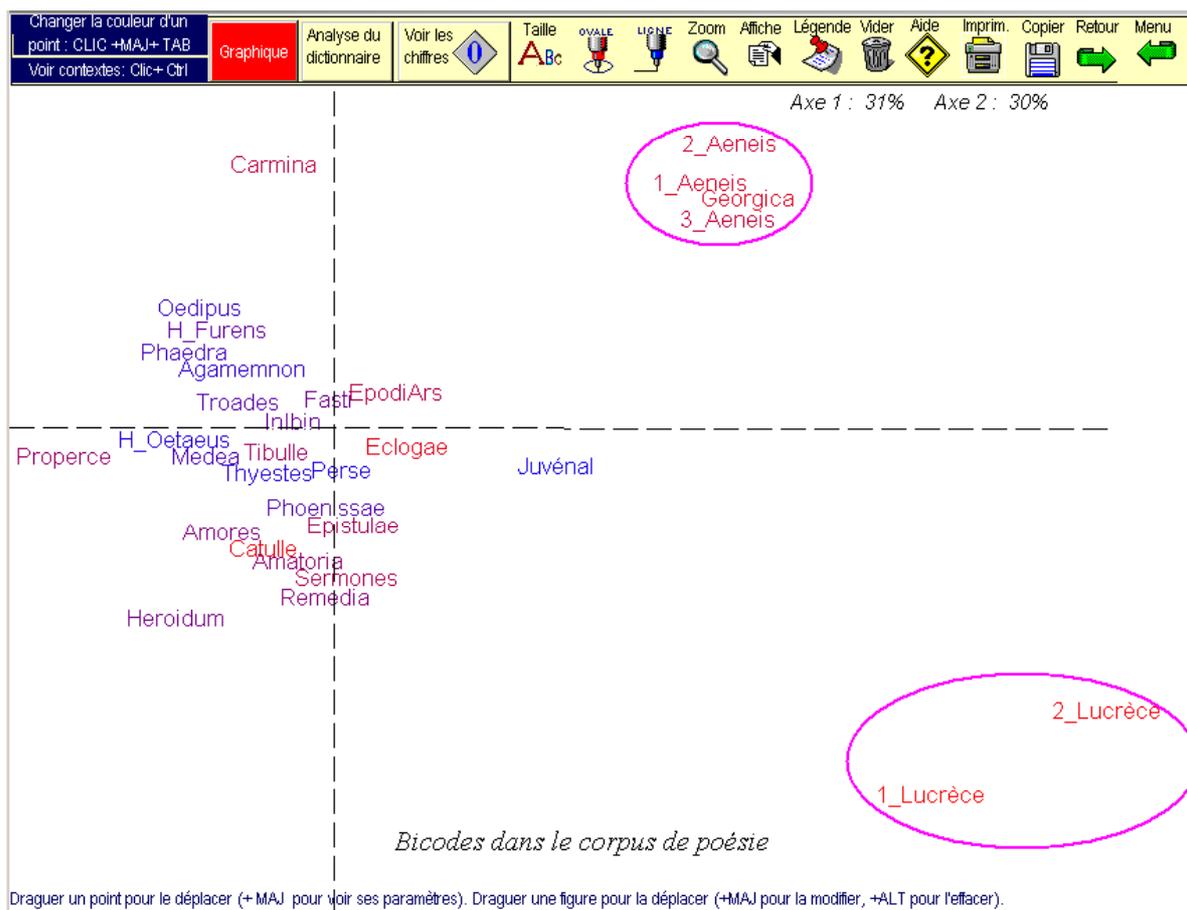
### *Analyse factorielle des correspondances pour l'ensemble des textes du corpus au regard de leur emploi des prépositions*

On observe ici très nettement la conjonction de deux facteurs d'analyse : l'opposition des genres discursifs et des auteurs d'une part, l'influence de la chronologie d'autre part. En effet, le long de l'axe horizontal, symbolique du premier facteur, on passe progressivement de la prose la plus classique à droite à la poésie à gauche (avec, au sein de cette dernière, une nette originalité pour les tragédies de Sénèque) et le long de l'axe vertical qui représente le deuxième facteur les œuvres s'échelonnent partiellement en fonction de leur date, Plaute et les Classiques étant situés dans la moitié inférieure tandis que Tacite, Sénèque et Pétrone accaparent la moitié supérieure. La distribution des variables, les prépositions, pourrait aussi

<sup>10</sup> Pour une explication plus précise, voir É. Évrard & S. Mellet (1998) : « Méthodes quantitative en langues anciennes », *LALIES* 18 : 111-155, en part. 140-144.

donner lieu à quelques commentaires ; on notera par exemple le regroupement dans la partie supérieure du graphe des formes *citra*, *ultra*, *infra*, *supra*, *extra*, *intra*.

Divers outils sont disponibles pour mettre en valeur les éléments principaux du graphique et pour en améliorer la lisibilité. Nous en donnons une illustration sur les deux figures ci-dessous qui représentent les proximités et distances des textes de la base POÉSIE.EXE au regard de la distribution des « bicodes », c'est-à-dire de toutes les successions possibles de deux codes grammaticaux (substantif-adjectif, substantif-verbe, substantif-préposition, substantif-adverbe, verbe-adverbe, adjectif-adverbe, etc.). Le premier graphe représente les premier et deuxième facteurs d'analyse :

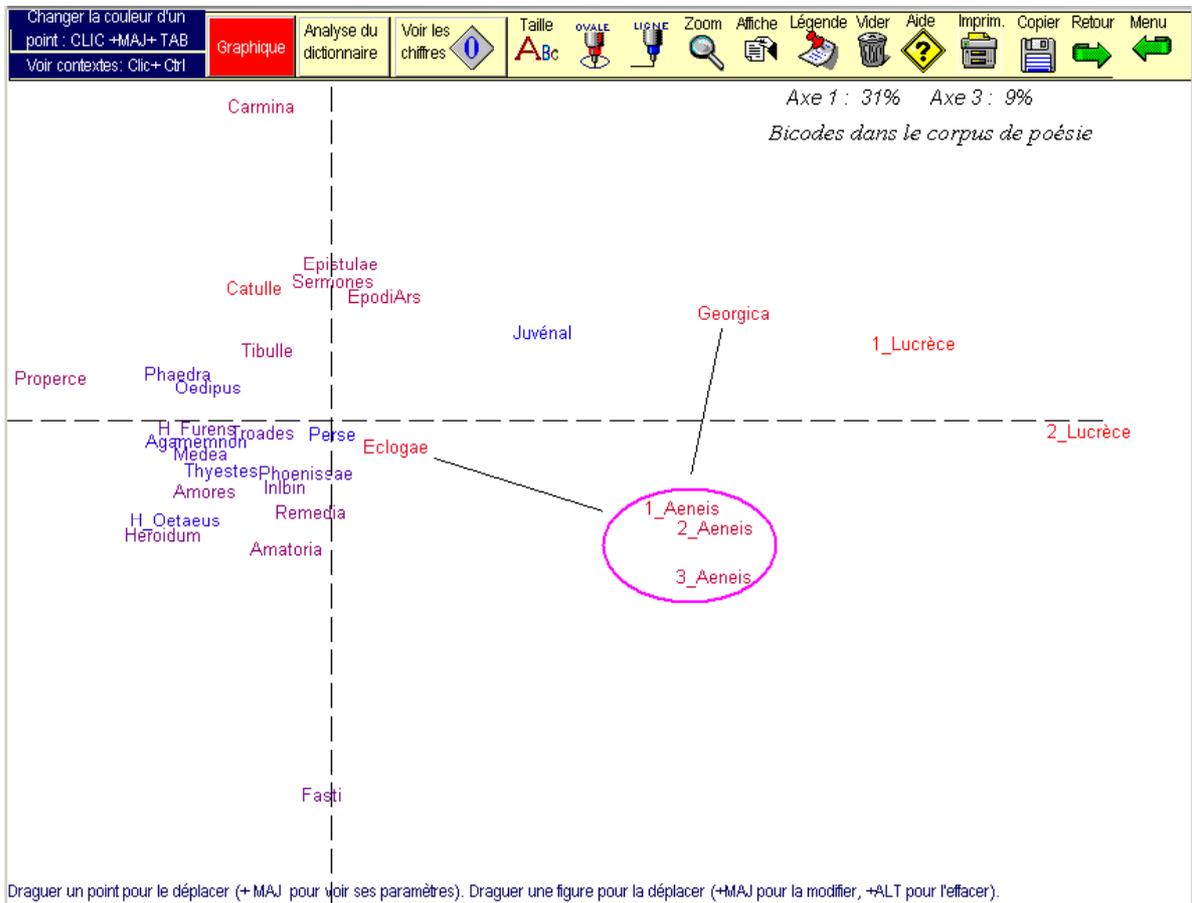


*Analyse factorielle des correspondances pour les textes du corpus POÉSIE au regard de la distribution des « bicodes » : axes 1 et 2*

