

Dans deux études remarquables consacrées à la statistique lexicale (1), Charles MULLER a étudié notamment certains aspects de la division en rôles dans le théâtre de Corneille. Ses travaux ont conduit à des résultats très intéressants et surtout à la mise au point de méthodes de travail. Nous avons l'intention d'utiliser certaines d'entre elles au cours d'une recherche sur les tragédies de Sénèque. Nous voudrions toutefois approfondir l'analyse au niveau philologique. C'est pourquoi nous avons mis au point un programme de fréquence par rôle (2) qui tient compte de la référence de chaque mot. Ceci permet de retourner facilement au texte et met par ailleurs en évidence le groupement des références selon les parties de l'oeuvre.

Nous avons utilisé comme point de départ le fichier mécanographique constitué par J. DENOOZ (3) à partir des travaux de fin d'études réalisés par des étudiants. Chaque carte contient les renseignements suivants :

colonnes	2 à 18	lemme
	19 à 38	forme
	41 à 43	code d'oeuvre
	44 à 47	numéro du vers
	48	indication éventuelle d'un vers double
	49 à 50	place du mot dans le vers
	51 à 53	place du mot dans la phrase
	54 à 63	codes d'analyse
	64	code du rôle

colonnes	65	code indiquant une variante dans l'apparat critique
	66	code de ponctuation
	76 à 80	numéro d'ordre dans l'index général des tragédies

Les autres colonnes sont vides.

De tous ces renseignements, nous ne gardons que le lemme, la référence, le rôle, le code de tradition manuscrite, la catégorie grammaticale ainsi que le contenu de la colonne 55, c'est-à-dire l'indication de la déclinaison, de la conjugaison, de la voix, du degré de comparaison des adjectifs et des adverbess. Nous croyons en effet que des tests ultérieurs peuvent indiquer si tel personnage emploie plus volontiers l'actif que le passif, le comparatif ou le superlatif que le positif ou bien si la distribution de ces sous-catégories est normale. De tels indices présentent une grande importance tant au niveau stylistique qu'au niveau psychologique.

Notre programme consiste en une série de comparaisons et de comptages de fréquences. Il comporte 163 instructions en langage PL 1 et fonctionne sur l'ordinateur 360/20 du Laboratoire d'Analyse statistique des Langues anciennes. Il fournit des cartes destinées à servir de données à d'autres programmes et il imprime une liste directement utilisable. Avant d'examiner ces résultats, il convient de décrire quelque peu les opérations techniques.

Le fichier général des tragédies est d'abord trié dans l'ordre des pièces afin de pouvoir traiter facilement et rapidement chaque oeuvre comme un

tout. Les enregistrements d'une même pièce se succèdent en ordre d'index, c'est-à-dire que les cartes se présentent dans l'ordre alphabétique des lemmes et, à l'intérieur de chaque lemme, dans l'ordre grammatical.

Lors du traitement du fichier, nous avons rencontré deux problèmes. Le premier concernait les mots pour lesquels existent des variantes critiques importantes, *crux*, mot ou leçon rejetés, mot impossible à analyser. Il était nécessaire, me semblait-il, de ne pas perdre toute trace de ces mots qui ne peuvent intervenir dans les comptages généraux. Ces cartes sont aisément repérées par un test sur la colonne 65 et, dès lors, l'ordinateur en imprime immédiatement le lemme, la référence, la catégorie grammaticale et le contenu de la colonne 55, le code de l'apparat critique et le rôle (4).

Le second problème provenait du manque d'uniformité dans la codification des rôles. Les premiers fichiers ayant été faits par différents chercheurs, les codes des rôles varient d'une pièce à l'autre, ce sont tantôt des chiffres, tantôt les premières lettres de l'alphabet, tantôt les dernières qui ont servi à les noter. Afin de faciliter la lecture et les calculs, une partie du programme donne automatiquement un nouveau code à chaque rôle, en comparant le code lu sur la carte avec le contenu d'une table préalablement initialisée. C'est ainsi que le rôle du chœur des *Troyennes*, qui est noté L dans le fichier, pour nous devient 12.

