

*ESSAI D'EXPLOITATION
DE DONNEES FOURNIES PAR DES MOYENS INFORMATIQUES SUR
LES SATIRES DE PERSE*

Nous avons examiné les documents suivants relatifs aux *Satires* de Perse : la liste de fréquence en ordre décroissant, les comptages grammaticaux, l'index avec les analyses grammaticales et la concordance lemmatisée. L'édition de référence est celle de D.Bô.

L'essai est partiel : parmi le grand nombre de voies de recherche qu'ouvrent les documents cités, nous avons retenu : la richesse du vocabulaire, les thèmes des *Satires* et l'emploi du verbe *DICO* qui est, après le verbe *SVM*, le verbe le plus utilisé par Perse.

Il ne s'agit pas ici d'une étude stylistique à proprement parler, mais plutôt d'une contribution à une telle étude en utilisant l'aide de l'informatique et de la statistique.

I. LA RICHESSE DU VOCABULAIRE DE PERSE

Pour étudier le vocabulaire (1), nous partons de trois données fournies par la liste de fréquence : le nombre total de mots du texte (les occurrences), le nombre de lemmes (nombre de mots différents du point de vue sémantique) et le nombre de lemmes qui n'apparaissent qu'une fois. Ces trois données sont

respectivement symbolisées par les lettres M, L et L1. Voici les chiffres qui nous intéressent pour les *Satires* de Perse : M = 4647, L = 1938, L1 = 1237. A partir de ces éléments, nous allons essayer de déterminer quelle est la richesse du vocabulaire de Perse (2).

Le premier type de rapport est la moyenne d'emploi des lemmes : M / L , nombre de mots divisé par nombre de lemmes. Plus cette moyenne est faible, plus le vocabulaire est riche. Pour Perse ce chiffre est de 2,40. C'est un chiffre très bas qui nous donne un premier indice de richesse. Mais il faut faire la réserve suivante : plus un texte est court, plus la moyenne d'emploi des lemmes risque d'être faible et c'est le cas des *Satires* de Perse. Cependant la comparaison avec un texte en vers de longueur avoisinante : les *Bucoliques* de Virgile (5294 mots, 1449 lemmes) fait apparaître non seulement que la moyenne d'emploi des lemmes y est nettement plus élevée : 3,65 mais que, pour un nombre de mots plus important chez Perse, Virgile a un nombre moins important de lemmes.

Il existe un deuxième type de rapport beaucoup plus précis car il tient compte des variations de longueur entre les textes et permet de comparer, dans certaines limites, des oeuvres de longueur différente : L / \sqrt{M} : nombre de lemmes divisé par la racine carrée du nombre de mots (3). Il est ainsi possible de confronter, sur le plan de la richesse lexicale, les *Satires* de Perse avec un grand nombre d'autres oeuvres. Le rapport L / \sqrt{M} donne un indice dont la moyenne est 22,14 (4) pour les seize oeuvres dont nous avons utilisé les relevés et qui représentent un corpus de plus de 220.000 mots. Plus cet indice est élevé, plus le vocabulaire de l'auteur considéré est riche et inversement. Pour établir ce corpus, nous avons cherché parmi les oeuvres dont nous avons les relevés, des oeuvres

appartenant au même genre que les *Satires* de Perse (Juvénal), des oeuvres contemporaines (Sénèque), des oeuvres en prose ou en vers antérieures ou postérieures aux *Satires* de Perse et nous les avons rassemblées dans un tableau chronologique en séparant la prose de la poésie :

	TEXTES (5)	MOTS	LEXEMES	INDICE
I. PROSE				
-	César, <i>De Bello Gallico</i> .	42826	2798	13,52
-	Salluste, <i>De Coniuratione Catilinae</i> .	10094	1773	17,65
	<i>De Bello Iugurthino</i> .	20282	2260	15,86
-	Tite Live, <i>Ab Vrbe Condita</i> *.	16665	2581	19,99
-	Sénèque, <i>Consolatio ad Marciam</i> .	8304	2204	24,07
	<i>Consolatio ad Helviam Matrem</i> .	6775	1861	22,61
-	Tacite, <i>Annales</i> *.	13795	2762	23,52
II. POESIE				
-	Plaute, <i>Amphitryon</i> .	9422	1521	15,66
-	Catulle, <i>Carmina</i> .	12366	2824	25,40
-	Virgile, <i>Bucoliques</i> .	5294	1449	19,91
	<i>Géorgiques</i> .	14263	3005	25,16
-	Horace, <i>Odes</i> .	12457	2861	25,63
-	Tibulle, <i>Elégies</i> .	12635	2629	23,38
-	Ovide, <i>Métamorphoses</i> *.	9852	2327	23,44
-	PERSE, <i>SATIRES</i> .	4647	1938	28,43
-	Juvénal, <i>Satires</i> .	25562	4809	30,07