On remarque d'abord que les variables (noms des « bicodes ») ont été supprimées de la représentation : trop nombreux en effet, et peu explicites, ces codes nuisaient à la lisibilité de la figure. Pour effacer ainsi soit les variables en lignes, soit les textes en colonnes, il suffit de solliciter le bouton AFFICHE du menu et de répondre à ses demandes de renseignement. Les groupements de textes intéressants ont été mis en évidence par des ellipses fuschia : il suffit pour ce faire de cliquer sur le bouton OVALE ; le positionnement de l'ellipse se fait en la draguant avec la souris et son plus ou moins grand étirement s'obtient en draguant et en appuyant simultanément sur la touche MAJUSCULE du clavier (on peut aussi l'effacer en la draguant et en appuyant simultanément sur la touche ALT). Un titre peut être placé sur le graphe grâce au bouton LÉGENDE (ici on a ajouté : « Bicodes dans le corpus de poésie »). La taille des caractères et l'épaisseur des traits peut être augmentée au moyen du bouton TAILLE. On remarque enfin que, spontanément, le logiciel

affecte une couleur variable aux noms des textes, couleur qui évolue du rouge au bleu en fonction de l'ordre chronologique choisi dans la base. On voit ainsi très clairement le rouge dominer à droite du graphe et le bleu à gauche (avec une exception notable : Catulle), reflétant l'évolution chronologique du corpus comme premier facteur d'analyse.

Ce phénomène est confirmé dans la seconde figure obtenue à partir du même tableau initial, mais donnant à voir les axes 1 et 3, c'est-à-dire les premier et troisième facteurs d'analyse respectivement sur l'axe horizontal et l'axe vertical :



*Analyse factorielle des correspondances pour les textes du corpus POÉSIE au regard de la distribution des « bicodes » : axes 1 et 3*

L'axe 3, porteur de 9% de l'information seulement, permet de séparer les trois livres de l'*Enéide* des *Géorgiques* et des *Bucoliques* (les traits qui les relient sont obtenus grâce au bouton LIGNE). Cet axe met aussi mieux en lumière l'originalité d'Ovide qui s'oppose à la fois à Catulle, Propertius, Tibulle et Horace, tandis que les tragédies de Sénèque restent groupées dans une zone centrale.

D'autres boutons encore sont disponibles sur cet écran : l'ANALYSE DU DICTIONNAIRE permet de refaire une analyse factorielle qui portera cette fois-ci sur les éléments du vocabulaire constitutif de chaque texte (formes graphiques ou lemmes) ; la distance entre deux textes du corpus est calculée en fonction du vocabulaire commun qu'ils partagent et du vocabulaire spécifique (ou exclusif) de chacun ; seules ces distances intertextuelles sont représentées sur l'AFC, les éléments du vocabulaire n'apparaissant pas pour des raisons évidentes de place et de lisibilité (voir p. 39).

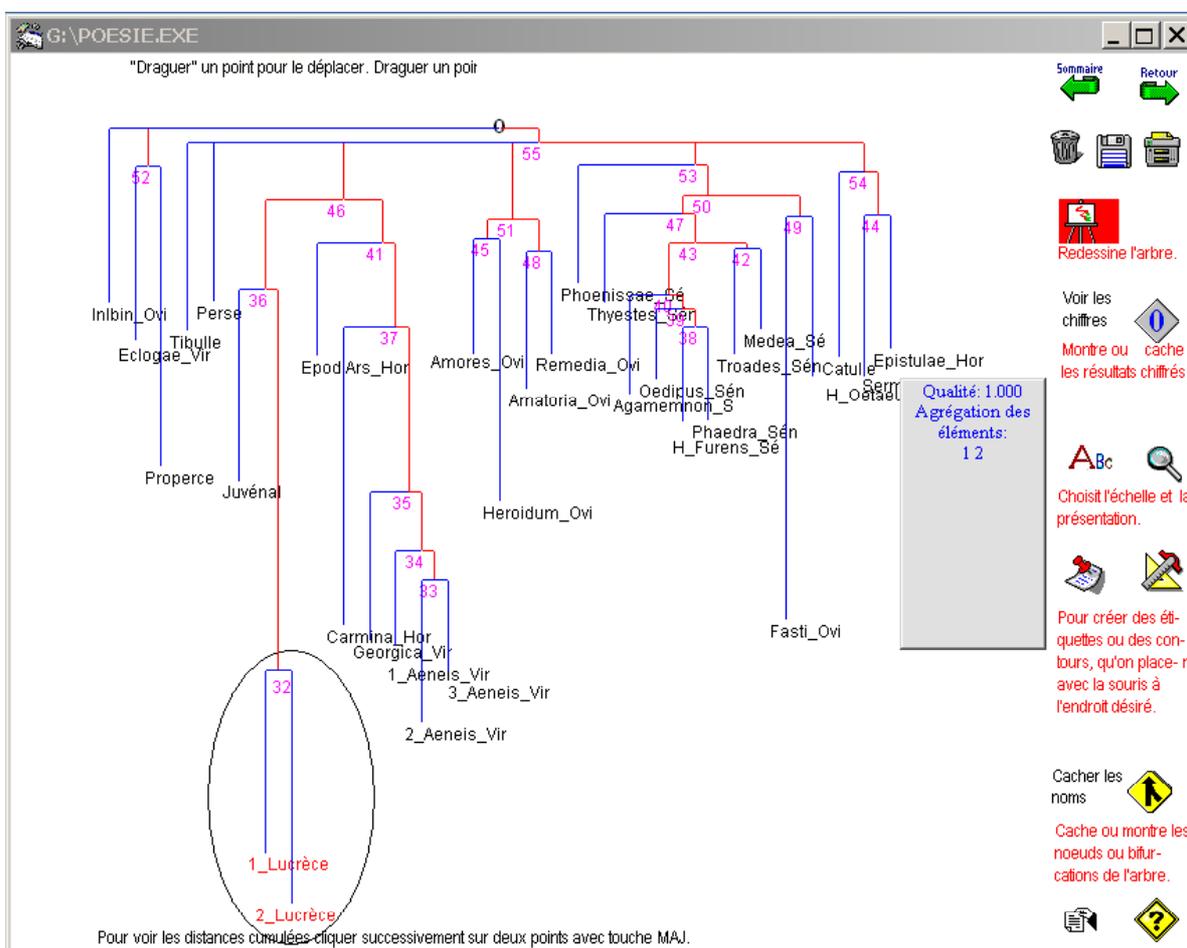
Le bouton VOIR LES CHIFFRES sollicite un éditeur de texte qui affiche l'ensemble des éléments chiffrés, tableaux et calculs ayant donné lieu à la représentation factorielle ; les plus expérimentés pourront y vérifier un certain nombre de paramètres explicatifs.

Le bouton GRAPHIQUE permet de refaire le graphique en choisissant d'afficher d'autres facteurs sur les axes orthogonaux. Enfin, le bouton COPIER indique comment faire une saisie de l'écran en cours pour le stocker dans le presse-papier.

## 2. L'analyse arborée

Il s'agit ici, non plus de cartographier des distances entre points sur un plan, mais d'opérer une classification hiérarchique ascendante des éléments étudiés en regroupant d'abord les deux plus proches, puis les deux suivants, etc. et en intégrant au processus de classement les groupements effectués à chaque étape précédente aux côtés des éléments isolés restants.

Cette méthode permet de calculer et représenter non seulement les distances entre des textes ou des variables, mais aussi les étapes des regroupements, appelés « nœuds » ; les graphes qui sont produits ressemblent à des chemins qui se rencontrent ou à des arbres aux multiples embranchements. La représentation dite « rectangulaire » *ne figure que* les étapes de la classification et les regroupements qui en sont issus. En voici ci-dessous un exemple qui reprend le même tableau de données utilisé pour l'analyse factorielle précédente :



Analyse arborée « rectangulaire » classant les textes de la base POÉSIE.EXE selon leur emploi des « bicoles »

Quelques outils ont été ménagés en vue d'améliorer la lisibilité du graphique :

1 - Quand il y a recouvrement, le programme donne à la souris le droit de déplacer les noms des textes.

2 - Les nœuds apparaissent ou non selon que le graphique est encombré ou non. Mais il est toujours possible de les montrer ou de les cacher (avant-dernier bouton jaune à partir du bas du menu). Il est à noter que ces nœuds sont numérotés dans l'ordre de leur création à partir du numéro du dernier élément classifié : ici le premier nœud créé porte le numéro 32 parce que la base comporte 31 textes. Le second nœud créé porte le numéro 33 et ainsi de suite. Observons une fois encore la pertinence des regroupements opérés.

