

Quelques observations sur la fréquence des substantifs et des prépositions dans la littérature latine

Joseph DENOOZ

La fréquence relative des parties du discours a été étudiée à plusieurs reprises déjà par les spécialistes de la linguistique quantitative. Retenons, par exemple, les observations faites par P. Guiraud dans *Les caractères statistiques du vocabulaire*¹, celles de Ch. Muller dans ses deux ouvrages sur le théâtre de Corneille² et l'étude de l'œuvre de Giraudoux, publiée plus récemment par Ét. Brunet³.

Pour la littérature latine, si l'on excepte les relevés numériques réalisés par J. Collart en 1959⁴, c'est Ét. Évrard qui, le premier, a posé le problème en termes statistiques, dans une communication faite au colloque international de linguistique appliquée qui s'est tenu à Nancy en 1964. Etant donné que cette communication a été à l'origine de plusieurs travaux, je voudrais en rappeler la conclusion :

Je dirai seulement que, pour certaines catégories, les variations de fréquence paraissent bien être liées à des recherches de style. C'est ainsi que j'ai relevé, dans la *Consolation à Marcia*, un passage où substantifs et adjektifs abondent d'une manière significative, tandis que les verbes y sont exceptionnellement assez peu nombreux. En

¹ P. GUIRAUD, *Les caractères statistiques du vocabulaire*, P.U.F., Paris, 1954, pp. 38-40 et 89-90.

² Ch. MULLER, *Essai de statistique lexicale*, *L'Illusion Comique de Pierre Corneille*, Klincksieck, Paris, 1964, pp. 25-30, et *Etude de statistique lexicale*, *Le vocabulaire du Théâtre de Pierre Corneille*, Larousse, Paris, 1967, pp. 111-116.

³ Ét. BRUNET, *Le vocabulaire de Jean Giraudoux*, *Structures et évolution*, Thèse d'état, Nice, 1976, pp. 353 sgg.

⁴ J. COLLART, *Quelques observations statistiques sur les parties du discours*, dans *Revue des Etudes latines*, 37 (1959), pp. 215-229.

fait, il s'agit d'un passage qui contient un discours fictif écrit dans un style emphatique. On peut dès lors se demander si la fréquence des catégories grammaticales ne pourrait pas entrer dans la définition des différents types de style.

A la suite d'Ét. Évrard⁵, on retiendra notamment l'analyse du *Corpus Tibullianum* de S. Govaerts⁶, les recherches de M. Dubrocard⁷ sur Juvénal et notre étude sur le théâtre de Sénèque⁸.

Ces travaux ont montré que chez un même auteur, à l'intérieur d'une même œuvre, la fréquence des catégories présente des variations significatives, variations qui ont été interprétées non pas en termes d'écart par rapport à la norme d'une langue, mais en termes de différences révélant des passages qui s'écartent de la tendance générale de l'œuvre.

On n'a jamais effectué jusqu'à présent, sur de grands ensembles, une recherche dont l'objectif serait de dégager les variations liées au genre littéraire, à l'époque, à l'auteur ou à une œuvre déterminée. Et pourtant, une telle enquête paraît porter en soi le fondement sur lequel devraient reposer des analyses détaillées d'ensembles restreints.

C'est la raison pour laquelle nous présentons ici une étude sur la répartition des catégories grammaticales chez une vingtaine d'auteurs latins, poètes et prosateurs. Nos relevés sont extraits de la banque de données du L.A.S.L.A. dans laquelle chaque mot est accompagné notamment de son analyse morphologique complète.

Le tableau 1 donne le détail des auteurs et des œuvres que nous avons retenus. On y trouve, en onze colonnes, les effectifs observés pour chaque catégorie grammaticale et pour l'ensemble d'une œuvre, ainsi que la totalité des effectifs des différentes catégories. Notre enquête porte sur 1 006 966 mots parmi lesquels 677 377 appartiennent à des œuvres en prose et 329 589 à la poésie.

⁵ Ét. ÉVRARD, *La fréquence des phénomènes grammaticaux est-elle constante ?* dans *Actes du premier colloque international de linguistique appliquée*, Nancy, 1966, pp. 157-162.

⁶ S. GOVAERTS, *Le Corpus Tibullianum, Index Verborum et relevés statistiques. Essai de méthodologie statistique*, Mouton, La Haye, 1966.

⁷ M. DUBROCARD, *L'utilisation des catégories grammaticales. Aspects méthodologiques. Application aux Satires de Juvénal*, dans *Collectif des utilisateurs de machines à des fins d'information et de documentation*, (1972), pp. 35-36, *L'utilisation des catégories grammaticales dans les Satires de Juvénal. Essai d'analyse factorielle*, dans *Annales de la Faculté des Lettres et Sciences humaines de Nice*, 35 (1979), (Philologie, Littérature et Histoire anciennes), pp. 259-273.