Le calcul de l'indice de richesse nous permet de déceler plusieurs lignes de partage dont l'une au moins était attendue : celle qui sépare la prose de la poésie; presque toutes les oeuvres en prose se situent en dessous de l'indice moyen de 22 et c'est l'inverse pour les oeuvres en vers. Il s'agit là d'un partage senti intuitivement par chacun. Mais deux autres éléments, beaucoup plus difficiles à manier se dégagent du tableau. Le premier est une progression irrégulière mais sensible (et indépendante de la longueur des textes) de l'indice de richesse dans l'ordre chronologique des oeuvres en prose. Cela pourrait montrer un accroissement du lexique au cours de la latinité en même temps qu'un goût des prosateurs pour utiliser un vocabulaire de plus en plus varié. Il faudrait pour vérifier cette hypothèse disposer d'un nombre plus important de relevés et aussi étudier les distributions de vocabulaire de chaque relevé, mais là n'est pas notre propos. Nous nous attacherons plutôt au deuxième élément qui montre, à l'intérieur de la classe "poésie", une sous-classe "satire" se distinguant par un vocabulaire particulièrement riche. Les indices de Perse et de Juvénal sont nettement plus élevés (28 et 30) que les indices des oeuvres poétiques les plus riches : les *Carmina* de Catulle et les *Odes* d'Horace (25). Cette hypothèse demanderait à être confirmée par les indices des *Satires* de Lucilius et d'Horace mais nous pouvons déjà chercher quelques éléments d'explication.

Le poète satirique est, comme les autres poètes, plus soucieux du choix de son vocabulaire que le prosateur. Aussi retrouvons-nous dans la satire les caractéristiques du vocabulaire poétique : les mots rares ou archaïques, les mots liés aux nécessités métriques etc. Ces lemmes viennent en fait s'ajouter à un vocabulaire fondamental à peu près identique à celui de la prose. Mais nous ne trouvons pas chez le poète satirique la même unité de thèmes que chez les autres poètes. Au contraire, la satire se caractérise par une multiplicité de thèmes : nous voyons

Perse parler de lectures publiques, de cuisine, de naufrage, de prières au temple, d'héritage, de jeunes gens aux bains ... (6), Juvénal s'attaquer aussi bien aux lectures publiques qu'à la sportule ou aux différents embarras de la vie à Rome. La *SATVRA* est bien encore à l'époque de Perse puis de Juvénal un "mélange" des genres et des thèmes et cela est probablement un élément d'explication de la grande richesse lexicale des oeuvres satiriques.

Le dernier type de rapport à l'intérieur d'une liste de fréquence concerne le pourcentage des lemmes de fréquence 1 dans le nombre total de lemmes (L1/L). Comme pour la moyenne d'emploi, la comparaison n'est possible qu'entre des textes de dimensions voisines. Nous nous contenterons donc de faire le rapprochement avec Virgile, *Bucoliques* (5294 M) et avec Sénèque, *Consolatio ad Helviam matrem* (6775 M).

Virgile : L1 / L =	51 %
Sénèque	= 56 %
PERSE	= 64 %

Perse possède donc relativement plus de lemmes de fréquence 1 que Virgile et Sénèque. C'est un indice de la dispersion de son vocabulaire, et l'on remarquera que pour ces trois oeuvres la progression du pourcentage des lemmes de fréquence 1 vient corroborer la progression de l'indice de richesse calculé plus haut.

D'une manière générale nous pouvons donc constater que tous les calculs (moyenne d'emploi, indice de richesse, pourcentage des lemmes de fréquence 1) vont dans le même sens : le vocabulaire des *Satires* de Perse est très riche, plus riche même que les vocabulaires pourtant très variés des *Odes* d'Horace et des *Carmina* de Catulle. La seule oeuvre qui de ce point de vue arrive au niveau

des *Satires* de Perse et même les dépasse est précisément l'oeuvre d'un autre poète satirique : Juvénal. Il y a là certainement un trait caractéristique de la "SATVRA".

II. LES THEMES DES SATIRES DE PERSE

La méthode proposée ici pour la recherche des thèmes se fonde essentiellement sur l'étude de la liste de fréquence décroissante des *Satires* de Perse et sur sa comparaison avec une seconde liste que nous appelons "liste de référence" et qui vient du dépouillement d'un corpus de plus de 320.000 mots (7). Les deux listes sont conçues de manière différente. Celle des *Satires* de Perse est une liste complète, c'est-à-dire que les 1938 lemmes rencontrés, y compris les noms propres, y ont été rangés dans l'ordre de fréquence décroissante. La liste de référence est le *Lexique de base latin* du C.R.I.D.E.L.A. (8), qui ne contient que 144 lemmes, et qui nous a paru être un bon outil de comparaison pour trois raisons principales.

La première tient à l'étendue dans le temps des oeuvres prises en compte (de Cicéron à Tacite) et à l'égale proportion des oeuvres en prose et de celles en vers.

La deuxième tient à la structure même de cette liste. En effet y apparaissent les lemmes qui répondent aux deux conditions suivantes (9) :

- Présence chez tous les auteurs étudiés.
- Fréquence au moins égale à la "fréquence-limite" (10).

Sont donc éliminés de cette liste des lemmes dont la haute fréquence serait liée à un contexte particulier chez tel ou tel auteur.

La troisième raison tient au fait suivant : si l'on ôte au corpus qui sert à établir la liste, la moitié des oeuvres qui le composent : Catulle, Cicéron, Horace, Sénèque, Tacite et Tite-Live qui n'apparaissent pas dans la première édition du lexique (1972), soit 187.263 mots sur 323.085, cela ne change strictement rien au nombre et à l'ordre des lemmes. C'est une preuve, nous semble-t-il suffisante, de la valeur de cette liste comme point de référence pour distinguer, dans le vocabulaire d'une oeuvre, ce qui lui est particulier.