3 - Lorsqu'on sollicite (en draguant) un de ces nœuds (par exemple le nœud 32 dans la figure ci-dessus), on fait apparaître en rouge toutes les feuilles (ou textes) qui dépendent de ce nœud (soit *1\_Lucrèce* et *2\_Lucrèce*), en même temps qu'une fenêtre grise apparaît indiquant le degré de cohésion (mesuré de 0 à 1) du groupe agrégé autour de ce nœud (ici 1, c'est-à-dire agrégation maximale des éléments 1 et 2). De même, en désignant successivement par un clic deux textes dont on veut mesurer la distance (enfoncer en même temps la touche MAJUSCULE), on fait apparaître une fenêtre indiquant la distance totale à parcourir entre les deux textes. L'ensemble des données de ce type peut, par ailleurs, être affiché en cliquant sur le bouton VOIR LES CHIFFRES.

4 - En effet, un bouton (losange gris affichant un zéro) permet alternativement de montrer ou de cacher les résultats quantifiés qui servent à la représentation graphique. Divers champs apparaissent alors qui pour chaque segment précisent le numéro d'ordre, le nom du texte, le nœud voisin, la distance du segment et les coordonnées du départ et de l'arrivée.

5 - D'autres outils, aux icônes explicites, permettent comme pour l'analyse factorielle (voir plus haut) d'enrichir le graphique avec des ellipses, des légendes et de l'agrandir plus ou moins pour occuper tout l'écran (le zoom - bouton ECHELLE - s'étend de 50% à 200%).

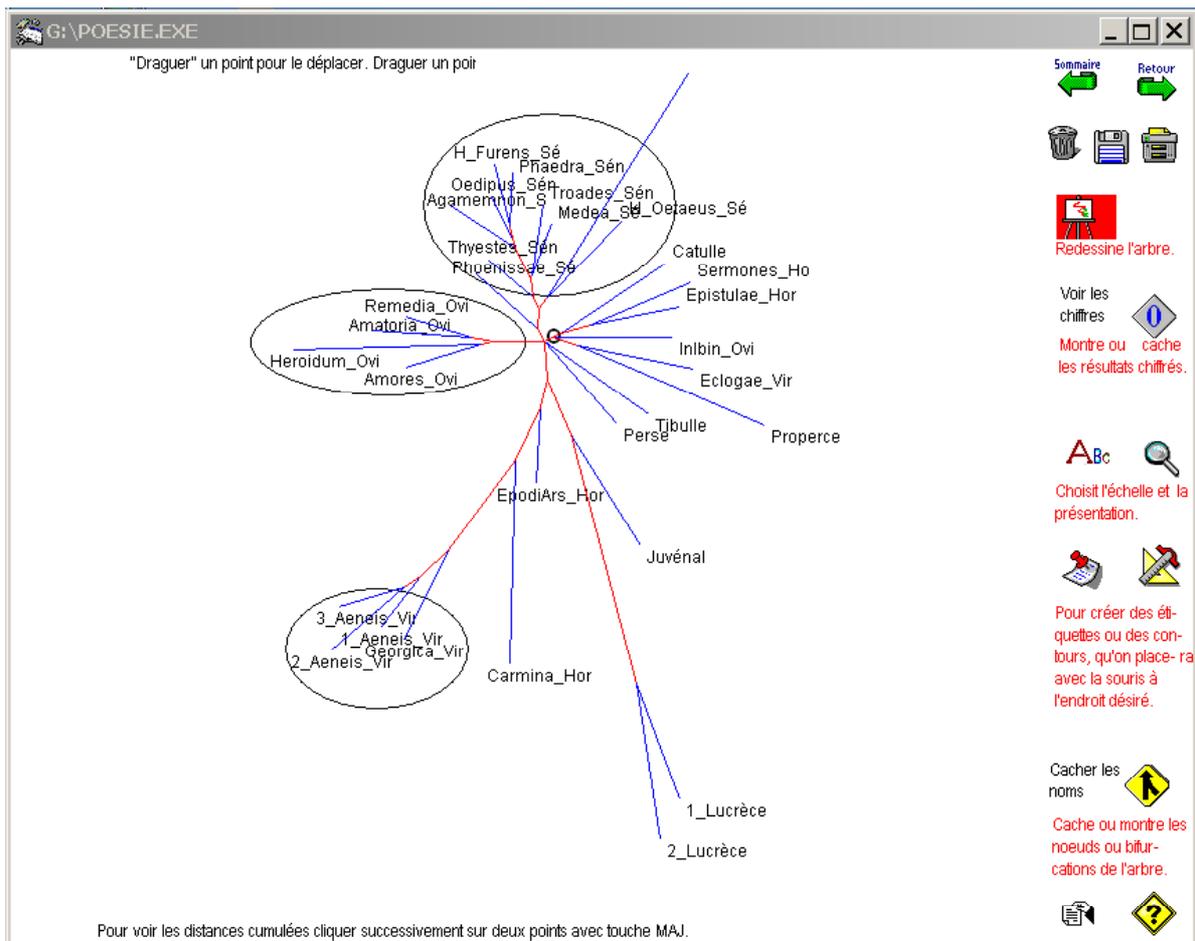
La représentation dite « radiale » figure à la fois les nœuds de regroupements et les *vraies distances* entre les éléments analysés. Elle repose sur une méthode de calcul développée par Xuan Luong<sup>11</sup>. Les distances sont donc plus faciles à interpréter dans cette présentation radiale, car elles sont directement proportionnelles à la longueur des parcours dessinés sur le graphique.

 Attention à ne pas mesurer les distances à vol d'oiseau. À chaque bifurcation le chemin emprunte une direction dont le sens importe peu, c'est la distance qui seule compte et qui se mesure par l'addition des segments de jonction.

Outre les distances, le niveau et la forme des embranchements joue un rôle important dans l'interprétation de ce type de représentation : on voit par exemple ici que les œuvres de Virgile (à l'exception des *Eglogues*) sont regroupées en un amas de petites branches très serrées et assez éloignées du nœud central de l'arbre. Ceci est l'indice d'une forte proximité entre ces œuvres et d'une originalité spécifique globale qui les différencie nettement des autres textes du corpus. De même pour les deux parties du texte de Lucrèce. La plupart des œuvres d'Ovide et les *Tragédies* de Sénèque sont aussi fortement regroupées ; elles se situent en revanche plus près du nœud central, ce qui suggère une moindre spécificité. Enfin les autres textes se répartissent dans une structure dite « en étoile » peu significative et peu interprétable.

---

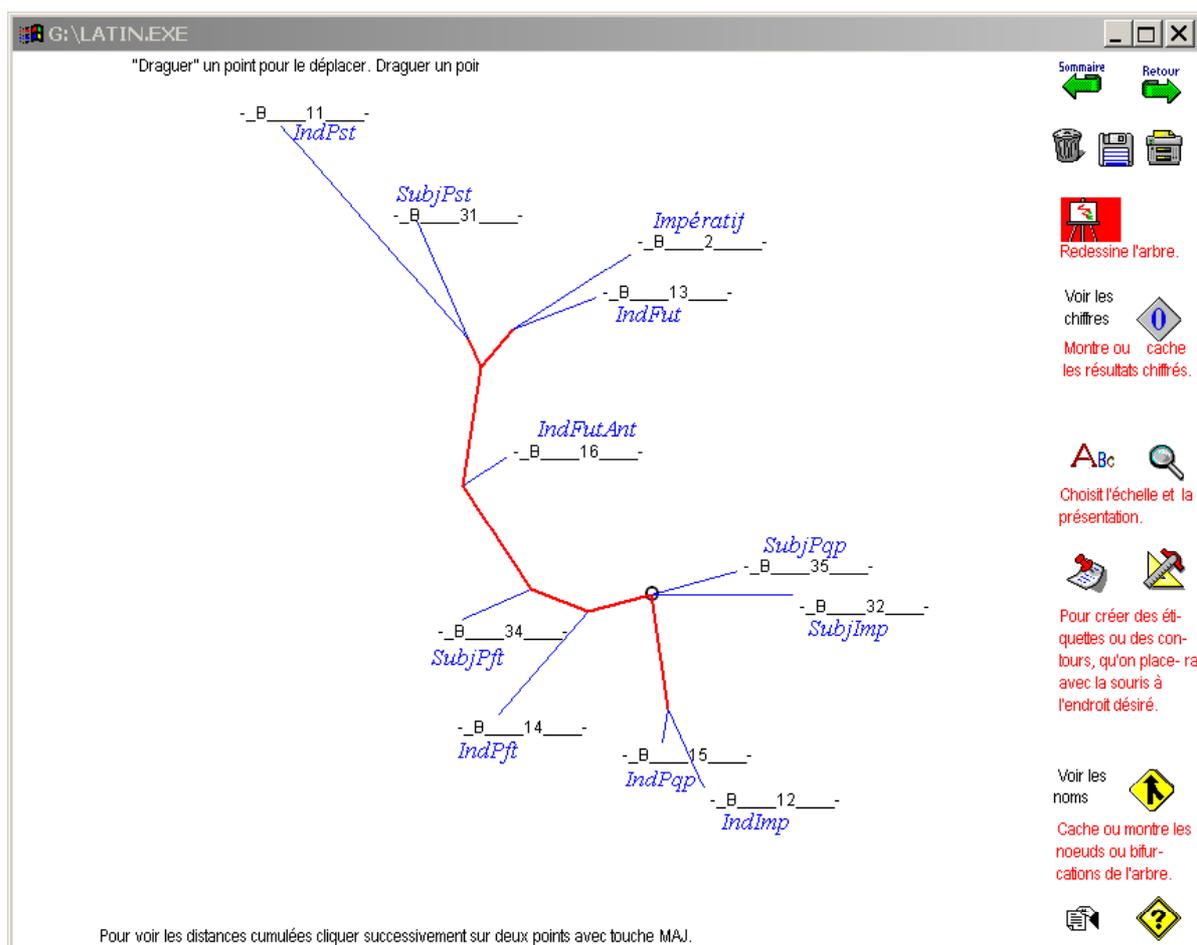
<sup>11</sup> Pour plus de détail, consulter Luong X. (éd.) (1989). *Analyse arborée des données textuelles, Tree Analysis of Textual Data*. Revue CUMFID 16, Nice.