⁸ J. DENOOZ, *Etude statistique sur la distribution des catégories grammaticales dans l'Hercule Furieux de Sénèque*, dans *Revue de l'Organisation internationale pour l'Etude des Langues anciennes*, 1974, pp. 25-49.

Tableau 1
Distribution des occurrences selon les catégories grammaticales
(Valeurs absolues)

	Subst.	Adj.	Nom.	Pron.	Verbes	Adv.	Prép.	C.c.	C.s.	Int.	Total
Ausone, Epigrammes	1199	395	54	481	966	259	88	241	108	6	3797
Catulle	3642	1695	122	1848	2821	1124	440	876	331	131	13030
César, B.G.	14351	2503	902	5888	10265	2868	3904	3240	1459	0	45388
Cicéron, Catilinaires	3192	844	72	2240	2560	1168	783	990	427	12	12266
Cicéron, De natura deorum	2303	581	111	1441	2008	949	465	830	303	5	6996
Cicéron, Pro Caelina	2322	506	59	1789	2462	1179	652	754	507	7	10217
Cicéron, Pro Rabirio	1155	261	10	554	612	289	248	247	124	0	3500
Cicéron, Pro lege Manilia	1822	547	74	1118	1249	613	459	554	187	3	6626
Cicéron, Tusc.	1498	573	45	1247	1509	730	396	650	289	7	6894
Histoire Auguste	13836	3352	650	5691	9009	4002	3331	3464	1624	9	45568
Horace, Epîtres	2812	1227	93	1093	2545	829	301	852	371	6	10129
Horace, Epodes	1024	474	16	264	653	201	95	255	70	22	3074
Horace, Odes	4803	2395	99	963	2654	782	412	1153	247	69	13582
Horace, Satires	3762	1837	160	1818	3414	1363	464	1030	589	45	14512
Juvénal	7944	2054	230	2835	5484	1704	831	2389	636	33	25140
Lucrèce	13469	4539	336	5102	11453	5338	2524	4887	1723	23	49554
Ovide, Art d'aimer, Rem., ...	6155	2252	164	2587	5698	1506	695	1478	558	61	21154
Ovide, Métam.	3268	1183	107	844	2448	603	456	1039	213	25	10186
Perse	1291	573	46	515	1149	373	154	358	139	42	4640
Plaute, Captifs	1384	606	49	1837	2187	1136	348	459	342	81	8429
Plaute, Casina	1130	491	37	1555	2030	1001	261	323	271	153	7252
Proprece	5843	2345	190	2809	4335	1524	692	1294	444	66	19542
Quinte-Curce	22293	4943	1096	8094	17504	8516	4264	5104	1782	37	71633
Salluste, Catilina	3451	935	56	1314	2152	840	706	822	326	5	10607
Salluste, Jugurtha	6567	1680	167	2515	4396	2060	1438	1843	630	5	21301
Sénèque, Cons. à Hélye	1652	597	65	1201	1512	641	408	431	199	9	6715
Sénèque, Cons. à Marcia	2106	751	76	1268	1924	767	434	644	216	6	8262
Sénèque, Cons. à Polybe	1193	461	36	1188	1324	572	287	416	157	6	5640
Sénèque, De beneficis	9227	3395	318	7856	12240	4635	2337	3288	1958	22	45276
Sénèque, De brev. Vitae	1393	587	57	1008	1470	629	387	395	171	11	6109
Sénèque, De clementia	2014	795	55	1205	1978	760	460	676	260	13	8216
Sénèque, De const. sapientis	1264	447	30	826	1259	473	291	534	183	3	5290
Sénèque, De Ira	5304	2149	149	3101	5614	2166	1196	1860	757	14	22410
Sénèque, De otio	407	132	26	360	473	197	140	144	70	0	1949
Sénèque, De providentia	1001	406	19	573	1070	368	238	271	131	3	4078
Sénèque, De tr. animal	1801	676	36	1102	1826	772	407	748	195	7	7568
Sénèque, De vita beata	1554	652	26	1204	1775	701	406	721	223	8	7270
Sénèque, Lettres à Lucilius	25468	9660	857	19743	31170	12044	6616	9209	3801	96	118664
Sénèque, Tragédies	20091	8132	418	6442	15017	3750	1577	3732	876	278	60314
Tacite, Agricola	2349	654	77	536	1274	630	376	732	148	0	6776
Tacite, Annales	33304	7955	658	6998	17888	6881	5452	8149	2642	14	90129
Tacite, Dialogue des Orateurs	2548	730	99	1271	1717	958	517	1196	287	13	9336
Tacite, Germanie	2026	497	53	475	966	591	335	662	133	0	5738
Tibulle	3710	1740	77	1521	2895	880	381	1024	319	53	12580
Tite-Live	5622	1264	211	1724	3734	1643	1285	1327	592	6	17408
Virgile, Bucoliques	1837	584	49	804	1188	454	204	401	129	45	5695
Virgile, Enéide	11071	3876	362	2740	6917	1993	1260	3408	418	123	32158
Virgile, Géorgiques	5100	2034	151	996	2935	1009	567	1746	254	29	14821
Vitruse	17887	5265	1519	4783	11385	4305	5375	5291	1717	0	57527
Totaux	290445	93330	10570	125365	231744	88802	55391	82241	29466	1612	1006966