Aussi bien dans la liste de référence que dans celle de Perse, nous établissons une distinction entre les "mots-outils" (conjonctions, prépositions, pronoms, particules de liaison, le verbe *SVM*...) et les "mots-pleins" (substantifs, verbes, adjectifs, adverbes de signification ...) non pas pour éliminer les uns ou les autres mais afin de distinguer, dans un premier temps, les faits de structure et les faits lexicaux. Dans un second temps, lorsqu'il s'agit de tirer des conclusions stylistiques, les deux séries de faits doivent être rapprochées. Nous verrons, dans la troisième partie, comment la haute fréquence du mot-outil *TV* et celle du mot-plein *DICO* sont liées dans les *Satires* de Perse.

En présence de ces deux listes, nous allons procéder à une double comparaison. La liste de référence comprend, sur ses 144 lemmes, 80 mots-pleins que nous confronterons aux 80 premiers mots-pleins théoriques (11) des *Satires*. Nous chercherons d'abord les lemmes de la liste des *Satires* qui ne se retrouvent pas sur la liste de référence et nous constituerons ainsi une troisième liste que nous appellerons "liste-réduite". Cette liste contiendra donc les lemmes dont la haute fréquence est particulière à Perse. Ensuite nous chercherons si, parmi les lemmes qui sont communs aux deux listes, certains présentent un écart de place significatif, en termes de probabilité, dans le sens positif pour Perse; il s'agirait donc

de lemmes fréquents dans la langue mais qui se trouveraient chez Perse avec une fréquence encore beaucoup plus élevée que celle que l'on aurait pu attendre.

Voici le début, jusqu'à la fréquence 7, de la liste des mots-pleins des *Satires* (12) :

1	DICO	(26)		VENIO	(10)	DIES	(7)
2	RES	(18)	17	DEVS	(9)	EXEO	(7)
	VIVO	(18)		MAGNVS	(9)	FAS	(7)
4	DO	(17)		PONO	(9)	IVBEO	(7)
5	O	(13)		VNVS	(9)	LOQVOR	(7)
	PVER	(13)	21	BONVS	(8)	METVO	(7)
7	FACIO	(12)		CAPVT	(8)	NVMMVS	(7)
	HIC (adv.)	(12)		EXCVTIO	(8)	PECTVS	(7)
	(IVPPITER)	(12)		FRANGO	(8)	PINGVIS	(7)
	VOLO	(12)		GRANDIS	(8)	POSSVM	(7)
11	POPVLVS	(11)		OMNIS	(8)	QVAERO	(7)
	VIDEO	(11)		POSCO	(8)	VETVS	(7)
13	MOS	(10)		VERBVM	(8)	VNGO	(7)
	PALLEO	(10)	29	AVRIS	(7)	VOTVM	(7)
	SCIO	(10)		CREDO	(7)		

Nous pouvons dire que ces lemmes représentent les "mots-thèmes" de Perse (18), c'est-à-dire ceux autour desquels s'organise sa pensée. Mais il serait vain de vouloir tirer des conclusions spécifiques pour Perse des hautes fréquences de *RES*, *DO* ou *FACIO* par exemple, car ces mots sont fréquemment employés par les auteurs latins et ils font également partie des mots-thèmes de la liste de référence. Aussi est-il nécessaire d'effectuer la comparaison avec cette dernière.

Voici le début de la liste de référence :

1	OMNIS	11	MVLTVS	21	IVBEO
2	MAGNVS	12	DICO	22	TEMPVS
3	RES	13	ANIMVS	23	TERRA
4	FACIO	14	DIES	24	VNVS
5	POSSVM	15	VENIO	25	CAVSA
6	LOCVS	16	SVPERVS	26	PETO
7	VIDEO	17	FERO	27	DEVS
8	HABEO	18	MITTO	28	HOMO
9	PARS	19	ARMA	29	MANVS
10	DO	20	PRIMVS	30	AGO

Après confrontation, nous obtenons cette liste-réduite des lemmes de Perse qui ne se retrouvent pas dans la liste de référence :

VIVO	18	AVRIS	7	CARMEN	6
O	13	CREDO	7	CENTVM	6
PVER	13	EXEO	7	COQVO	6
HIC (adv.)	12	FAS	7	CRASSVS	6
(IVPPITER)	12	LOQVOR	7	DISCO	6
MOS	10	METVO	7	DONO	6
PALLEO	10	NVMMVS	7	DVCO	6
SCIO	10	PECTVS	7	FIO	6
PONO	9	PINGVIS	7	INQVIO	6
CAPVT	8	VNGO	7	NOSCO	6
EXCVTIO	8	VOTVM	7	RADO	6
FRANGO	8	AIO	6	RIDEO	6
GRANDIS	8	ALBVS	6	RVMPO	6
POSCO	8	AVRICVLA	6	SAPIO	6
VERBVM	8	AVRVM	6	SVRGO	6

TANGO	6	FIGO	5	PES	5
TENER	6	GENIVS	5	PINGO	5
TOLLO	6	INGENS	5	PIPER	5
VERVM	6	INTVS	5	POETA	5
VETO	6	LAVDO	5	RECVSO	5
ACER	5	LIBER	5	SVBEO	5
AMICVS	5	LINGVA	5	VENA	5
CANTO	5	(MARCVS)	5	VITIVM	5
DEXTER	5	MISER	5	VOX	5
FERVEO	5	NIHILVM	5		
FIBRA	5	OCVLVS	5		

Cette liste-réduite qui représente une proportion importante (plus de 75 %) de la liste des cent un premiers mots-pleins des *Satires* nous offre donc des lemmes qui sont relativement répétés chez Perse (dix-huit fois pour le premier : *VIVO* et cinq fois pour le dernier : *VOX*) et qui sont soit des lemmes spécialisés, soit des lemmes rares puisqu'ils n'apparaissent pas dans la liste de référence. Nous avons là un outil très précieux pour la recherche des thèmes, surtout en effectuant des regroupements. Mais avant d'examiner quelques champs sémantiques, il est nécessaire de compléter la confrontation.