Le but du programme est de répartir et de compter les occurrences par rôle. Nous arrivons donc maintenant à la partie principale, les comparaisons et les comptages. La première opération concerne les lemmes. Les données

de la première carte lue sont transférées dans une zone où elles sont comparées à celles de la carte suivante; si les lemmes diffèrent, il se produit une variation d'indice. L'ordinateur ajoute 1 dans les trois compteurs, celui du nombre total de mots, celui du nombre d'occurrences du lemme et celui du nombre d'occurrences du lemme dans le rôle correspondant. La référence est envoyée dans une table qui est imprimée et remise à blanc dès qu'un nouvel enregistrement ne peut plus y trouver place. Ainsi, une référence se rapportant au rôle 12 ira occuper la 12e zone de la table, lorsqu'apparaîtra une seconde référence au rôle 12, le contenu de la table sera imprimé. La seconde opération concerne les catégories grammaticales, nous les testons afin de compter séparément les formes actives, passives et déponentes des verbes, les comparatifs et les superlatifs des adjectifs et des adverbes.

Lorsque toutes les occurrences d'un même lemme ont été traitées, l'ordinateur imprime la fréquence totale du mot ainsi que sa fréquence dans les différents rôles et il perfore les cartes-fréquence, dont le nombre peut atteindre trois pour le même mot (fréquence générale, fréquence du comparatif, fréquence du superlatif, par exemple). Voici le dessein de la carte-fréquence :

colonnes	1 à 17	lemme
	19	code indiquant le nombre de cartes pour ce lemme
	21	code précisant le type de la carte, ce code est 1 lorsqu'il s'agit des fréquences globales, 2 pour les verbes passifs et les comparatifs, 3 pour les déponents et les superlatifs
	23 à 25	nombre total d'occurrences du lemme
colonnes suivantes		nombre d'occurrences par rôle (5).

Enfin quatre chiffres numérotent la carte qui est encore distinguée par un code d'oeuvre correspondant à la place occupée par la pièce dans le manuscrit E.

La liste imprimée fournit, elle, le lemme, les numéros de rôle, les références pour chaque rôle, disposées en colonnes, le total général, le total par rôle et, dans le cas où ces compteurs ne sont pas vides, les totaux pour les compteurs particuliers des sous-catégories. Une page en est reproduite en annexe. Le classement des lemmes est alphabétique : outre la facilité de manipulation des références, cela nous permet de retrouver n'importe quel mot pour voir sa fréquence et le situer ainsi dans des listes de regroupement par fréquence que nous pouvons obtenir à partir des cartes. Les cartes peuvent aussi servir au calcul des vocabulaires réel et théorique, base d'un test du χ^2 .

Lorsque toutes les données ont été traitées, l'ordinateur imprime le nombre de mots de la pièce (6).

Ce programme a été conçu de manière à servir au calcul de fréquences reposant sur d'autres critères que les codes de rôles. Ainsi, il pourrait permettre d'étudier si tel mot apparaît plus souvent à une place déterminée dans le vers.

* * *

Examinons maintenant quelques utilisations possibles des résultats obtenus.