Le tableau 2 donne, en pourcentage, la fréquence de chacune des catégories dans les quarante-neuf parties du corpus. Ce tableau fait apparaître, dans chacune des colonnes, d'importantes différences. Ainsi, les substantifs représentent seulement 15,58 % des occurrences dans Plaute, *Casina*, alors qu'on en trouve 36,95 % dans *Les Annales* de Tacite. Autrement dit, lorsque Plaute emploie 10 noms, Tacite en utilise 23 ou 24. Des écarts aussi marqués peuvent être observés pour chaque catégorie grammaticale. Le record est atteint par les prépositions qui représentent 9,34 % des occurrences du *De architectura* de Vitruve, contre 2,32 % des *Epigrammes* d'Ausone, ce qui signifie que là où Ausone a besoin d'une préposition, Vitruve en utilisera plus de quatre. À l'opposé de cette différence maximale, on trouve les verbes pour lesquels le chiffre le plus faible est 16,84 % dans Tacite, *Germanie*, et le plus élevé 27,99 % dans Plaute, *Casina*; le rapport entre les deux étant de 1,66, les verbes constituent la catégorie la plus stable.

Pour évaluer la manière dont les catégories se répartissent dans les textes de notre corpus, nous avons calculé pour chacune d'elles sa fréquence théorique et la valeur de χ^2 en utilisant pour norme les effectifs observés dans l'ensemble des œuvres. Nous connaissons de la sorte pour chaque catégorie, l'écart positif ou négatif entre l'effectif observé et la fréquence théorique ainsi que la valeur du χ^2 et pour chaque œuvre le χ^2 global. Je ne donnerai pas le détail des valeurs, et je me limiterai à une synthèse des résultats.

Les quarante-neuf χ^2 globaux que nous avons obtenus, conduisent à rejeter l'hypothèse nulle. Le χ^2 le plus faible est celui du *De Providentia* de Sénèque qui vaut 83, ce qui, pour un nombre de degrés de liberté égal à 8, donne une probabilité très inférieure à 0,00001. Dès lors, pour le phénomène étudié, qui se situe sans doute à un niveau inconscient ou, en tous cas, incontrôlé de la création littéraire, les écarts sont tels qu'on ne peut jamais les attribuer au seul hasard.

Pour tenter de mettre en évidence les éléments qui pourraient influencer les variations de fréquence observées, nous avons constitué un tableau pour chaque partie du discours.

Chacun des neuf tableaux reprend, pour une catégorie grammaticale déterminée, les valeurs de χ^2 de cette catégorie dans les différentes œuvres du corpus et les présente en fonction d'un classement qui répond à un double critère : le sens de l'écart observé entre effectif réel et effectif théorique, d'une part, les valeurs croissantes du χ^2 , d'autre part. De cette manière, nous obtenons des listes dans lesquelles apparaissent au début les œuvres qui présentent pour une catégorie, le déficit le plus significatif et à la fin les textes où la fréquence observée est la plus anormalement excédentaire.

Tableau 2
Distribution des occurrences selon les catégories grammaticales
(Pourcentages)