Le deuxième aspect de la comparaison avec la liste de référence devrait faire ressortir des lemmes présents dans cette dernière mais avec un écart de place significatif (14). En fait le calcul révèle que seul le verbe *DICO* est intéressant de ce point de vue, car il est le seul à présenter un écart réduit significatif. Deux autres lemmes seulement mériteraient d'être signalés, sans que l'on puisse toutefois en tirer des conclusions sûres : *VOLO* (É.R. = 2,79) (15) et *POPVLVS* (É.R. = 1,58) (16). Cette deuxième partie de la comparaison ne nous apprend

donc que relativement peu de choses, mais en reprenant la liste-réduite établie dans la première partie pour y regrouper les lemmes qui s'inscrivent dans le même champ sémantique, nous voyons que la haute fréquence particulière du verbe *DICO* s'inscrit à l'intérieur d'un champ sémantique très bien représenté dans cette liste par les substantifs et par les verbes : *VERBVM*(8), *AVRIS*(7), *LOQVOR*(7), *AIO*(6), *AVRICVLA*(6), *INQVIO*(6), *LINGVA*(5), *VOX*(5). Nous reviendrons sur ce champ sémantique avec l'étude du verbe *DICO*.

En examinant la liste-réduite, nous voyons que la plupart des autres champs sémantiques importants de Perse s'ordonnent autour du thème de la vie physique (17). Nous discernons ainsi cinq thèmes principaux : le "corps", le "gras", l' "énorme", la "cuisine", la "pâleur".

Les substantifs qui désignent concrètement le corps : *CAPVT*(8), *AVRIS*(7), *PECTVS*(7), *AVRICVLA*(7), *FIBRA*(5), *LINGVA*(5), *OCVLVS*(5), *PES*(5), *VENA*(5) forment un champ sémantique dont l'importance est étonnante. D'autant plus qu'il faut le compléter par des substantifs désignant le corps qui sont extérieurs à la liste-réduite et qui se trouvent en très grand nombre : *ARTICVLVS*, *NARIS*, *BARBA*, *BILIS*, *BVCCA*, *PILVS*, *VNGVIS*, *OS (ossis)*, *OS (oris)*, *CVTIS*, *PVLMO*, *LACTES*, *NERVVS*, *GAUSAPE*, *GVTTVR* (18),... L'examen du contexte montre un phénomène intéressant qui peut expliquer le privilège dont jouit ce champ sémantique à l'intérieur des *Satires* : l'association constante chez Perse de l'expression de la laideur physique et de celle de la laideur morale.

Le champ sémantique du "gras", représenté dans la liste-réduite par *VNGO*(7) et *CRASSVS*(6) est toujours lié à un contexte dépréciatif. Dans les *Satires*, ce thème qui s'applique souvent à la cuisine (mais quelquefois aussi au corps, aux objets : une fenêtre, le lit d'un mort, et d'une manière figurée au jour) doit

être complété par le champ sémantique des lemmes se rapportant à la cuisine, représenté dans la liste-réduite par les lemmes *COQVO*(6) et *PIPER*(5) et, dans la liste complète des *Satires*, par un grand nombre de lemmes de basse fréquence : *CAVLIS*, *VRTICA*, *FVMOSVM SINCIPVT*, *EXTA ANSERIS*, *OLLA*, *CENARE*, *SILIQVAE*, *POLENTA* ... Encore une fois l'examen du contexte indique un emploi nettement dépréciatif de ces lemmes, confirmé par la liaison avec le champ sémantique du "gros" et de l' "énorme", représenté dans la liste-réduite par trois adjectifs : *GRANDIS*(8), *PINGVIS*(7) et *INGENS*(5) (19).

Le thème de la pâleur, représenté dans la liste-réduite par *PALLEO*(10) et *ALBVS*(6) employés presque toujours dans un contexte dépréciatif est complété dans la liste entière des mots-pleins par des lemmes employés dans un contexte laudatif ou neutre comme *IMPALLESCO*, *CANDIDVS*, *CANVS*, *CANITIES* ...

Signalons enfin deux autres champs sémantiques qui ne sont pas directement liés au thème de la vie physique mais qui sont intéressants parce qu'ils sont uniquement formés par des verbes : les verbes exprimant une action vive : *EXCVTIO*(8), *FRANGO*(8), *RVMPO*(6), *SVRGO*(6), *TOLLO*(6), ou un état d'effervescence : *FERVEO*(5) et les verbes se rapportant au savoir ou à la sagesse : *SCIO*(10), *DISCO*(6), *NOSCO*(6), *SAPIO*(6). Le premier champ sémantique est lié à une présence continue de la violence verbale et physique dans les *Satires*, le second à l'importance du savoir et de la sagesse pour Perse qui les présente comme les seuls remèdes aux tares de la vie physique. Il est intéressant de constater que ce thème abstrait tient une place relativement importante alors que nous avons vu jusqu'à présent la prépondérance des thèmes concrets.

