

QUELQUES OBSERVATIONS  
A PROPOS DE LA COURBE CUMULATIVE DE LA  
DISTRIBUTION DU VOCABULAIRE CHEZ UN POETE LATIN : TIBULLE

Il y a quelques années, j'ai publié, dans le cadre des travaux du Laboratoire d'Analyse statistique des Langues anciennes, un index du *Corpus Tibullianum* et divers relevés statistiques relatifs à la grammaire, au vocabulaire et au style de cette oeuvre. \*

Je voudrais revenir ici sur un point particulier et montrer, à propos de l'oeuvre de Tibulle, l'intérêt que présente la courbe cumulative de la distribution du vocabulaire pour l'interprétation des structures générales du style d'un auteur. La technique employée pour l'établissement de cette courbe est identique à celle qui a été mise au point par l'équipe du L.A.S.L.A. pour l'étude du vocabulaire des trois *Consolations* de Sénèque : d'après le tableau de distribution du vocabulaire, les différentes classes présentant un même effectif sont réunies de manière à former des groupes distincts. Par exemple, il y a, chez Tibulle, 17 classes d'effectif 1, c'est-à-dire 17 mots différents qui, apparaissant dans le texte avec des fréquences différentes, fournissent 1473 occurrences; 6 classes d'effectif 2, c'est-à-dire 12 mots qui, ayant deux par deux la même fréquence, représentent 458 occurrences, et ainsi de suite jusqu'à une classe d'effectif 983, soit 983 mots qui n'apparaissent qu'une fois et fournissent donc 983 occurrences.

Les données sont cumulées, à la fois au niveau des effectifs et au niveau des occurrences, et reportées sur les coordonnées d'un graphique semi-logarithmique, les effectifs se trouvant en abscisse, les occurrences, en ordonnée.

---

\* S. GOVAERTS, *Le Corpus Tibullianum. Index verborum et Relevés statistiques. Essai de méthodologie statistique*, La Haye, MOUTON & Co, 1966.

On a pu constater, à propos des trois *Consolations* de Sénèque, que le résultat affectait l'allure générale, soit d'une droite, soit d'une ligne brisée formée de plusieurs droites de pente différente. L'accentuation de la pente est en proportion inverse de la richesse du vocabulaire. D'autre part, on observe que la première partie de la courbe correspond, dans le vocabulaire, à la zone des mots-outils et des mots-thèmes, la seconde partie à la zone des mots de caractérisation\*.

L'intérêt, sur le plan stylistique, de l'interprétation de telles courbes, proposée à titre d'hypothèse par L. DELATTE\*\*, m'a incitée à entreprendre une recherche parallèle à propos du vocabulaire de Tibulle.

Le graphique qu'on trouvera en annexe permet de juger des résultats de ces calculs. A côté de la courbe de Tibulle, j'ai reproduit une des courbes du vocabulaire de Sénèque, pour deux raisons d'ordre différent.

Tout d'abord, l'analyse et l'interprétation stylistique de telles courbes ne sont pas chose aisée. Il est évident qu'une comparaison met toujours beaucoup mieux en relief les caractéristiques particulières des résultats et permet de formuler des observations plus précises, en éliminant une certaine part d'arbitraire.

D'autre part, il m'a paru intéressant de rapprocher les distributions de vocabulaire d'un texte en prose et d'une oeuvre poétique. Bien entendu, ce n'est pas une comparaison limitée à deux auteurs qui permettra de formuler une

---

\* P. GUIRAUD, *Les caractères statistiques du vocabulaire*, Paris, 1954.

\*\* L. DELATTE, *Un Laboratoire d'Analyse statistique des Langues anciennes à l'Université de Liège*, dans le *Bulletin trimestriel de l'Association des Amis de l'Université de Liège*, 36<sup>e</sup> année, 1 (1964), pp. 16-18.

conclusion définitive quant à ces deux domaines de la langue. Mais l'expérience valait la peine d'être tentée : dans un domaine aussi neuf, les hypothèses de travail sont fécondes, même si, par la suite, d'autres observations obligent à y renoncer.

Parmi les trois oeuvres de Sénèque étudiées jusqu'à présent, j'ai choisi, comme point de comparaison, la *Consolation à Marcia*, qui présente les meilleures conditions pour l'étude que je me propose. En effet, le texte de Tibulle et celui de Sénèque sont de longueur à peu près égale, l'effectif total comptant respectivement 8142 et 8384 occurrences. Le vocabulaire de Sénèque comprend 2204 mots différents, celui de Tibulle 2041, d'où une moyenne d'emploi de 3,78 chez le premier auteur, de 3,98 chez le second. Cette dernière donnée révèle, chez Tibulle, un vocabulaire un peu moins riche, mais ne permet pas de préciser dans quelle mesure, car la moyenne est un indice d'interprétation difficile.