Après avoir classé le vocabulaire des personnages de *Thyeste* en ordre de fréquence, nous avons fait un premier calcul pour nous donner une idée très générale des phénomènes qui pouvaient apparaître : un rapport entre le nombre d'occurrences de chaque lemme dans un rôle et le nombre total d'occurrences de ce mot dans la pièce. Ainsi, nous voyons qu'*ego* a la moitié de ses occurrences dans le rôle de Thyeste, alors que ce rôle ne comporte que 21,94 % des mots de la pièce. Pour cerner de plus près la réalité, il faut étudier la probabilité donnée par le développement du binôme de Newton, c'est la méthode que nous allons utiliser dans cet article (7). Nous envisagerons la probabilité d'avoir au moins autant d'occurrences que le nombre d'occurrences rencontré. La probabilité d'avoir 17 occurrences d'*ego* dans le même rôle est de 0,00030; elle est nettement inférieure au seuil de 0,05 généralement admis comme significatif. C'est donc un mot intéressant à étudier. En nous reportant à la table des références, nous sommes frappés par la répartition des emplois de ce mot : beaucoup de références disséminées à travers toute la pièce pour Thyeste (rôle 1), dont deux dans le même vers, un peu moins pour Atrée (rôle 2), quatre pour l'ombre de Tantale (rôle 3), assez groupées, deux pour le messager (rôle 7) et une seule pour le chœur (rôle 8). Les probabilités pour Atrée, l'ombre de Tantale et les rôles où le mot n'apparaît pas ne sont pas significatives, mais pour le messager la probabilité d'avoir un aussi petit nombre d'occurrences d'*ego* est fort faible, elle ne dépasse pas 0,02788. Nous pouvons penser que celui-ci s'en tient à son récit, sans trop faire intervenir ses réactions. Nous parlerons du chœur plus loin.

D'un point de vue stylistique, les deux références chez Thyeste au vers 485 attirent l'attention : en nous reportant au contexte, nous voyons que la répétition est effectivement significative :

Pro me nihil iam metuo : vos facitis mihi Atrea timendum.

“ Pour moi je ne crains plus rien : c'est à cause de vous qu'à moi-même Atrée fait peur.”

De même, les références au rôle de l'ombre de Tantale : aux vers 81 et 82 *mihi* est en fin de vers et le *me* du vers 86 commence sa réplique suivante.

Les références, telles qu'elles nous sont données, permettent évidemment des études plus raffinées, par exemple sur l'emploi des pronoms dans la pièce ou chez tel personnage, ou encore sur la répétition de mots... Elles permettraient aussi, en s'attachant aux répliques plus longues, d'élargir le travail qu'a fait SEIDENSTICKER (8) sur la reprise de certains mots par les interlocuteurs dans un échange de répliques brèves.

D'autre part, la rareté d'*ego* dans le rôle du chœur attire l'attention, la probabilité d'en avoir un au plus est de 0,00021. Nous pourrions supposer que le chœur parle de lui-même au pluriel, reportons-nous donc à *nos* : trois occurrences seulement, mais groupées (v. 875, 878, 879). Le chœur déplore que la fin du monde arrive sur lui, c'est un emploi collectif qui désigne plutôt l'ensemble des humains. Voyons si l'emploi des possessifs nous renseignera davantage sur ce problème : *meus* n'apparaît qu'en 398 dans le même petit passage où se trouvait déjà *ego* (P = 0,00010 d'avoir un emploi ou

moins); *noster* (v. 621) est à nouveau un collectif ($P = 0,00891$ d'avoir un emploi ou moins). Le chœur répugne donc à se désigner comme une personne.

Pour le rôle de Thyeste, nous avons 15 emplois de *meus* sur un total de 37 dans la pièce, la probabilité d'en avoir autant ou plus est de 0,00857, donc fort faible. Les références sont dispersées, trois cependant sont groupées (999, 1001, 1001) avec de nouveau un effet de style :

“Quel est ce tumulte qui secoue mes entrailles ? ... et mon cœur gémit d'un gémissement qui n'est pas mien.”

Nous avons 20 emplois de *meus* chez Atrée ($P = 0,00015$ d'avoir ce chiffre ou plus), répartis de même manière que chez Thyeste; nous avons plusieurs références groupées dans la scène où il accueille son frère, comme si Sénèque voulait lui faire mieux affirmer son pouvoir et sa personnalité.