III. LE VERBE "DICO" DANS LES SATIRES DE PERSE

Deux éléments invitent à étudier de plus près l'emploi du verbe *DICO* dans les *Satires* de Perse. D'abord l'examen de la liste des lemmes absents de la liste de référence fait ressortir la présence importante de lemmes appartenant aux champs sémantiques de la parole et de l'audition (*VERBVM, AVRIS, LOQVOR, AIO, AVRICVLA, INQVIO, LINGVA, VOX*). Il s'agit donc de lemmes que Perse a privilégiés par rapport à un état moyen de la langue représenté par notre liste de référence. Certes ces huit lemmes ne sont pas tous sur le même plan et il est possible de faire des distinctions selon des critères sémantiques et syntaxiques. Cependant le fait qu'ils aient tous au moins un dénominateur commun et qu'ils fassent partie des lemmes de fréquence anormalement élevée nous invite à les rassembler momentanément.

Ensuite un des lemmes fondamentaux de ce champ sémantique, le verbe *DICO*, figure en tête de la liste complète des mots-pleins de PERSE. Il s'agit certes d'un verbe courant dans la langue latine mais nous avons vu que l'écart avec la fréquence théorique établie d'après la liste de référence était très significatif. Il est intéressant également de remarquer l'écart de place du verbe *AVDIO* : quarante-troisième rang chez Perse, soixante-douzième rang dans la liste de référence. Mais l'écart est bien moins significatif car *AVDIO* n'est employé que six fois chez Perse.

En nous appuyant sur ces deux éléments, nous voudrions pousser encore les recherches dans deux directions. Nous avons jusqu'ici effectué les confrontations avec la liste de référence. Or cette liste qui nous présente un état moyen et "banalisé" de la langue manque de précision pour une étude approfondie, et

il peut être utile de comparer Perse individuellement avec d'autres auteurs. D'autre part nous pouvons également faire des comparaisons à l'intérieur même du texte de Perse en utilisant ce dernier comme norme. Ainsi pourrions-nous chercher quel est le statut syntaxique du verbe *DICO* par rapport à celui des autres verbes puisque nous disposons de comptages grammaticaux très précis sur toutes les catégories de lemmes.

Pour la confrontation avec d'autres auteurs, nous utiliserons quatre oeuvres en prose et quatre oeuvres en vers parmi celles dont nous avons les listes de fréquence : Plaute, *Amphitryon*, César, *De Bello Gallico*, Salluste, *Catilina*, Sénèque, *Consolatio ad Helviam matrem*, Virgile, *Géorgiques*, Tibulle, *Elégies*, Ovide, *Métamorphoses*, Juvénal, *Satires*. Il s'agit donc de vérifier individuellement avec chacun de ces auteurs, choisis dans des genres divers et à des époques différentes, si la haute fréquence du verbe *DICO* est bien un phénomène propre à Perse comme nous l'a montré la comparaison avec la liste de référence. Pour cela nous calculons dans chaque oeuvre la fréquence relative sur cent mots du verbe *DICO* selon la formule :

$$\frac{\text{Fréquence réelle}}{\text{Nombre de mots}} \times 100$$

Plaute	:	1,01	Virgile	:	0,15
César	:	0,23	Tibulle	:	0,34
Salluste	:	0,30	Ovide	:	0,46
Sénèque	:	0,14	Juvénal	:	0,32
PERSE	:	0,55			

Il semble donc, au vu de cette confrontation et si l'on excepte la pièce de Plaute,

que l'emploi du verbe *DICO* soit un phénomène particulier chez Perse. Cela confirme parfaitement les résultats obtenus uniquement avec la liste de référence, puisqu'en choisissant huit oeuvres aussi différentes de genre et d'époque, et dont la moitié (Plaute, Sénèque, Tibulle, Juvénal) ne fait pas partie de cette liste, nous n'en trouvons qu'une qui dépasse les *Satires* pour la fréquence de *DICO*, les autres présentant une fréquence relative beaucoup plus faible. Nous reviendrons un peu plus loin sur le rapprochement de Plaute et de Perse mais avant de tenter quelques éléments d'explication, passons au deuxième type de confrontation, celle effectuée à l'intérieur du texte de Perse.

Comme nous connaissons la distribution des modes pour l'ensemble des verbes, il est intéressant de voir si le verbe *DICO* s'écarte de la distribution moyenne et, dans ce cas, sur quels points (20). La confrontation révèle que pour *DICO*, Perse utilise de préférence les modes subjonctif et impératif, et beaucoup moins le mode indicatif. Ce phénomène apparaît mal dans l'observation directe des données car les chiffres réels ne révèlent pas de différences spectaculaires entre chaque catégorie : 6 pour l'indicatif, 8 pour le subjonctif et 5 pour l'impératif. Mais le calcul nous montre ici qu'en suivant le modèle théorique de distribution nous aurions pu attendre environ deux fois plus d'occurrences à l'indicatif et deux fois moins au subjonctif et à l'impératif.