Quelle que soit son importance, cette caractéristique du vocabulaire de Tibulle apparaît nettement lors de l'examen du graphique, qui permet en outre de déterminer à quel niveau elle se manifeste.

Observons d'abord l'allure générale des deux courbes. Elle affecte chez Tibulle, comme chez Sénèque, la forme d'une ligne brisée, composée de deux droites de pente différente dont la première, rappelons-le, correspond à la zone des mots-outils et des mots-thèmes, la seconde à la zone des mots de caractérisation.

Notons encore que le point de départ de la courbe est situé plus près des origines chez Tibulle que chez Sénèque. Si l'on se reporte aux données numériques, on constate que les classes d'effectifs 1 et 2 groupent, chez

Sénèque, 41 mots qui fournissent 2743 occurrences, soit 32,72 % de l'effectif total; chez Tibulle, elles ne comptent que 29 mots qui fournissent 1931 occurrences, soit 23,72 % de l'effectif total. Au niveau des mots-outils, le vocabulaire de Tibulle est donc moins concentré et, par conséquent, plus riche que celui de Sénèque.

Etant donné que la moyenne d'emploi révèle chez Tibulle un vocabulaire dans l'ensemble moins riche que chez Sénèque, le caractère que je viens de souligner doit nécessairement être compensé par un appauvrissement à un autre niveau. C'est ce qu'indique, d'ailleurs, la pente de la courbe, nettement plus accusée chez Tibulle, à tel point que, malgré le niveau différent de l'origine, elle finit presque par rejoindre la courbe de Sénèque. Le rapprochement entre les deux courbes s'accroît à partir de la seconde zone. C'est donc au niveau des mots de caractérisation que le vocabulaire de Tibulle est plus pauvre, aussi bien en lui-même que par référence à Sénèque.

D'autre part, un examen plus attentif permet de déceler dans la courbe de Tibulle certaines particularités. Chacune des deux zones que je viens de définir est constituée à son tour par deux droites de pente différente. Il convient donc de distinguer dans le vocabulaire de Tibulle, non plus deux zones, mais quatre champs distincts, plus restreints, qui semblent être le reflet d'une structure du vocabulaire plus complexe, tant au niveau des mots-outils et des mots-thèmes qu'au niveau des mots de caractérisation. J'ai tenté, par une analyse plus détaillée de la courbe et en me reportant à la liste de fréquence, de traduire cette particularité sur le plan stylistique.

Le premier champ est constitué par une droite de pente assez douce. Il comprend 72 mots différents (soit 3,53 % du vocabulaire), fournissant 2837 occurrences (soit 34,85 % de l'effectif total). Ce sont la majorité

des mots-outils et quelques mots, de sens très général, qui peuvent leur être assimilés, comme *facio, dico, possum, multus*. Ce champ contient, en outre, de nombreux mots-thèmes et, en particulier, tous ceux qui sont en même temps des mots-clés.

Une droite de pente un peu plus raide, comprise entre le 72e mot, d'une part, et le 251e, d'autre part, constitue le second champ. Celui-ci groupe donc 179 mots (soit 8,77 % du vocabulaire total) qui fournissent 1753 occurrences (soit 21,53 % de l'effectif total). On y trouve encore, de loin en loin, un mot-outil, mais il contient surtout des mots-thèmes et, en particulier, ceux que nous qualifierions volontiers de mots-clés occasionnels. Il s'agit de mots fréquents, mais dont la dispersion n'est pas homogène, la plupart des occurrences étant concentrées dans quelques élégies. Dès lors, on peut les considérer comme mots-clés de ces élégies, mais non comme mots-clés de l'oeuvre. Au point de vue psychologique, ils ne traduisent pas un trait profond et constant de la personnalité de l'auteur; mais sont plutôt l'expression d'une réaction émotionnelle ou intellectuelle, vive, mais passagère, devant une situation donnée. L'appauvrissement du vocabulaire de Tibulle révélé par l'accentuation de la pente, dans ce second champ, est dû à la répétition de tels mots : tout se passe comme si l'auteur, pour exprimer le même concept, préférerait répéter un mot-thème plutôt que de recourir à un synonyme, choisi parmi les mots de caractérisation. Rappelons que, chez Sénèque, on observait l'attitude opposée, cet auteur évitant systématiquement la répétition des mots-thèmes.