*Analyse arborée « radiale » classant les textes de la base POÉSIE.EXE selon leur emploi des « bicodes »*

Les mêmes outils de présentation et d'interprétation du graphique que précédemment sont toujours disponibles. Signalons en outre qu'en cas de recouvrement, le programme donne à la souris le droit de déplacer les textes (ou feuilles du graphe), à condition de respecter le point d'ancrage (ou nœud) auquel ils se rattachent et la longueur du segment terminal (en bleu) qui leur est propre. Il suffit de draguer le nom du texte à déplacer et de le faire pivoter jusqu'à l'endroit désiré, en s'arrêtant quand la longueur requise est atteinte et que le programme l'autorise (l'icône du curseur change de forme et le trait bleu reste en pointillés quand le trajet est trop court ou trop long).

Notons enfin que le programme offre aussi de calculer les distances entre les variables (lignes de la matrice numérique) – et non pas seulement entre les textes (colonnes de la matrice). En voici une illustration dans laquelle on a représenté les affinités ou éloignements entre les principales formes verbales selon l'usage qu'en font les différents textes de la base générale :



### Analyse arborée « radiale » de la distribution des principales formes verbales dans la base LATIN.EXE

Les résultats sont conformes aux attentes, opposant les branches des présents et des futurs d'un côté à celles des imparfaits et plus-que-parfaits de l'autre, le parfait – au milieu – hésitant comme prévu entre énonciation historique et discours. Les regroupements plus fins sont aussi intéressants, manifestant tantôt les affinités temporelles (imparfait avec plus-que-parfait), tantôt les proximités modales (indicatif futur et impératif)<sup>12</sup>.

### LA CONNEXION LEXICALE (OU DISTANCE LEXICALE)

Il s'agit cette fois-ci d'évaluer la proximité ou la distance des textes les uns par rapport aux autres en fonction du vocabulaire qu'ils partagent et du vocabulaire spécifique à chacun d'eux. Cette fonction est accessible à partir du menu principal, en sollicitant le bouton DISTRIBUTION du menu statistique (vertical en bas à gauche de la page d'accueil), puis le

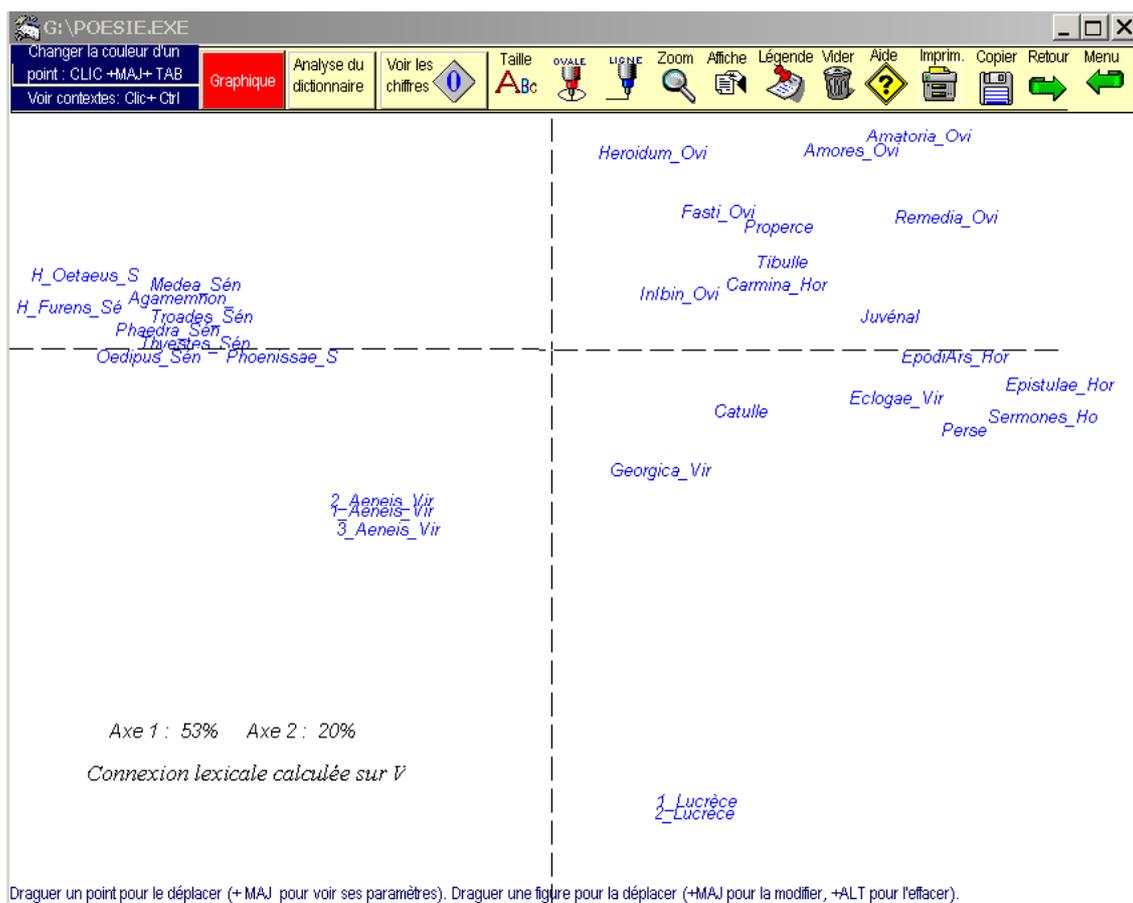
<sup>12</sup> Il est bien évident que la rédaction d'un manuel impose de choisir des exemples simples et aisément interprétables pour faciliter l'exposé et rendre accessible l'outil présenté ; en contrepartie ces exemples n'ont généralement aucune valeur heuristique et n'apprennent rien au lecteur. Il serait abusif d'en conclure que l'outil n'est susceptible d'apporter aucune information nouvelle et déploie des moyens excessifs au regard des résultats obtenus : c'est au chercheur, comme toujours, de se montrer inventif et de poser les bonnes questions face auxquelles il sera vite convaincu de l'utilité du logiciel.

bouton DISTANCE. Pour deux textes dont on cherche à apprécier la connexion, un mot contribue à rapprocher ces deux textes s'il est commun aux deux et à augmenter la distance s'il est privatif et ne se rencontre que dans un seul (voir p. 34).