	Subst.	Adj.	N.n.	Pron.	Verbes	Adv.	Prép.	C.c.	C.s.	Int.
Ausone, Epigrams	31.58	10.40	1.42	12.67	25.44	6.82	2.32	6.35	2.84	0.16
Catulle	27.95	13.01	0.94	14.18	21.65	8.63	3.38	6.72	2.54	1.01
Cesar, B.G.	31.62	5.51	1.99	12.99	22.62	8.31	8.60	7.14	3.21	0.00
Cleéron, Catilinares	25.98	6.07	0.59	18.23	20.84	9.49	6.37	8.06	3.48	0.10
Cleéron, De natura deorum	25.60	6.46	1.23	16.02	22.32	10.55	5.17	9.23	3.37	0.06
Cleéron, Pro Cæcina	22.73	4.95	0.58	17.31	24.10	11.54	6.38	7.33	4.96	0.07
Cleéron, Pro Rabirio	33.00	7.46	0.29	15.83	17.49	8.26	7.09	7.06	3.54	0.00
Cleéron, Pro lega Manilia	27.50	8.26	1.12	16.87	18.85	9.25	6.93	8.36	2.82	0.05
Cleéron, Tuso.	21.73	8.31	0.65	18.09	21.69	10.50	5.74	9.43	3.47	0.10
Histoire Auguste	30.38	7.38	1.43	12.49	21.09	8.78	7.31	7.60	3.56	0.02
Horace, Epîtres	27.76	12.11	0.92	10.79	25.13	8.18	2.97	8.41	3.06	0.06
Horace, Epodes	33.31	15.42	0.52	8.59	21.24	6.54	3.09	8.30	2.28	0.72
Horace, Odes	35.38	17.63	0.73	7.03	19.54	5.78	3.03	8.53	1.82	0.51
Horace, Satires	25.92	12.68	1.10	12.53	23.73	9.39	3.20	7.10	4.06	0.31
Juvénal	31.60	12.15	0.91	11.28	21.81	6.78	3.31	9.50	2.53	0.13
Lucrece	27.18	9.38	0.68	10.30	23.11	10.69	5.09	9.68	3.48	0.05
Ovide, Art d'alher, Rem., ...	29.10	10.65	0.78	12.23	26.94	7.12	3.29	6.99	2.64	0.29
Ovide, Metam.	32.03	11.61	1.05	8.29	24.03	5.92	4.48	10.20	2.09	0.25
Perse	27.82	12.35	0.99	11.10	24.76	8.04	3.32	7.72	3.00	0.91
Plaute, Captifs	16.42	7.19	0.58	21.79	25.95	13.48	4.13	5.45	4.06	0.96
Plaute, Casina	15.53	6.77	0.51	21.44	27.99	13.60	3.60	4.45	3.74	2.11
Propreza	29.90	12.00	0.97	14.37	22.18	7.80	3.54	6.62	2.27	0.34
Quinte-Durce	31.12	6.99	1.53	11.30	24.44	9.10	5.95	7.13	2.49	0.05
Salluste, Catilina	32.54	8.81	0.53	12.39	20.29	7.92	6.66	7.75	3.07	0.05
Salluste, Jugurtha	30.83	7.89	0.78	11.81	20.64	9.87	6.75	8.65	2.98	0.02
Sénèque, Cons. à Hélvia	24.60	8.89	0.97	17.83	22.52	9.55	6.08	6.42	2.96	0.13
Sénèque, Cons. à Marcia	25.49	9.03	0.92	15.59	23.29	9.28	5.88	7.79	2.61	0.07
Sénèque, Cons. à Polye	21.15	8.17	0.64	21.06	23.43	10.14	5.09	7.38	2.78	0.11
Sénèque, De beneficis	20.33	7.50	0.70	17.35	27.03	10.24	5.16	7.26	4.32	0.05
Sénèque, De brev. vita	22.80	9.61	0.93	16.50	24.06	10.30	6.33	6.43	2.80	0.18
Sénèque, De clementia	24.51	9.68	0.67	14.67	24.07	9.25	5.60	8.23	3.16	0.16
Sénèque, De const. sapientis	23.89	8.45	0.57	15.61	23.60	8.94	5.50	10.09	3.03	0.06
Sénèque, De Ira	23.67	9.59	0.66	13.84	25.05	9.67	5.34	8.75	3.38	0.06
Sénèque, De otio	20.88	6.77	1.33	18.47	24.27	10.11	7.18	7.33	3.59	0.00
Sénèque, De providentia	24.55	9.96	0.47	14.05	26.24	9.02	5.79	6.65	3.21	0.07
Sénèque, De tr. animal	23.80	8.93	0.48	15.46	24.13	10.20	5.38	9.68	2.58	0.09
Sénèque, De vita beata	21.88	8.97	0.38	16.56	24.42	9.84	5.58	9.92	3.07	0.11
Sénèque, Lettres à Lucilius	21.46	8.14	0.72	16.64	26.27	10.15	5.58	7.76	3.20	0.08
Sénèque, Tragédies	33.31	13.48	0.69	10.68	24.90	6.22	2.61	6.19	1.45	0.48
Tacite, Agricola	34.67	9.65	1.14	7.91	18.80	9.30	5.55	10.80	2.18	0.00
Tacite, Annales	38.95	8.83	0.95	7.75	19.65	7.63	6.05	9.04	2.93	0.02
Tacite, Dialogue des Orateurs	27.29	7.82	1.06	13.61	18.39	10.26	5.54	12.81	3.07	0.14
Tacite, Germanie	35.31	8.66	0.92	8.28	16.84	10.30	5.84	11.54	2.32	0.00
Tibulle	29.49	13.83	0.81	12.03	23.01	6.84	3.03	8.14	2.54	0.42
Tite-Live	32.30	7.26	1.21	9.90	21.45	9.44	7.38	7.62	3.40	0.03
Virgile, Bucoliques	32.26	10.25	0.88	14.12	20.88	7.97	3.58	7.04	2.27	0.79
Virgile, Enéide	34.43	12.05	1.13	6.52	21.51	6.17	3.92	10.60	1.30	0.38
Virgile, Géorgiques	34.41	13.72	1.02	6.72	19.80	6.81	3.83	11.78	1.71	0.20
Vitrue	31.03	9.15	2.64	8.31	19.79	7.48	9.34	9.20	2.93	0.00

Le tableau 3 donne le classement auquel nous avons abouti pour les substantifs. Il permet de voir que les *Lettres à Lucilius* de Sénèque présentent, pour cette catégorie, le déficit le plus significatif et *Les Annales* de Tacite l'excédent le plus remarquable.