Dans les autres rôles, les probabilités sont normales, sauf dans le cas du messager où la rareté de *meus* s'explique par les mêmes motifs que celle d'*ego*.

Afin d'étayer ces résultats, nous pouvons aussi étudier la fréquence et la répartition de *noster* et de *nos* dans les rôles autres que le chœur. Le peu d'emplois de *nos* (6 dans toute l'oeuvre) ne permet toutefois pas d'en tirer beaucoup de conclusions, et la table indique qu'il a une répartition normale, confirmée par le test du binôme. Quant à *noster*, nous constatons une accumulation au début de la pièce, dans le rôle d'Atrée. Sa fréquence n'est pas

anormale, ($P = 0,07746$), ce qui est significatif c'est sa dispersion irrégulière et son sens particulier : 6 occurrences sur les 8 équivalent à *meus*, dont deux dans la scène d'accueil, cela renforce ce que nous avons observé précédemment. Pourquoi ensuite Atrée abandonne-t-il l'emploi de ce possessif ? Sans doute, l'utilisation du pluriel de majesté se justifie-t-elle davantage au début de la pièce : s'adressant à un courtisan ou à son frère revenant d'exil, Atrée peut avoir tendance à les traiter de haut, avec plus d'emphase. Ou bien la différence est-elle due à une habitude momentanée de Sénèque ou tout simplement au hasard ?

Pour le rôle de Thyeste, nous avons aussi une probabilité normale, mais le groupement est inverse : 4 occurrences à la fin de la pièce pour 1 au milieu. La première et la dernière équivalent au possessif du singulier. Il n'y a donc pas la même particularité de langage que chez Atrée.

Chez le courtisan, la probabilité atteint le seuil de signification ($P = 0,05462$). Le premier emploi est assez original : le courtisan s'identifie à son maître :

“Mais par quelles ruses sera-t-il pris et conduit dans nos rêts pour y mettre le pied ?” (v. 286-7).

Le second est un pluriel de majesté, encore à l'imitation de son maître ? car il suit juste un emploi similaire chez Thyeste (v. 333-334).

Le rôle du messenger ne comporte pas d'emploi de *noster*, ce qui est statistiquement anormal, mais se comprend toujours par le sujet : il n'a pas lieu

d'utiliser ce possessif.

Les autres rôles sont normaux, l'ombre de Tantale, roi aussi, utilise le pluriel de majesté.

Du point de vue de la construction de la pièce, sans toutefois vouloir émettre de conclusion générale à partir de ces quelques cas, nous pouvons dire que, parmi les trois rôles les plus longs, deux sont tenus par des personnalités bien marquées, mais que le chœur ne semble pas se présenter comme une personne. Son peu de personnalisation avait déjà été remarqué par les critiques. Sénèque accentue en cela la tendance d'Euripide en opposition avec la manière des deux autres grands Tragiques grecs.

L'emploi du pluriel de majesté et la fréquence de *meus, ego, noster* chez Thyeste et Atrée indiquent aussi un trait de caractère qui peut servir de point de départ à une analyse psychologique.

De plus, il serait intéressant de voir si cette fréquence se retrouve dans d'autres rôles de tyrans chez Sénèque.

Une telle comparaison dépasse le cadre de cet article. Nous pourrions toutefois examiner rapidement la fréquence des pronoms et des possessifs de la première personne chez un autre tyran, par exemple Lycus dans *Hercule furieux*. Nous constatons que leur fréquence n'a là rien d'anormal, mais si

nous nous reportons au texte, nous voyons que les 5 *noster* et les 2 *nos* sont à nouveau des pluriels de majesté. D'autre part, le groupement de ces mots dans le rôle d'Hercule attire notre attention. L'application du test binomial donne les probabilités suivantes : *ego* 0,00007; *meus* 0,00178; *noster* 0,03578; *nos* 0,44218. Les 13 emplois de *noster* dans ce rôle et un des deux emplois de *nos* sont des pluriels de majesté. La proportion entre les pluriels de majesté et les pluriels véritables est nettement plus élevée que dans les autres rôles, excepté celui de Junon. Le pluriel de majesté est bien employé conformément à son nom (9) : on le trouve dans la bouche du roi et de Junon. Cela rejoint les observations que nous avons faites à propos de Thyeste.