Hyperbase propose deux méthodes de calcul<sup>13</sup> :

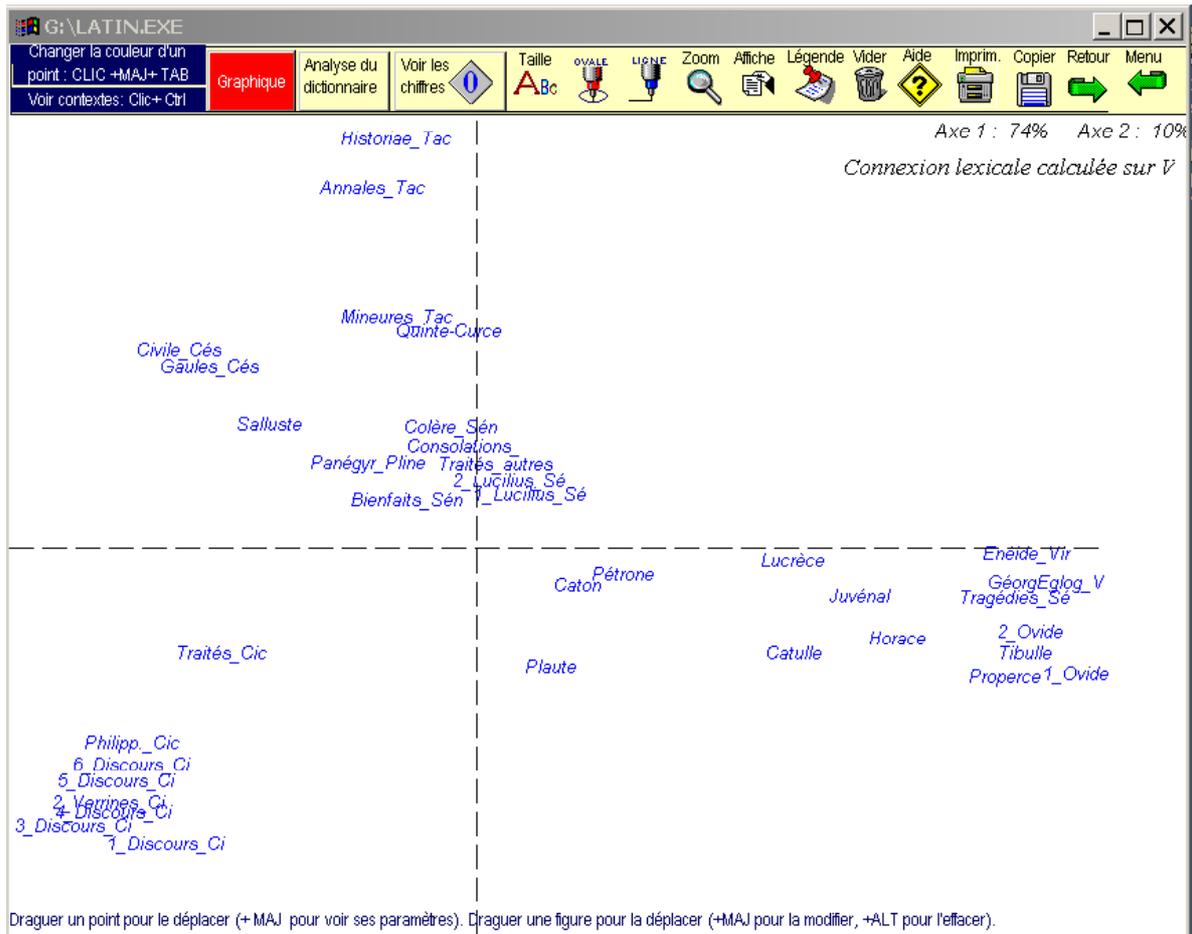
- l'une suit la méthode Jaccard qui ne se préoccupe pas de fréquence et pour un mot donné ne considère que sa présence ou son absence dans le texte considéré (bouton FACTORIELLE SUR V) ;
- l'autre prend en compte la distribution des fréquences et considère que plus l'écart de fréquence d'un mot commun aux deux textes mais inégalement réparti est grand, plus il contribue à différencier les deux textes en questions (bouton FACTORIELLE SUR N).

C'est donc ici l'ensemble du dictionnaire de la base qui est soumis à la représentation graphique par le programme d'analyse factorielle. Le programme laisse le choix d'effectuer les calculs sur les formes graphiques, sur les lemmes, et même sur les codes grammaticaux affectés à chaque forme. Les regroupements sont généralement spectaculaires, comme on le constate dans les deux figures ci-dessous consacrées respectivement à la connexion lexicale des textes de la base POÉSIE.EXE calculée sur les lemmes par la méthode Jaccard et à la connexion lexicale de tous les textes de la base générale LATIN.EXE calculée sur les formes graphiques par la même méthode. La représentation est ici celle de l'analyse factorielle, mais elle peut aussi prendre la forme d'une analyse arborée.



*Connexion lexicale des textes de la base POÉSIE, calculée sur la présence/absence des lemmes*

<sup>13</sup> Sur les calculs de distances intertextuelles et leurs applications, voir le n°2 de la revue CORPUS.

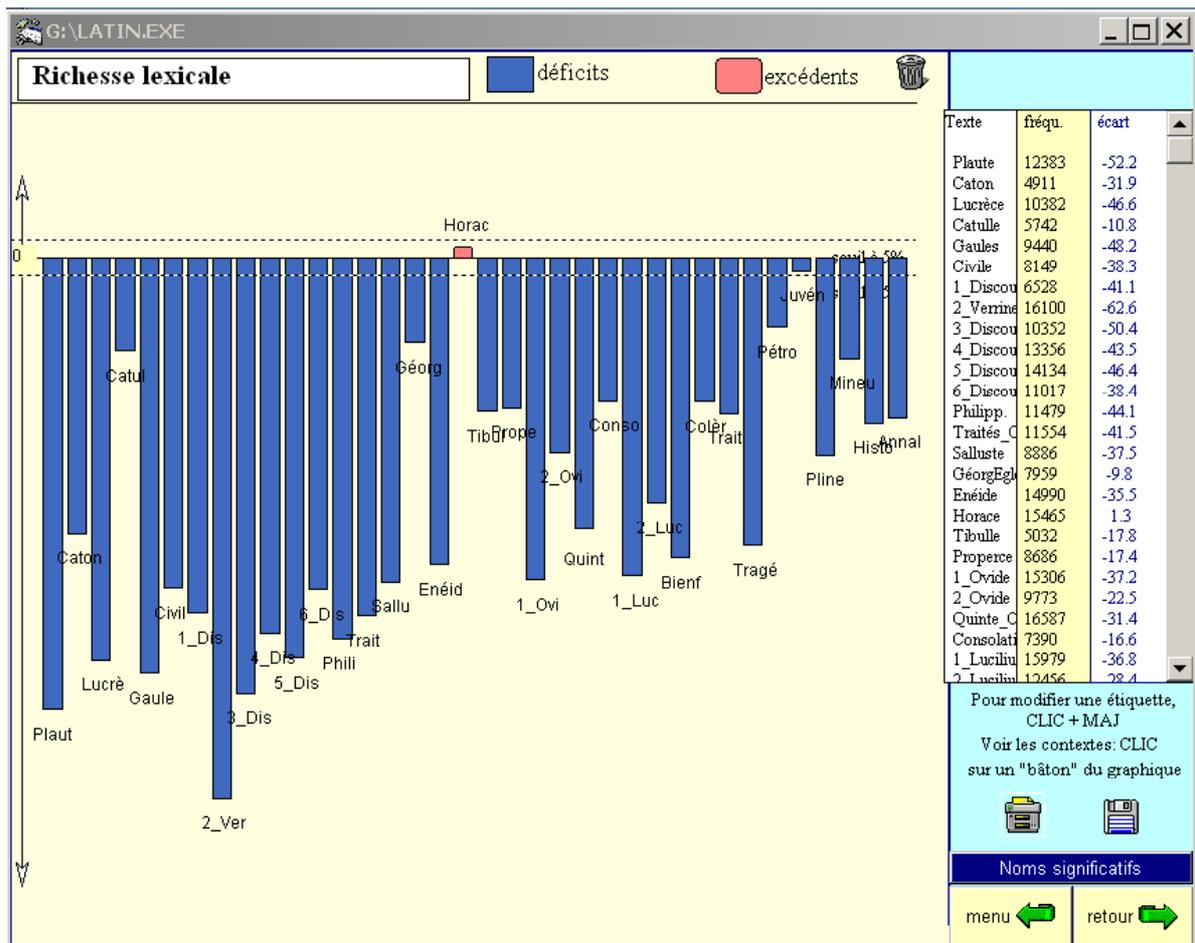


*Connexion lexicale des textes de la base générale, calculée sur la présence/absence des formes graphiques*

## RICHESSSE LEXICALE, HAPAX, ACCROISSEMENT

Le programme de préparation des traitements statistiques, entre autres tâches, constitue le tableau de distribution des classes de fréquences, le relevé des hapax (ou mots employés une seule fois) et bien d'autres résultats qui intéressent la structure du vocabulaire. Pour voir et imprimer ces tableaux, solliciter le bouton DISTRIBUION de l'écran d'accueil qui conduit à une page spécifique où sont consignés les résultats statistiques acquis dans cette perspective.