Peut-être convient-il, pour éviter tout malentendu, d'insister sur le fait que les signes qui accompagnent les valeurs de χ^2 n'ont aucune signification mathématique. Ils ont simplement pour effet de rendre le classement plus explicite en répartissant les œuvres en fonction du sens de l'écart entre effectif réel et effectif calculé.

Bien qu'une étude approfondie des différents tableaux mériterait d'être faite, nous avons choisi de porter notre attention sur deux d'entre eux : celui des substantifs, parce qu'il concerne la catégorie la plus importante et celui des prépositions pour lesquelles nous avons remarqué précédemment que les pourcentages d'emploi présentaient proportionnellement les écarts les plus nets.

Commençons par les substantifs. Dans la partie inférieure du tableau, nous trouvons les auteurs et les œuvres qui utilisent un nombre de substantifs supérieur à la fréquence théorique. Si l'on fait abstraction du *Pro Rabirio* de Cicéron, on y rencontre la plupart des poètes, ce qui n'est guère surprenant, et tous les historiens sans exception : Salluste, Tite-Live, César, Quinte-Curce, Tacite et même une œuvre aussi problématique et disparate que l'*Histoire Auguste*.

Quintilien, au livre X, chapitre 1, de l'*Institution oratoire* notait déjà cette proximité des poètes et des historiens lorsqu'il écrivait :

Car l'histoire a beaucoup d'affinité avec la poésie et n'en diffère qu'en ce qu'elle n'est pas assujettie à la mesure. Elle se propose de narrer et non de prouver. Ce n'est point un débat actuel, un combat présent qu'elle engage, c'est un récit qu'elle transmet à la mémoire de la postérité. C'est pourquoi elle prévient, par la hardiesse des expressions et des figures, l'ennui inséparable des longues narrations.

Nos calculs étaient les affirmations de Quintilien en montrant, pour un critère précis, la proximité de la prose historique et de la poésie; la raison en est évidemment que l'une et l'autre relèvent d'un style narratif et descriptif.

La présence de Vitruve dans le bas du tableau se comprend aisément dans la mesure où le *De Architectura*, même s'il n'a aucune prétention littéraire, est écrit dans une langue où abondent les termes techniques et les descriptions précises et détaillées.

Le tableau 3 suggère encore d'autres commentaires.

Ainsi, peut-on attribuer au seul hasard le regroupement d'Ovide, *Art d'aimer*, *Remèdes à l'amour* avec les élégiaques Tibulle et Properce, qui tous trois se rejoignent au milieu de notre liste avec des valeurs de χ^2 très proches