L'examen du texte fait apparaître que tout ce qui vient d'être vu correspond à un trait du style de Perse qui donne fréquemment à son oeuvre la forme d'un dialogue, ce qui pourrait expliquer le rapprochement avec Plaute pour la fréquence relative de *DICO*. Au lieu de raconter simplement ou d'attaquer sur un plan moral et abstrait, il met des personnages en scène et fait vivre le texte par un jeu continu de répliques. La haute fréquence du pronom personnel *TV* montre

l'importance de l'interlocuteur dans les *Satires*. En effet ce pronom, avec ses soixante-huit occurrences arrive en cinquième position dans la liste de fréquence complète (mots-outils compris) alors qu'il est absent de la liste de référence. Tout cela explique l'utilisation de *DICO* et de ses synonymes (21) pour introduire les répliques, et également la présence de *AVDIO*. Ces verbes entraînent dans leur sillage les substantifs désignant les organes ou les moyens de la parole et de l'audition : *LINGVA, VOX, VERBVM, AVRIS, AVRICVLA* ... Et ceci d'autant plus facilement que, comme nous l'avons vu pour le vocabulaire du corps, Perse privilégie le concret par rapport à l'abstrait. Ainsi préfère-t-il toujours une expression concrète avec un personnage qui parle à une pensée exprimée par l'auteur lui-même dans un vocabulaire nécessairement abstrait. Perse construit donc son texte un peu comme une représentation théâtrale, à cette différence près que ses dialogues sont fictifs. Il n'y a souvent pas de réplique réelle mais Perse en émet l'hypothèse, d'où l'emploi du subjonctif dans ce cas : I, 23 : *TVN...DICAS : "OHE" ?* (Est-ce que tu dirais : "ohé" ?); V, 171 : *SI VOCET, ... DICAS : "QVIDNAM IGITVR FACIAM ?"* Cf III, 56; III, 78; V, 158; III, 43; IV, 46; V, 189).

Deux types d'emploi du verbe *DICO* retiennent l'attention et représentent à eux seuls seize exemples sur les vingt-six.

Le premier (douze exemples) est l'emploi du verbe *DICO* comme terme introducteur d'un discours direct :

I, 28 : *AT PVLCHRVM EST DIGITO MONSTRARI ET DICIER : "HIC EST"*.
 (Mais il est beau d'être montré du doigt et d'entendre dire : "c'est lui !")
 Cf. I, 128; I, 23; V, 171; I, 56; VI, 58; III, 78.

Il est curieux de constater que, dans cinq exemples sur les douze, le verbe *DICO*

est placé après ce discours direct, tournure qu'il est impossible de transposer directement en français :

III, 88 : *"INSPICE, NESICIO QUID TREPIDAT MIHI PECTVS ET AEGRIS
FAVCIBVS EXSVPERAT GRAVIS HALITVS; INSPICE, SODES,"
QVI DICIT MEDICO...*

("Examine-moi, je sens une vague inquiétude dans la poitrine et de ma
gorge malade s'exhale une haleine forte, examine-moi, s'il te plaît",
celui qui parle ainsi au médecin ...)

Cf. V, 158; III, 42; IV, 2; V, 113.

Le deuxième type d'emploi est *DICO* à l'impératif suivi d'une interrogation directe. Il est caractéristique, d'une part de la préférence de Perse pour le discours direct : au schéma "Dis-moi si tu viendras demain", Perse préfère systématiquement le schéma "Viendras-tu demain ? Dis-le moi.", d'autre part d'une volonté marquée de provoquer un dialogue en forçant l'interlocuteur à parler :

VI, 51 : *AN PROHIBES ? DIC CLARE.* (Me le défends-tu, parle clairement !)

Cf. II, 69; IV, 3. L'exemple suivant, qui n'est pas une interrogation directe, se rapproche toutefois du schéma précédent :

II, 22 : *HOC IGITVR, QVO TV IOVIS AVREM IMPELLERE TEMPTAS,
DIC AGEDVM STAIO.* (Ce par quoi tu essaies d'influencer l'oreille
de Jupiter, va, dis-le à Staius !)

Le "dire" tient donc chez Perse une place prépondérante et les verbes *DICO*, *AIO*, *INQVIO*, *LOQVOR* ont un rôle tout à fait privilégié. Perse attache incontestablement beaucoup d'importance à la mise en scène et à la prédication.

NOTES

- (1) Nous définissons le vocabulaire comme l'actualisation par le texte d'un lexique virtuel. Les deux éléments sont très distincts et un auteur peut, bien que possédant un lexique virtuel très riche, n'en utiliser qu'une petite partie et avoir ainsi un vocabulaire pauvre; inversement il peut aussi utiliser pleinement un lexique virtuel peu étendu. Nous ne nous intéressons ici qu'au vocabulaire.
- (2) Les rapports entre les trois données citées posent un grand nombre de problèmes statistiques dont C. Müller fait l'inventaire dans *l'Initiation à la statistique linguistique*, Paris, Larousse, 1968, p. 156 sqq. Voici ce que nous entendons par "richesse relative du vocabulaire" : la nature du vocabulaire d'une oeuvre peut être théoriquement comprise entre deux extrêmes; la plus grande pauvreté théorique est un seul lemme employé M fois, la plus grande richesse théorique : M lemmes employés une seule fois chacun; nous essaierons de déterminer quelle est la place relative des *Satires* de Perse, par rapport à d'autres oeuvres, sur cet axe allant d'un pôle à l'autre.
- (3) Cf. P. Guiraud, *Les Caractères Statistiques du Vocabulaire*, Paris, P.U.F., 1954, p. 52-53 et *Problèmes et Méthodes de la Statistique Lexicale*, Paris, P.U.F., 1960, p. 84-90. P. Guiraud qui persiste dans ce dernier