Le bouton RICHESSE ET HAPAX assure le dénombrement des formes différentes relevées dans chaque texte. Et, en s'appuyant sur le tableau de distribution des fréquences (accessible en cliquant sur le bouton correspondant) et sur l'étendue relative des textes, un programme mesure la part du vocabulaire théoriquement présent dans chacun des textes en vertu de la loi binomiale et compare cet effectif attendu à celui qu'on observe en réalité ; la différence est appréciée, une fois encore, au moyen d'un écart réduit (affiché dans la colonne la plus à droite) et donne une image de la richesse lexicale des textes conçus comme des sous-ensembles du corpus. Cette mesure peut, au choix, être effectuée sur les formes graphiques ou sur les lemmes et les résultats obtenus peuvent être représentés par un histogramme :

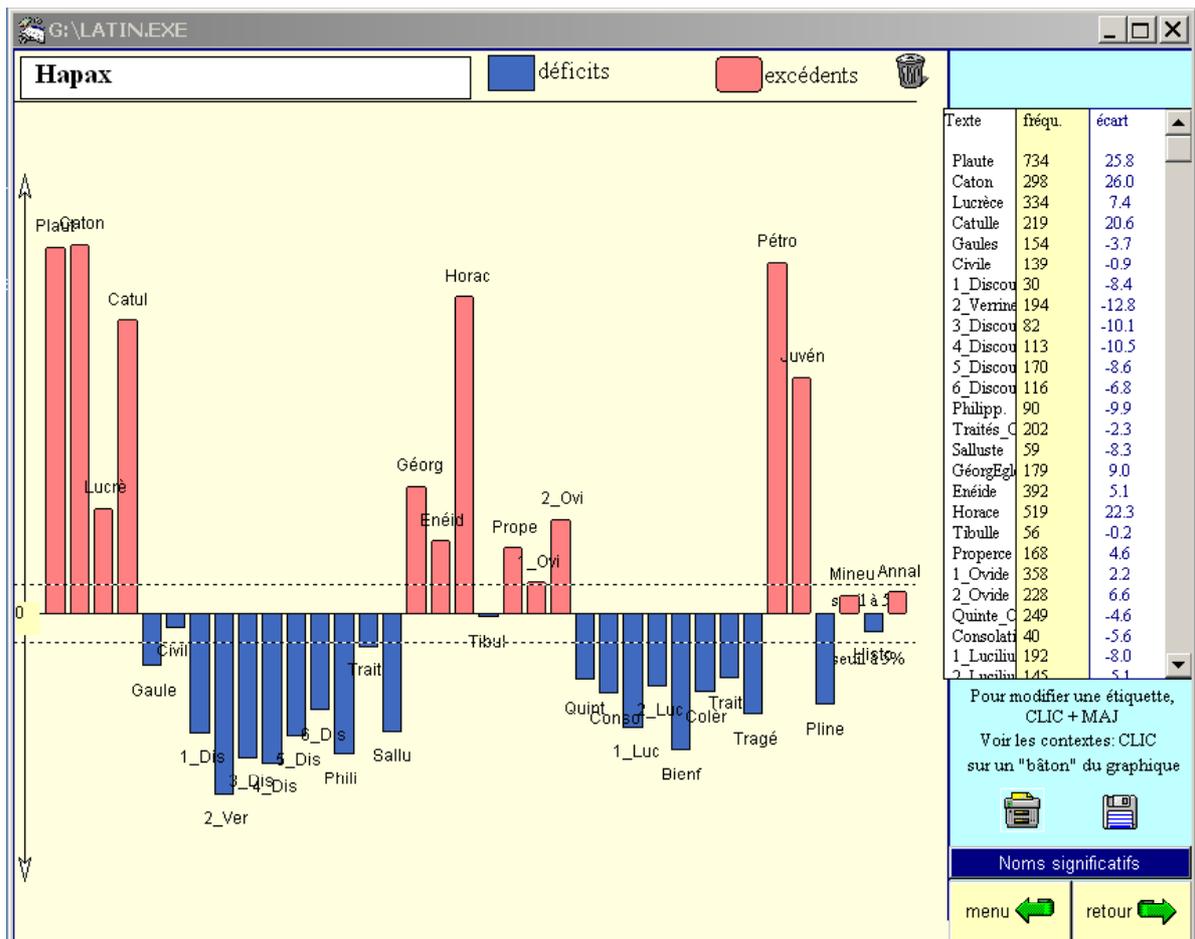


*Histogramme de la richesse lexicale calculée sur les formes graphiques (base LATIN.EXE)*

On ne s'étonnera pas de constater que tous les écarts sont négatifs, à l'exception notable de l'œuvre d'Horace. Ceci « traduit le phénomène partout observé de la spécialisation lexicale : les mêmes mots se trouvent dans les mêmes textes, en vertu des contraintes, thématiques ou stylistiques, de la situation de discours »<sup>14</sup>. Ce sont, généralement, les œuvres poétiques qui offrent le vocabulaire le plus riche.

Le même menu donne accès à la représentation de la distribution des hapax (formes et lemmes), c'est-à-dire des formes ou des lemmes qui ont été rencontrés une seule fois dans le corpus, et conséquemment dans un seul texte. La méthode est ici plus simple et se rattache à la loi normale. On aboutit pareillement à des écarts réduits qui servent d'ordonnées au programme de courbe.

<sup>14</sup> É. Brunet, *Le vocabulaire de Victor Hugo*, Genève : Slatkine, 1988, p. 27.



*Distribution des hapax dans la base LATIN.EXE*

Le bouton ACCROISSEMENT CHRONOLOGIQUE donne une visée dynamique de la structure du vocabulaire dans le corpus étudié, puisqu'il permet d'évaluer le cumul progressif des formes et le renouvellement – de plus en plus ralenti – du vocabulaire. La direction la plus naturelle de cette évolution est la chronologie, mais le chemin inverse qui prend le temps à rebours peut aussi révéler des ruptures intéressantes. La représentation graphique peut se faire œuvre par œuvre, ou par tranches chronologiques.

Enfin, toujours dans la même page ouverte par le bouton DISTRIBUTION du menu principal, le bouton ÉVOLUTION CHRONOLOGIQUE donne accès aux listes de formes et de lemmes dont l'emploi est en accroissement significatif au fil du temps ou, au contraire, en décroissance significative (la pertinence de cette évolution étant mesurée par un coefficient de corrélation variant entre  $-1$  et  $+1$ ). Les mêmes résultats peuvent être affichés par ordre alphabétique (bouton ÉVOLUTION ALPHABÉTIQUE) afin de faciliter la recherche d'une forme ou d'un lemme précis.

G:\HISTOIRE.EXE

→ alpha. **L'évolution du lexique (hiérarchique)** seuil Sommaire

Clicker sur un mot pour voir les contextes

	Progression	Fréquence	Forme	Régression	Fréquence	Forme
FORMES	+ 0.765	408	quia	- 0.794	483	omnibus
	+ 0.758	157	super	- 0.787	265	suos
	+ 0.752	141	principis	- 0.783	1925	se
	+ 0.749	187	nomen	- 0.779	221	re
	+ 0.742	96	olim	- 0.778	2059	ex
	+ 0.732	9991	et	- 0.766	1244	ab
	+ 0.727	310	mox	- 0.760	539	his
	+ 0.722	51	perinde	- 0.758	174	magno
	+ 0.707	318	an	- 0.756	169	una
	+ 0.703	195	cuncta	- 0.752	373	eos
LEMES	+ 0.697	81	specie	- 0.749	164	possent
	+ 0.685	157	donec	- 0.746	172	omni
	+ 0.685	59	preces	- 0.744	245	parte
	+ 0.682	46	diuersa	- 0.738	192	consilio
	+ 0.675	45	uetus	- 0.732	256	proelio
	+ 0.669	115	modum	- 0.731	463	omnes
	+ 0.669	83	ue	- 0.730	442	eorum
	+ 0.669	34	recens	- 0.722	146	fuga
	+ 0.819	145	validvs	- 0.830	2183	omnis
	+ 0.797	229	et1	- 0.787	2980	svil
+ 0.796	146	compono	- 0.772	1187	facio	
+ 0.790	78	infensvs	- 0.772	88	consvesco	
+ 0.781	277	adversvs3	- 0.764	2448	ex	
+ 0.766	162	omitto	- 0.761	180	contendo	
+ 0.765	408	qvia	- 0.758	1715	res	
+ 0.761	438	nomen	- 0.756	2821	ab	
+ 0.754	133	crimen	- 0.755	2430	hic1	
+ 0.745	84	insignis	- 0.749	1478	magnvs	
+ 0.744	147	svper2	- 0.745	105	reliqvi	
+ 0.742	96	olim	- 0.740	99	privsqvam	
+ 0.739	192	traho	- 0.733	112	contineo	
+ 0.739	73	vvlgo1	- 0.732	1315	possvml	
+ 0.735	80	praecipvvs	- 0.731	280	proficiscor	
+ 0.734	116	promptvs2	- 0.730	148	convenio	
+ 0.727	310	mox	- 0.724	207	impero	
+ 0.727	21	assevero	- 0.724	194	pvgno	

*Evolution du vocabulaire chez les historiens latins*

## SPÉCIFICITÉS

Ce bouton permet de calculer, d'une part quel est le vocabulaire ou quelles sont les catégories grammaticales spécifiques d'une base partielle par rapport à la base générale LATIN.EXE, d'autre part quel est le vocabulaire ou quelles sont les catégories grammaticales spécifiques de chacune des œuvres constitutive d'une base par rapport à l'ensemble de cette base. C'est le corpus de la base qui sert alors de norme pour les textes qui la constituent.