Tableau 3
Substantifs

SÉNÈQUE	<i>Lettres à Lucilius</i>	2 241,51	(-)
SÉNÈQUE	<i>De beneficiis</i>	1 124,45	(-)
PLAUTE	<i>Captifs</i>	450,93	(-)
PLAUTE	<i>Casina</i>	442,37	(-)
SÉNÈQUE	<i>De Ira</i>	208,17	(-)
SÉNÈQUE	<i>De vita beata</i>	140,61	(-)
CICÉRON	<i>Pro Caecina</i>	132,55	(-)
CICÉRON	<i>Tusc.</i>	120,77	(-)
SÉNÈQUE	<i>Cons. à Polybe</i>	115,77	(-)
SÉNÈQUE	<i>De brev. vitae</i>	77,28	(-)
SÉNÈQUE	<i>De tr. animi</i>	66,85	(-)
SÉNÈQUE	<i>De clementia</i>	53,48	(-)
LUCRÈCE		47,50	(-)
SÉNÈQUE	<i>De const. sapientis</i>	44,98	(-)
HORACE	<i>Satires</i>	42,95	(-)
SÉNÈQUE	<i>De otio</i>	42,75	(-)
SÉNÈQUE	<i>Cons. à Helvia</i>	41,93	(-)
CICÉRON	<i>Catilinaires</i>	34,96	(-)
CICÉRON	<i>De natura deorum</i>	32,86	(-)
SÉNÈQUE	<i>Cons. à Marcia</i>	32,20	(-)
SÉNÈQUE	<i>De providentia</i>	26,04	(-)
TACITE	<i>Dialogue des Orateurs</i>	7,81	(-)
CICÉRON	<i>Pro lege Manilia</i>	4,14	(-)
HORACE	<i>Epîtres</i>	4,14	(-)
CATULLE		3,58	(-)
PERSE		1,65	(-)
OVIDE	<i>Art d'aimer, Rem., ...</i>	0,46	(+)
TIBULLE		1,81	(+)
PROPERCE		7,53	(+)
AUSONE	<i>Epigrammes</i>	9,88	(+)
CICÉRON	<i>Pro Rabirio</i>	20,82	(+)
HORACE	<i>Epodes</i>	21,16	(+)
VIRGILE	<i>Bucoliques</i>	22,91	(+)
SALLUSTE	<i>Jugurtha</i>	29,12	(+)
	<i>Histoire Auguste</i>	36,54	(+)
OVIDE	<i>Métamorphoses</i>	37,07	(+)
SALLUSTE	<i>Catilina</i>	50,23	(+)
JUVÉNAL		66,23	(+)
TITE-LIVE		71,94	(+)
TACITE	<i>Agricola</i>	79,85	(+)
TACITE	<i>Germanie</i>	83,17	(+)
VITRUVE		100,91	(+)
CÉSAR		121,07	(+)
QUINTE-CURCE		128,75	(+)
VIRGILE	<i>Géorgiques</i>	159,21	(+)
HORACE	<i>Odes</i>	199,90	(+)
VIRGILE	<i>Enéide</i>	347,35	(+)
SÉNÈQUE	<i>Tragédies</i>	417,18	(+)
TACITE	<i>Annales</i>	2 054,43	(+)

les unes des autres ? Quelle en est la raison ? Est-ce parce que, contemporains, ils ont fréquenté le même cercle littéraire ? Est-ce parce que le thème principal des trois œuvres est l'amour ? Est-ce l'influence déterminante du genre élégiaque ?

Il est intéressant de noter qu'Ovide lui-même était conscient de respecter les règles de l'élégie lorsqu'il écrit, aux vers 395 et 396 des *Remèdes à l'amour* :

Et l'élégie confesse ne pas m'être moins redévable
Qu'à Virgile, la noble épopée⁹.

Ces deux vers permettent sans doute de comprendre pourquoi *Art d'aimer* et *Les Remèdes* sont plus proches de Tibulle et de Properce que des *Métamorphoses*, l'autre œuvre d'Ovide que nous avons étudiée et qui raconte les légendes grecques sous une forme épique.

Dans la partie supérieure du tableau 3, se trouvent les œuvres dans lesquelles les substantifs sont déficitaires. Il s'agit principalement de Cicéron et de Sénèque. Parmi les œuvres de ces derniers, nous trouvons Plaute, Lucrèce et *Les Satires* d'Horace.

Le fait que *Captifs* et *Casina* se classent en troisième et en quatrième positions n'est guère surprenant : dans les comédies, les répliques vives et pleines de verve sont calquées sur la conversation familière et vulgaire où les mots grammaticaux, pronoms, adverbes et interjections sont très fréquents. A cet égard, il est significatif que dans *Casina*, 2,11 % des occurrences soient des interjections alors que cette catégorie ne représente que 0,4 %, c'est-à-dire cinq fois moins, dans les tragédies de Sénèque.

En ce qui concerne Lucrèce, la place qu'il occupe au milieu des œuvres de Sénèque le Philosophe, ne s'expliquerait-elle pas par le contenu du *De rerum natura* ? Ce poème, en effet, est d'abord un exposé du système scientifique et philosophique d'Epicure. C'est peut-être pour cette raison que la fréquence des substantifs y est conforme à l'usage de la prose philosophique.

N'est-ce pas pour la même raison thématique que *Les Satires* d'Horace se trouvent également parmi les œuvres de Sénèque et, en outre, plus proches de Lucrèce que des autres œuvres d'Horace, *Epîtres*, *Epodes* et *Odes*. Dans *Les Satires*, en effet, nous trouvons à la fois des leçons de morale, une parodie des stoïciens et des pièces philosophiques.

Revenons aux œuvres dans lesquelles les substantifs sont plus nombreux que prévus, pour examiner le classement de Virgile et de Tacite.

Pour le premier, nous observons la proximité des *Géorgiques* et de l'*Enéide*, proximité qui se confirme d'ailleurs dans toutes les catégories grammaticales, à

⁹ Ovide, *Rem.* 395-296.

Tantum si nobis elegi debere fatentur
Quantum Vergilio nobile debet epos.

l'exception des verbes, des adverbes et des prépositions. Quant aux *Bucoliques*, elles ne rejoignent jamais ni les *Géorgiques*, ni l'*Enéide*.