ouvrage, après bien des recherches, à considérer cet indice comme "le plus simple et le plus précis" de tous ceux qu'il a eu l'occasion d'examiner, le considère toutefois comme utilisable seulement entre 10.000 et 50.000 mots environ. Nous pensons pour notre part que sa limite inférieure d'utilisation peut être abaissée jusqu'à 5000 mots si, au niveau de l'interprétation, nous tenons compte pour les textes compris entre 5000 et 10.000 mots d'une variation possible. La table-type établie par P. Guiraud dans les *C.S.V.* p. 52 nous montre du reste qu'entre 3000 et 40.000 mots-forts, c'est-à-dire entre environ 6000 et 80.000 mots, les variations de l'indice non corrigé ne dépassent pas 1,2. Tous les calculs de P. Guiraud portent sur la langue française et nous manquons malheureusement de l'équivalent pour la langue latine, mais nous verrons au cours de ce travail que la plupart des conclusions tirées par P. Guiraud ou C. Müller sur le français peuvent également s'appliquer au latin. E. Burnet dans un article intitulé *Le traitement des faits linguistiques et stylistiques sur ordinateur. Texte d'application : Giraudoux in Statistique et linguistique. Actes et Colloques n° 15*, Paris, Klincksiek, 1974, utilise un système plus compliqué et certainement plus précis pour comparer du point de vue de la richesse lexicale les différentes oeuvres de Giraudoux : il se sert de la formule binomiale (Cf. C. Müller, *o.c.*, p. 38 sqq) pour calculer le vocabulaire théorique puis il le compare avec les effectifs réellement observés. Mais nous avons refait les calculs avec l'indice L / \sqrt{M} et nous aboutissons à un classement qui, dans ses grandes lignes est semblable au sien.

- (4) Il est à noter que cet indice moyen est très proche de celui cité par P. Guiraud pour la langue française dans *Problème et méthodes ...* et

qui est de 22.

(5) - Les renseignements ont été pris dans :

- . C.R.I.D.E.L.A. (Centre de Recherches, Information et Documentation pour l'enseignement des Langues Anciennes), *Lexique de Base Latin*, 2e édition, Liège, 1974.
- . L. Delatte, E. Evrard, S. Govaerts, *Sénèque. Consolation à Helvia. Index verborum. Relevés statistiques*, La Haye, Mouton, 1963. (*Consolation à Marcia*, id., 1964).
- . M. Dubrocard, *Juvénal. Satires. Index verborum. Relevés Statistiques*, Hildesheim / New York, Olms, 1976.
- . A. Maniet, A. Paquot, *Plaute, Amphitryon. Index uerborum. Lexiques inverses. Relevés lexicaux et grammaticaux*, Hildesheim / New York, Olms, 1970.
- . S. Govaerts, *Le corpus Tibullianum*, La Haye, Mouton, 1966.

- Les textes marqués d'un astérisque n'ont pas été dépouillés en entier; la liste des extraits se trouve en annexe 1.

(6) H. Bardon, in *Latomus* XXXIV, 3, Juil.-Sept. 1975, va même jusqu'à parler, à propos des *Satires* de Perse, de procédés de "collage" comme en peinture; en effet la variété du vocabulaire reflète la variété de la structure des *Satires* où il ne faut pas chercher une unité de développement. Au contraire Perse procède par "collage" d'images non liées entre elles et très différentes de sujet.

- (7) La liste des textes dépouillés se trouve en Annexe 2.
- (8) Centre de Recherches, Information et Documentation pour l'Enseignement des Langues Anciennes, ... (o.c.).
- (9) Les noms propres sont exclus de cette liste.
- (10) Est appelée "fréquence-limite" le niveau de fréquence correspondant au moment où, sur la liste de fréquence en ordre décroissant, on atteint le pourcentage de 85 % des occurrences. Elle varie, bien entendu, suivant les oeuvres : 10 pour César, *Bellum Gallicum*, 5 pour Salluste, *Jugurtha*, 3 pour Virgile, *Bucoliques*... Cette structure, éliminant l'apparition de lemmes propres à un auteur ou à un contexte particulier est bonne pour dégager une sorte de norme ou fonds commun de la langue latine sur une période assez large. Par exemple le lemme *CASTRA, -orum*, qui a une fréquence 273 dans César, *Bellum Gallicum*, 40 dans Salluste, *Jugurtha*, 9 dans Virgile, *Enéide* et Salluste, *Catilina*, 3 dans Virgile, *Géorgiques*, 1 dans Virgile, *Bucoliques*, 0 dans Ovide, *Métamorphoses*... ne figure pas dans la liste. C'est un lemme propre à un auteur et surtout à un contexte particulier. Son caractère trop spécialisé l'exclut d'une liste de base.
- (11) En fait il n'est pas intéressant de se limiter strictement au chiffre de 80, car, à un certain niveau de la liste des mots-pleins des *Satires*, le rang n'a plus qu'une signification alphabétique : ainsi, du rang 70 au rang 101, tous les lemmes ont la fréquence 5. Ce sont donc les cent un premiers mots-pleins de la liste des *Satires* que nous comparerons aux

quatre-vingts mots-pleins de la liste de référence.