Dans le premier cas de figure (une base partielle comparée à la base générale) deux présentations sont, là encore, possibles : si l'on choisit la présentation hiérarchique (« Lemmes triés » ou « Formes triées » dans le rectangle bleu ciel central), les deux listes sont triées d'après la valeur absolue de l'écart réduit, de façon à mettre en relief ce qui est le plus significatif, dans le sens des excédents comme dans celui des déficits. Ci-dessous on trouvera la liste des lemmes en excédent et en déficit significatifs dans la base TRAITES.EXE comparée à la base LATIN.EXE. La présentation alphabétique est moins pertinente, mais plus pratique pour retrouver une forme dont on subodore qu'elle pourrait faire partie du vocabulaire spécifique d'une base. Dans les deux cas cependant on peut demander au programme, grâce au bouton CHERCHER, de vérifier si un mot figure parmi le vocabulaire spécifique, positif ou négatif.

N°	écart	corpus	texte	mot	N°	écart	corpus	texte	mot
54.74	1193	747	BENEFICIVM		-20.83	25801	1941	QVE	
36.28	39989	7006	SVM1		-16.21	6441	335	ATQVE1	
29.09	20415	3722	NON		-15.83	88604	8844	.	
28.40	4300	1101	ENIM2		-15.44	13637	1015	EGO	
26.90	38546	6203	QVI1		-12.09	4653	279	MEVS	
25.23	1936	583	AVTEM		-12.04	3650	193	NEQVE	
25.09	227	148	IRASCOR		-11.79	2298	87	ROMANVSA	
24.86	328	183	INGRATVS		-11.14	1616	45	MILES	
23.94	1753	527	QVIDAM		-11.06	12589	1073	TV	
23.20	665	270	VOLVPTAS		-10.44	1285	30	EXERCITVS1	
22.68	275	153	VUTILITAS		-10.08	2461	127	VOS	
21.83	438	198	SAPIENS1		-10.02	1725	68	ARMA	
21.28	2570	647	ALIQVIS		-9.85	1710	69	SENATVS	
20.63	357	167	VUTILIS		-9.71	1113	26	CASTR2	
20.44	769	272	IRA		-9.38	2447	137	BELLVM	
20.38	1392	407	MATVRA		-9.17	2477	143	POPVLVS1	
19.19	583	217	AMICITIA		-8.84	1968	104	VRBS	
18.84	13729	2314	ILLE		-8.80	2969	193	NVNC	
18.65	5225	1044	QVAM1		-8.68	1230	46	CONSVL	
18.57	1802	464	DEBEO		-8.54	915	24	HIC2	
18.14	286	132	SENECTVS1		-8.28	1083	39	NOX	
17.92	3989	830	ANIMVS		-8.23	3985	299	IAM	
17.78	549	198	HVMANVS		-8.21	1910	108	HOSTIS	
17.41	40004	5795	ET2		-8.02	1717	94	MITTO	
17.37	2867	634	QVIDEM		-7.99	870	26	PES	
17.23	685	225	VITIVM		-7.95	1314	61	VESTER	
17.21	2724	607	BONVS		-7.77	2669	183	DIES	
17.19	1752	436	ACCIPIO		-7.74	1345	66	NE2	
17.15	907	272	INIVRIA		-7.66	2454	165	TVM	
17.00	6399	1185	NEC2		-7.52	825	27	CRIMEN	
16.68	177	92	ELOQVENTIA		-7.47	2797	200	VENIO	
16.50	83	58	HONESTAS		-7.46	4457	361	PER	
16.39	138	78	ORATOR		-7.45	882	32	AQVA	
16.23	530	182	PROSVM1		-7.35	959	39	MILLE	
16.18	717	223	OFFICIVM		-7.18	1361	74	IVDEX	

### Spécificités, en lemmes, de la base TRAITÉS.EXE par rapport à la base générale

Dans le second cas de figure (spécificités d'un texte par rapport à l'ensemble des textes constitutifs de la base à laquelle il appartient), on n'a accès qu'aux seuls excédents et seule la présentation hiérarchique est disponible. Cependant le bouton CHERCHER est toujours là pour permettre de vérifier si un mot est présent dans la liste : simultanément à la mise en surbrillance inverse du mot cherché, s'ouvre une fenêtre donnant le profil du mot au sein des œuvres, si du moins l'emploi de ce mot est suffisamment caractéristique pour franchir le seuil significatif dans au moins un des textes du corpus.

Comme les listes dépassent généralement les possibilités de l'écran, le bouton EDITER permet de les restituer sur l'imprimante dans leur intégralité. Rappelons que le seuil significatif généralement admis est aux alentours de la valeur 2 (en laissant 5 chances sur 100 au hasard). Nous avons préféré nous arrêter à la valeur 3 en adoptant un seuil plus sévère.

On a choisi comme exemples ci-dessous la liste des formes et lemmes spécifiques du *de Officiis* dans la base TRAITÉS.EXE et deux listes de codes grammaticaux spécifiques : celle de Caton dans la base générale et celle du livre 7 de la *Guerre des Gaules* dans la base des historiens (où l'on remarquera le positionnement en début de liste des formes de présent de l'indicatif dû à la fréquence importante du présent de narration dans ce livre).

G:\TRAITES.EXE

Refaire résumé Officiis(forme) Mots Phrases Codes Officiis(lemme) Recherche Sommaire

CLIC sur un mot: Recherche du mot dans les spécificités Officiis CLIC+MAJ: Recherche du mot dans les textes

N°	écart	corpus	texte	mot	N°	écart	corpus	texte	mot
2	16.5	583	266	autem	2	24.6	2389	879	IS
2	14.0	468	208	id	2	16.5	583	266	AVTEM
2	13.7	160	103	igitur	2	13.7	160	103	IGITVR
2	13.0	105	77	honestum	2	13.5	58	55	HONESTAS
2	12.6	276	138	ea	2	12.5	194	109	HONESTVS
2	10.0	336	137	atque	2	11.8	217	113	RESPUBLICA
2	9.1	164	80	hominum	2	10.9	223	109	OFFICIVM
2	8.8	606	198	de	2	10.6	84	58	IVSTITIA
2	8.6	193	86	Neque	2	10.6	605	215	HOMO
2	8.3	640	201	etiam	2	10.1	226	105	GENVS1
2	8.2	175	78	iis	2	9.9	335	136	ATQVE1
2	8.0	71	43	genere	2	9.8	1941	518	QVE
2	7.5	97	50	eam	2	9.4	68	47	SOCIETAS
2	7.3	239	91	esset	2	9.4	26	26	PANAETIVS0
2	7.3	158	68	maxime	2	9.3	6203	1395	QVI1
2	7.1	53	33	iure	2	8.9	606	198	DE
2	7.1	23	20	iustitiae	2	8.6	193	86	NEQVE
2	7.0	49	31	ii	2	8.4	640	201	ETIAM
2	6.9	22	19	decorum	2	8.1	25	23	IVSIVRANDVM
2	6.7	918	248	esse	2	7.9	946	268	RES
2	6.7	49	30	iustitia	2	7.9	165	73	TVM
2	6.7	17	16	honestate	2	7.5	29	24	RELIQVVS
2	6.4	1613	393	cum	2	7.2	66	38	NEZ
2	6.3	58	32	omnino	2	7.2	19	18	COMMVNITAS
2	6.0	12	12	honestatis	2	7.1	16	16	DECORVM
2	5.7	23	17	item	2	6.9	15	15	CONTENTIO2
2	5.7	11	11	iure iurando	2	6.8	1192	309	SVM2
2	5.7	11	11	honestatem	2	6.6	47	29	VENDO
2	5.4	30	19	eas	2	6.6	101	47	GLORIA
2	5.4	10	10	Panaetio	2	6.5	48	29	ARBITROR
2	5.3	50	26	officium	2	6.4	41	26	MVLTVTVD0
2	5.3	143	53	eorum	2	6.4	140	58	IVS1
2	5.2	1107	269	enim	2	6.3	58	32	OMNINO
2	5.2	109	43	homines	2	6.3	29	21	COGNITIO
2	5.1	9	9	duplex	2	6.3	13	13	APPETITVS