Faut-il en conclure que Tibulle est moins sensible à de telles répétitions ou encore que c'est non seulement son vocabulaire, mais son lexique, qui est moins riche ? Je pencherais plutôt pour une troisième explication. On pourrait voir dans ce trait une manifestation inconsciente d'un phénomène

psychologique : le même concept s'exprimerait sous une forme identique parce qu'il revêt un caractère d'obsession, qui se manifeste même au niveau verbal.

Le 3e champ du vocabulaire de Tibulle est compris entre les limites du 251e et du 1058e mot. Il comprend 807 mots différents (soit 39,54 % du vocabulaire) et fournit 2569 occurrences (soit 31,55 % de l'effectif total). Il correspond à la quasi totalité des mots de caractérisation. La pente de la droite qui le constitue s'accroît à tel point que la courbe de Tibulle rejoint presque celle de Sénèque. En fait, elle irait même couper cette dernière si, brusquement, au niveau du 1058e mot, la pente ne s'écartait de la verticale. Je reviendrai dans un instant sur ce point. Ce 3e champ nous montre que la zone la plus pauvre du vocabulaire de Tibulle est celle des mots de caractérisation, jusqu'aux mots de fréquence 2 inclusivement.

Le dernier champ est constitué par une droite dont la pente s'oriente vers l'horizontale. Il correspond à la classe d'effectif 983, c'est-à-dire aux 983 mots employés une seule fois dans l'oeuvre. Il constitue 48,16 % du vocabulaire et fournit 12,07 % de l'effectif total des occurrences. Le brusque changement d'orientation de la pente révèle donc qu'il y a enrichissement du vocabulaire de Tibulle, au niveau des mots employés une seule fois. La proportion de ces mots est, par rapport à l'ensemble des mots de caractérisation, anormalement élevée. Sans doute faut-il voir là une autre manifestation, soit d'une répugnance à répéter les mots de caractérisation, soit d'une tendance à répéter certains mots-thèmes. En tout cas, ce fait semble bien confirmer l'hypothèse selon laquelle le lexique de Tibulle n'est pas moins riche que celui de Sénèque, par exemple, et qu'il faut plutôt interpréter la relative pauvreté de son vocabulaire au niveau des mots de caractérisation comme un fait de style, puisque cette caractéristique ne

joue plus au niveau des mots employés une seule fois.

En résumé, on peut dire que la courbe, chez Tibulle, est beaucoup moins régulière que chez Sénèque et que la structure du vocabulaire semble, chez lui, moins organisée, sujette à des fluctuations plus nombreuses. L'interprétation, sur le plan stylistique, de cette particularité est délicate. On pourrait peut-être l'attribuer à la manière dont l'oeuvre de Tibulle a été conçue et composée, c'est-à-dire élégie par élégie. Dès lors, ce n'est pas l'oeuvre entière qu'il faudrait considérer comme un tout cohérent, organisé en fonction d'une idée maîtresse, mais chacune des pièces prise séparément qui constituerait un ensemble.

Or, il semble évident qu'un auteur ne compose pas de la même manière une oeuvre de quelque 8000 mots, comme la *Consolation à Marcia*, par exemple, et une oeuvre qui n'en compte que 500, ce qui correspond à l'effectif moyen d'une élégie de Tibulle.