Les écarts traduisent-ils une évolution de Virgile vers une langue de plus en plus expressive ou bien l'influence des modèles grecs, influence plus sensible dans les *Eglogues*? A moins qu'ici encore ce ne soit le genre littéraire qui joue un rôle prépondérant dans la classification?

Les résultats obtenus pour Tacite méritent une analyse détaillée.

L'œuvre la moins riche en substantifs est le *Dialogue des Orateurs* qui se situe à la fois dans la même zone que Cicéron, *Pro lege Manilia* et qui est proche aussi des *Epîtres* d'Horace, de Catulle et de Perse.

Dans le bas du tableau, nous trouvons l'un à la suite de l'autre et parmi les historiens, l'*Agricola* et la *Germanie*. Enfin, à la dernière ligne, avec un excédent de substantifs tout à fait exceptionnel *Les Annales*.

Le classement et les valeurs de χ^2 coïncident à plusieurs points de vue avec ce que nous connaissons des œuvres de Tacite ou avec ce que les philologues en ont dit. Tout d'abord, les œuvres sont rangées en ordre chronologique : le *Dialogue*, pense-t-on, a été écrit en 81 et publié en 96; vient ensuite, en 98, l'*Agricola*, suivi la même année de la *Germanie*. Enfin, dans la dernière partie de sa vie, après avoir publié *Les Histoires* pour lesquelles nous n'avons malheureusement pas de relevé, Tacite, maître dans un style qu'il a porté à sa perfection, écrit *Les Annales*.

En ce qui concerne la langue et le style de Tacite, notre classement correspond étrangement à ce qu'en disent les philologues. Ainsi, ceux-ci ont parfois douté de l'authenticité du *Dialogue* parce que celui-ci est trop cicéronien. Or, les chiffres rapprochent ce texte des œuvres de Cicéron.

Néanmoins, ne doit-on pas tenir compte du fait que le *Dialogue* est aussi relativement proche des poètes et qu'il présente déjà, à l'état latent, certains traits caractéristiques de la prose tacitienne.

La critique a maintes fois souligné l'évolution de ces traits caractéristiques de l'*Agricola* jusqu'aux *Annales*, il faut bien reconnaître que les valeurs de χ^2 sont, à cet égard, très significatives.

Les écarts observés pour les trois œuvres historiques semblent montrer que Tacite, partant d'un niveau de langue propre à ce genre littéraire, a porté à leur paroxysme les particularités fondamentales de ce genre.

Le tableau 4, celui des prépositions, va confirmer dans une large mesure, les observations que nous venons de faire. En voici un rapide commentaire.

Tout d'abord, la distinction entre poésie et prose est très nette. *A priori*, cette observation n'a rien de surprenant, on a souvent mentionné la tendance que manifestent les poètes à éliminer les mots grammaticaux et tout

Tableau 4
Prépositions

SÉNÈQUE	<i>Tragédies</i>	913,53	(-)
JUVÉNAL		220,32	(-)
OVIDE	<i>Art d'aimer, Rem.,...</i>	188,97	(-)
HORACE	<i>Odes</i>	150,23	(-)
VIRGILE	<i>Enéide</i>	146,46	(-)
HORACE	<i>Satires</i>	139,79	(-)
TIBULLE		139,77	(-)
PROPERCE		136,45	(-)
HORACE	<i>Epîtres</i>	117,66	(-)
CATULLE		107,01	(-)
VIRGILE	<i>Géorgiques</i>	75,47	(-)
AUSONE	<i>Epigrammes</i>	70,05	(-)
PLAUTE	<i>Casina</i>	47,73	(-)
PERSE		40,00	(-)
VIRGILE	<i>Bucoliques</i>	37,96	(-)
HORACE	<i>Epodes</i>	32,40	(-)
PLAUTE	<i>Captifs</i>	29,00	(-)
OVIDE	<i>Métamorphoses</i>	19,31	(-)
LUCRÈCE		14,97	(-)
SÉNÈQUE	<i>De beneficiis</i>	9,52	(-)
CICÉRON	<i>De natura deorum</i>	1,82	(-)
SÉNÈQUE	<i>Cons. à Polybe</i>	1,71	(-)
SÉNÈQUE	<i>De Ira</i>	1,11	(-)
SÉNÈQUE	<i>De tr. animi</i>	0,19	(-)
SÉNÈQUE	<i>De const. sapientis</i>	0,00	(+)
TACITE	<i>Dialogue des Orateurs</i>	0,02	(+)
TACITE	<i>Agricola</i>	0,02	(+)
SÉNÈQUE	<i>De vita beata</i>	0,09	(+)
SÉNÈQUE	<i>De clementia</i>	0,14	(+)
SÉNÈQUE	<i>De providentia</i>	0,64	(+)
CICÉRON	<i>Tusc.</i>	0,76	(+)
TACITE	<i>Germanie</i>	1,14	(+)
SÉNÈQUE	<i>Lettres à Lucilius</i>	1,21	(+)
SÉNÈQUE	<i>Cons. à Marcia</i>	1,98	(+)
SÉNÈQUE	<i>Cons. à Helvia</i>	4,12	(+)
SÉNÈQUE	<i>De brev. vitae</i>	7,74	(+)
SÉNÈQUE	<i>De otio</i>	10,18	(+)
CICÉRON	<i>Pro Caecina</i>	14,41	(+)
CICÉRON	<i>Pro Rabirio</i>	15,67	(+)
CICÉRON	<i>Catilinaires</i>	16,94	(+)
CICÉRON	<i>Pro lege Manilia</i>	24,79	(+)
SALLUSTE	<i>Catilina</i>	25,95	(+)
QUINTE-CURCE		26,64	(+)
TACITE	<i>Annales</i>	49,22	(+)
SALLUSTE	<i>Jugurtha</i>	60,37	(+)
TITE-LIVE		111,62	(+)
	<i>Histoire Auguste</i>	270,83	(+)
CÉSAR		792,81	(+)
VITRUE		1 545,04	(+)