*Le vocabulaire spécifique du de Officiis dans les traités*

G:\LATIN.EXE

Refaire résumé Caton(forme) Mots Phrases Codes Caton(lemme) Recherche Sommaire

Caton

2	7.7	952	41	_0	
2	32.8	2514	296	_d18	Numéral, Cardinal, Indéclinable,
2	32.8	723	620	_b3_1_231200	Verbe, 3e conjugaison, Singulier, Impératif, Futur, Actif, 2e personne,
2	32.8	277	215	_b1_1_231200	Verbe, 1re conjugaison, Singulier, Impératif, Futur, Actif, 2e personne,
2	32.8	259	219	_b5_1_231200	Verbe, Conjugaison mixte, Singulier, Impératif, Futur, Actif, 2e personne,
2	28.6	141	88	_b2_1_231200	Verbe, 2e conjugaison, Singulier, Impératif, Futur, Actif, 2e personne,
2	26.8	1451	172	_d132	Numéral, Cardinal, Accusatif, Pluriel,
2	25.1	25739	682	_a231	Substantif, 2e déclinaison, Accusatif, Singulier,
2	21.8	893	114	_d131	Numéral, Cardinal, Accusatif, Singulier,
2	21.0	12555	389	_c1311	Adjectif, 1re classe, Accusatif, Singulier, Positif,
2	19.2	936	101	_a68	Substantif, 5e déclinaison, Indéclinable,
2	18.4	75	42	_b4_1_231200	Verbe, 4e conjugaison, Singulier, Impératif, Futur, Actif, 2e personne,
2	16.9	16432	399	_a131	Substantif, 1re déclinaison, Accusatif, Singulier,
2	16.8	72	37	_b6_1_1312sx	Verbe, Conjugaison anormale, Singulier, Indicatif, Futur, Actif, 2e personne,
2	14.0	140	36	_b6_1_1313sx	Verbe, Conjugaison anormale, Singulier, Indicatif, Futur, Actif, 3e personne,
2	12.9	17489	358	_a332	Substantif, 3e déclinaison, Accusatif, Pluriel,
2	12.5	122	30	_b6_1_1313ln	Verbe, Conjugaison anormale, Singulier, Indicatif, Futur, Actif, 3e personne,
2	12.4	23	18	_b3_1_231300	Verbe, 3e conjugaison, Singulier, Impératif, Futur, Actif, 3e personne,
2	11.6	8760	209	_a132	Substantif, 1re déclinaison, Accusatif, Pluriel,
2	11.2	498	44	_d332	Numéral, Distributif, Accusatif, Pluriel,
2	11.0	13780	278	_c1321	Adjectif, 1re classe, Accusatif, Pluriel, Positif,
2	10.9	67	21	_b6_1_231200	Verbe, Conjugaison anormale, Singulier, Impératif, Futur, Actif, 2e personne,
2	10.6	24302	417	_a232	Substantif, 2e déclinaison, Accusatif, Pluriel,
2	10.6	872	53	_b3_1_311300	Verbe, 3e conjugaison, Singulier, Subjonctif, Présent, Actif, 3e personne,
2	10.4	368	36	_d2611	Numéral, Ordinal, Ablatif, Singulier,
2	10.4	26	15	_b6_1_1313wt	Verbe, Conjugaison anormale, Singulier, Indicatif, Futur,

*Les codes spécifiques de l'œuvre de Caton dans la base générale*

Refaire résumé		7_Gaules(forme)		Mots	Phrases	Codes	7_Gaules(lemme)	Cherche	Imprimer	Sommaire
8	2.5	118	9	_0						
8	7.3	1239	90	_b3_1_111300	Verbe, 3e conjugaison, Singulier, Indicatif, Présent, Actif, 3e personne,					
8	6.6	4359	217	_a262	Substantif, 2e déclinaison, Ablatif, Pluriel,					
8	6.4	686	56	_b3_2_111300	Verbe, 3e conjugaison, Pluriel, Indicatif, Présent, Actif, 3e personne,					
8	5.6	3715	179	_a362	Substantif, 3e déclinaison, Ablatif, Pluriel,					
8	5.3	25335	933	_r	Préposition,					
8	5.3	621	46	_l_62	Pr. indéfini, Ablatif, Pluriel,					
8	4.8	1095	65	_b362_442_ad	Verbe, 3e conjugaison, Ablatif, Pluriel, Participe, Parfait, Passif,					
8	4.7	309	27	_b562_442_ad	Verbe, Conjugaison mixte, Ablatif, Pluriel, Participe, Parfait, Passif,					
8	4.6	575	40	_d132	Numéral, Cardinal, Accusatif, Pluriel,					
8	4.5	1984	99	_g_31	Pr. réfléchi, Accusatif, Singulier,					
8	4.5	1194	67	_a561	Substantif, 4e déclinaison, Ablatif, Singulier,					
8	4.3	901	53	_i_62	Pr. démonstratif, Ablatif, Pluriel,					
8	4.3	26	7	_b2_2_3113xk	Verbe, 2e conjugaison, Pluriel, Subjonctif, Présent, Actif, 3e personne,					
8	4.2	7093	283	_a261	Substantif, 2e déclinaison, Ablatif, Singulier,					
8	4.2	820	49	_l_61	Pr. indéfini, Ablatif, Singulier,					
8	4.2	661	42	_i_42	Pr. démonstratif, Génitif, Pluriel,					
8	4.1	1177	63	_d18	Numéral, Cardinal, Indéclinable,					
8	4.1	203	19	_c1622	Adjectif, 1re classe, Ablatif, Pluriel, Comparatif,					
8	4.0	8923	344	_a361	Substantif, 3e déclinaison, Ablatif, Singulier,					
8	4.0	251	21	_c1613	Adjectif, 1re classe, Ablatif, Singulier, Superlatif,					
8	4.0	141	15	_d332	Numéral, Distributif, Accusatif, Pluriel,					
8	4.0	32	7	_b1_1_3223bn	Verbe, 1re conjugaison, Singulier, Subjonctif, Imparfait, Passif, 3e personne,					
8	3.9	1109	59	_b361_442_ad	Verbe, 3e conjugaison, Ablatif, Singulier, Participe, Parfait, Passif,					
8	3.9	339	25	_b1_2_111300	Verbe, 1re conjugaison, Pluriel, Indicatif, Présent, Actif, 3e personne,					
8	3.8	1168	61	_a252	Substantif, 2e déclinaison, Datif, Pluriel,					
8	3.8	369	26	_b162_442_ad	Verbe, 1re conjugaison, Ablatif, Pluriel, Participe, Parfait, Passif,					
8	3.8	9	4	_b4_2_112300	Verbe, 4e conjugaison, Pluriel, Indicatif, Présent, Passif,					

### *Les codes spécifiques du livre 7 de la Guerre des Gaules parmi les historiens*

Les icônes représentant des doigts pointés permettent de passer aux spécificités du texte suivant ou précédent dans la base : c'est une autre façon de les faire défiler, à côté de la sélection par accession à la liste donnée dans le rectangle bleu turquoise central de la barre de menu. Par ailleurs, le bouton REFAIRE LE RESUME permet de recalculer, en en modifiant éventuellement les paramètres, les phrases d'un texte qui contiennent le plus de mots spécifiques et sont donc, à ce titre, représentatives du texte, en résumant les thématiques et le style. Cette fonction est également accessible à partir du bouton PHRASES-CLES du menu principal (en bas à gauche de l'écran d'accueil).

## LES PHRASES-CLÉS

La technique utilisée consiste à prendre appui sur les spécificités lexicales et à retenir les paragraphes qui contiennent les spécificités les plus nombreuses et les plus significatives.

Le nombre absolu des spécificités enregistrées dans un paragraphe dépend du seuil adopté pour en établir la liste. Généralement ce seuil correspond à la probabilité de 0,05 laissée au hasard. Nous l'avons adopté également, avec un correctif qui limite à 200 le nombre des éléments considérés pour un même texte. Ces éléments sont évidemment les premiers de la liste triée des spécificités. Mais leur nombre dans une phrase doit être



## TABLE DES MATIÈRES

---

<b>Présentation.....</b>	<b>2</b>
<b>Installation .....</b>	<b>3</b>
<b>Le menu principal .....</b>	<b>4</b>
<b>Chapitre 1 : l'exploration libre .....</b>	<b>7</b>
<b>Chapitre 2 ; l'exploitation documentaire .....</b>	<b>12</b>
La fonction CONTEXTE .....	12
La fonction CONCORDANCE.....	13
Objet de la recherche.....	14
Les listes.....	20
<b>Chapitre 3 : l'exploitation statistique .....</b>	<b>24</b>
Partition et statistique .....	24
Les graphiques.....	25
L'environnement thématique .....	30
Les analyses mutidimensionnelles des listes.....	31
Analyses factorielles .....	31
Analyses arborées.....	35
La connexion lexicale (ou distance lexicale) .....	38
Richesse lexicale, hapax, accroissement .....	40
Spécificités .....	43
Phrases-clés .....	46