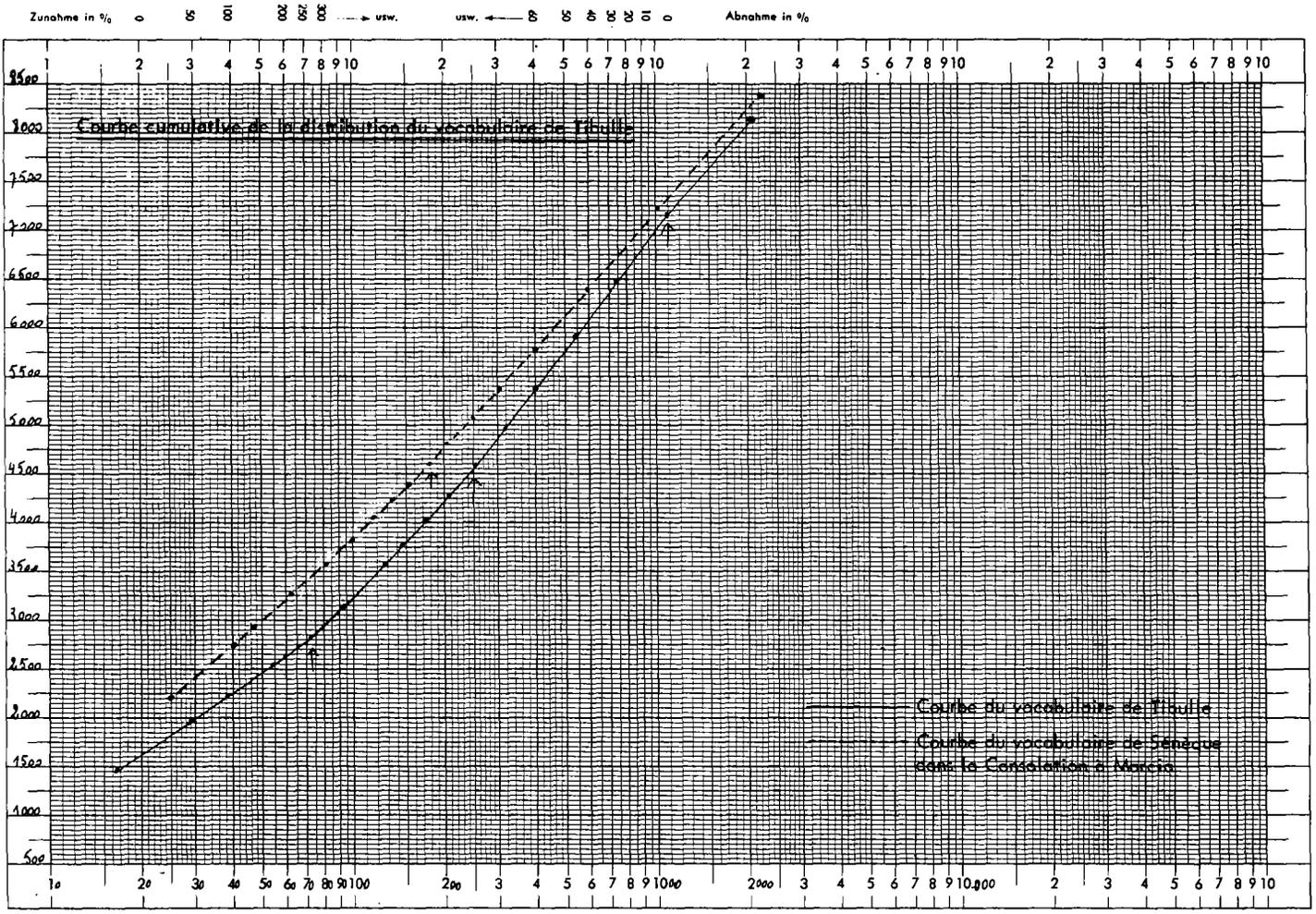
Quelle que soit la valeur de cette hypothèse, il faut se garder de tirer de la comparaison des courbes de Tibulle et de Sénèque des conclusions générales, relativement à la distribution du vocabulaire en prose et en poésie, ou même, de voir dans les différences observées des caractéristiques de style propres à chacun des deux auteurs. C'est seulement lorsque le chercheur disposera, dans ce domaine, d'une documentation abondante, qu'il sera à même de préciser ces divers points et de voir dans quelle mesure se confirme une hypothèse de travail.

Université de Liège  
Laboratoire d'Analyse statistique  
des Langues anciennes

S. GOVAERTS

TABLEAU DE DISTRIBUTION DU VOCABULAIRE

Fréquence	Effectifs	Nombre d'occurrences	Fréquence	Effectifs	Nombre d'occurrences
216	1	216	24	4	96
172	1	172	23	2	46
156	1	156	21	4	84
141	1	141	20	3	60
125	1	125	19	4	76
88	1	88	18	7	126
87	1	87	17	5	85
80	1	80	16	9	144
75	1	75	15	11	165
63	1	63	14	9	126
53	2	106	13	6	78
50	1	50	12	13	156
48	2	96	11	19	209
43	1	43	10	13	130
41	1	41	9	26	234
40	1	40	8	36	288
39	2	78	7	43	301
36	2	72	6	62	372
34	1	34	5	86	430
33	1	33	4	127	508
32	3	96	3	195	585
31	4	124	2	337	674
30	2	60	1	983	983
29	1	29		—	—
27	3	81		2041	8142



COPYRIGHT CARL SCHLEICHER & SCHÜLL-REINBECK HAN. Nr. 369<sup>1</sup>/<sub>2</sub>:6

Eine Achse logar. geteilt von 1 bis 10000, Einheit 62,5 mm, die andere in mm mit Prozentmaßstab