particulièrement les prépositions. Marouzeau le note dans son *Traité de stylistique latine* lorsqu'il écrit :

Par exemple, il est visible que les poètes, soucieux d'enfermer dans leurs vers "moins de mots que de sens", recherchent les constructions légères, réduites au minimum d'artifice grammatical, et disent volontiers : *se inferre campis* pour ... *in campos*, ...¹⁰

Indépendamment de la distinction entre poésie et prose, le tableau 4 présente plusieurs aires de cohérence.

Au centre, nous trouvons réunis les textes de prose philosophique de Sénèque et de Cicéron, avec pour seule interférence les trois œuvres mineures de Tacite qui, soit dit en passant, sont encore classées en ordre chronologique. Les écarts entre *De Beneficiis*, première œuvre du groupe et *De otio* qui ferme la série, sont relativement peu importants.

Après ces deux philosophes, nous trouvons regroupés, en une suite parfaite, les quatre discours de Cicéron que nous avons étudiés et, enfin, au bas du tableau, tous les historiens.

L'abondance de prépositions chez ces derniers s'explique par le fait que ces mots outils servent à introduire, en particulier, les multiples circonstances de lieu, de temps, etc. indispensables à la précision et à la clarté du récit historique.

Les observations faites à partir des substantifs et des prépositions devraient évidemment être enrichies par l'étude des autres catégories grammaticales.

Néanmoins, cette première approche révèle déjà quelques faits que je résumerai comme suit :

- a) la fréquence des parties du discours présente dans toutes les œuvres que nous avons examinées, des variations significatives par rapport à la moyenne générale de notre corpus;
- b) l'interprétation des écarts fait ressortir le rôle prépondérant du genre littéraire. Nous en avons trouvé des indices tels que la séparation de la prose et de la poésie, le comportement semblable des œuvres historiques, le regroupement des élégiaques;
- c) il est possible également que la thématique d'une œuvre influence certaines catégories. Ainsi, la fréquence des substantifs dans Lucrèce rapproche le *De rerum natura* de la prose philosophique;
- d) le style d'une œuvre doit également être pris en considération dans la mesure où les variations peuvent refléter une évolution de l'auteur : *Les Annales* de Tacite en sont une illustration.

¹⁰ J. MAROUZEAU, *Traité de stylistique latine*, Paris, Les Belles Lettres, 4^e éd., 1962, p. 211.

On remarquera que les résultats auxquels nous aboutissons ne peuvent en aucun cas être utilisés pour la critique d'authenticité. Celle-ci ne peut vraisemblablement reposer sur un critère aussi général que celui auquel nous nous référons. En outre, nous n'avons pas décelé chez les historiens, par exemple, de modifications liées à l'époque, et pourtant entre César et l'*Histoire Auguste* plusieurs siècles se sont écoulés.

Pour réaliser une étude approfondie de la répartition des occurrences par catégorie grammaticale, il conviendrait de travailler dans deux directions. La première consiste simplement à étendre le corpus à d'autres auteurs, à d'autres œuvres et à d'autres genres littéraires. La deuxième direction de recherche m'a conduit à appliquer aux relevés d'autres tests statistiques. C'est ainsi que nous avons mené une étude reposant sur la méthode d'analyse des données du professeur J.-P. Benzécri. Nous avons présenté les résultats de cette recherche dans un des numéros des *Cahiers* publiés par le Laboratoire de Statistique de l'Université Pierre et Marie Curie¹¹.

¹¹ J. DENOOZ, *Application des méthodes d'analyse factorielle à la fréquence des catégories grammaticales en latin*, dans *Les Cahiers de l'Analyse des Données*, 1988, 1, pp. 19-40.