- (12) Le chiffre qui précède le lemme indique son rang dans l'ordre de fréquence décroissante des mots-pleins, celui qui suit, sa fréquence.
- (13) Cf. P. Guiraud, *C.S.V.*, (o.c.), p. 64 sqq.
- (14) Comme pour la première partie de la comparaison, nous procédons en sens unique : de même que nous n'avons pas relevé les mots de la liste de référence absents chez Perse, de même nous ne nous occupons pas des mots présents avec un écart négatif dans la liste de Perse, mais ce pourrait être également une voie de recherche intéressante. Nous utilisons, pour déclarer qu'un écart est significatif ou non, la même méthode que P. Guiraud pour sa recherche des mots-clé (Cf. *C.S.V.*, p. 65 sqq.) mais nous ne pouvons travailler que sur les seize premiers mots-thèmes, c'est-à-dire sur les seize premiers mots-pleins des *Satires* si nous voulons disposer de fréquence suffisamment élevées. Nous calculons d'abord leur fréquence théorique selon la formule :
$$\frac{(\text{Fréquence dans la liste de ref.}) \times (\text{Nombre de mots des Sat.})}{(\text{Nombre de mots de la liste de référence})}$$

puis l'écart absolu avec la fréquence réellement observée dans les *Satires*, et enfin l'écart réduit (= écart absolu divisé par la racine carrée de la fréquence théorique). C'est à partir de la valeur de l'écart réduit que nous pouvons considérer un écart comme significatif ou non, en consultant les tables d'interprétation en termes de probabilité de la valeur des écarts réduits (Cf. C. Müller, *Initiation à la Statistique linguistique* (o.c.)). Voici un exemple de calcul pour les deux premiers lemmes de la liste de Perse :

	Fréquence théorique	Fréquence réelle	Ecart absolu	Ecart réduit
<i>DICO</i> :	12,52	26	13,47	3,80
<i>RES</i> :	24,22	18	-6,22	-1,26

De ces résultats nous pouvons conclure que *DICO* présente un écart significatif et est donc à retenir (probabilité entre 0.001 et 0.0001) mais que *RES* a une fréquence à peu près attendue (probabilité entre 0.30 et 0.20).

- (15) Probabilité entre 0,01 et 0,001.
- (16) Probabilité entre 0,20 et 0,10
- (17) Sur l'expression du concret dans les *Satires* de Perse, cf. H. Bardon, *Perse et la réalité des choses*, in *Latomus*, XXXIV, 2, 1975, p. 319 sqq.
- (18) Pour l'étude des thèmes, il n'est pas intéressant de se limiter à l'examen des lemmes de fréquence élevée. Aussi nous servons-nous simplement de la liste-réduite comme d'un outil de recherche que nous complétons par l'étude de tous les lemmes de Perse, y compris ceux de fréquence un.
- (19) Comme *VNGO* et *CRASSVS*, ces lemmes sont quelquefois employés de manière figurée.
- (20) Voici un tableau dans lequel est présenté la comparaison entre le modèle théorique de la distribution de *DICO* et la distribution réelle. Ce tableau seul ne suffirait pas pour appuyer la démonstration à cause du

petit nombre d'occurrences de chaque catégorie, mais il peut être utilisé malgré sa fragilité statistique, comme une donnée parmi d'autres :

	IND.		IMP.		SUB.		PART.		ADJ. V.		INF.	
	M	R	M	R	M	R	M	R	M	R	M	R
	12,34	6	1,95	5	4,19	8	3,65	0	0,27	1	3,51	6
Ecart absolu	-6,76		3,05		3,81		-3,65		0,73		2,49	
Ecart réduit	-1,80		0,81		1,01		-0,97		0,19		0,66	

M = modèle théorique

R = chiffre réel

- (21) Une étude des fréquences relatives de *LOQVOR*, *AIO* et *INQVIO* révèle que Perse emploie généralement plus ces verbes que les autres auteurs.

ANNEXE 1

LISTE DES EXTRAITS :

- Tite Live, *Ab Vrbe Condita* : Préface; I, 19; II, 23-24, 27-30, 31-33, 54-58; III, 15; IV, 1-5, 59-60; V, 37-43; XXI, 30-38; XXII, 1-7; XXIII, 4-6; XXV, 24-31.
- Tacite, *Annales* : II, 42-43, 53-61, 69-73, 82-84; III, 1-18; IV, 1-15, 52-54; VI, 50-51; XIII, 1-5; 12-21; XIV, 1-10.
- Ovide, *Métamorphoses* : I, 1-150; II, 1-332; IV, 55-166, 604-803; VI, 146-312; VIII, 183-235, 611-724; X, 550-739; XI, 85-193; XV, 871-879.

ANNEXE 2

Liste des textes formant le corpus du lexique de base du C.R.I.D.E.L.A.

- Phèdre : *Fables* (extraits).
César : *Bellum Gallicum*.
Ovide : *Métamorphoses* (extraits).
Salluste : *De Coniuratione Catilinae*.
Bellum Iugurthinum.
Virgile : *Bucoliques*.
Géorgiques.
Enéide (six premiers chants).
Catulle : *Carmina*.
Cicéron : *De diuinatione* (extraits).
De Finibus (extraits).
De legibus (extraits).
De natura deorum (extraits).
De officiis (extraits).
De re publica (extraits).
Tusculanae disputationes (extraits).
Horace : *Odes*.
Sénèque : *Epistulae morales ad Lucilium*.
Tacite : *Annales* (extraits).
Histoires (extraits).
Tite Live : *Ab Vrbe Condita* (extraits).