Nous pouvons dire que, pour les deux pièces examinées, l'affirmation de soi marquée par l'emploi des pronoms de la première personne n'est pas le fait du tyran en tant que tyran, mais celui d'un caractère que Sénèque veut présenter comme conscient de sa personnalité.

Prenons maintenant un cas de l'utilisation des résultats donnés par les compteurs spéciaux, par exemple celui des occurrences de la voix passive. Sur la liste de fréquence de chaque rôle, nous avons fait imprimer le total des emplois à la voix passive dans le rôle et dans l'oeuvre, et cela pour chaque verbe. *Timeo* a ainsi 42,11 % de ses occurrences dans le rôle de Thyeste, la probabilité d'avoir 8 occurrences sur 19 dans la même partie est déjà faible (0,03962), mais celle d'avoir dans ce rôle les 5 occurrences de la voix passive est évidemment bien plus faible encore (0,00052). Pourquoi Thyeste emploie-t-il ainsi le passif de ce verbe ? Y aurait-il là l'indice d'une préférence plus générale pour le passif ? Non, car la pièce

compte 284 occurrences au passif, 73 d'entre elles se trouvent dans le rôle de Thyeste; en appliquant ici le test du χ^2 , nous obtenons une valeur de 2,50 pour 1 degré de liberté, ce qui correspond à une probabilité comprise entre 0,10686 et 0,12134. Si nous regardons de plus près l'utilisation de *timeo*, nous voyons que les références sont groupées dans la scène où Thyeste se laisse convaincre par son fils de retourner près de son frère. Les passifs, sauf dans un cas, n'ont pas Thyeste pour sujet, ce sont des phrases du genre *id timendum est*, c'est Thyeste qui a peur.

Cette accumulation de références dans une même scène nous conduit à nous demander si Thyeste n'y emploie pas d'autres mots appartenant au même champ sémantique. Ici aussi la table des références peut nous aider : nous voyons en effet qu'il y prononce deux fois *timor*, deux fois *metuo*, une fois *metus*, une fois *pavere*.

Pour en revenir à la voix passive, nous constatons qu'elle a une fréquence trop élevée dans le rôle de l'ombre de Tantale (effectif réel = 22; effectif théorique = 14; $0,02535 < P < 0,02846$). Nous pouvons penser que Sénèque signifie bien par là que Tantale est sorti des Enfers contre son gré, qu'il est le jouet de l'action d'autrui.

En examinant ainsi à première vue les résultats fournis par les programmes pour une pièce, *Thyeste*, nous avons découvert quelques cas qui semblaient anormaux. Un examen plus poussé de ces cas nous a montré qu'effectivement la fréquence d'*ego*, *meus*, *nos*, *noster* dans les différents rôles n'est

pas aléatoire. De même, l'emploi de la voix passive contribue à renforcer un aspect de la situation où se trouve un personnage. En retournant au texte, nous avons été aussi amené à constater une abondance des pluriels de majesté. Nous avons alors fait quelques comparaisons avec une autre pièce, *l'Hercule furieux*, ce qui a renforcé nos observations, les emplois comparés se retrouvant dans les deux oeuvres. Ces résultats se prêtent à des interprétations plus poussées tant sur le plan du style que sur le plan de la psychologie.

Nous espérons ainsi avoir montré que les programmes que nous avons réalisés peuvent, déjà à ce stade, fournir des résultats intéressants.

Université de Liège
Laboratoire d'Analyse statistique
des Langues anciennes

Agnès GERMAIN

NOTES

- (1) C. MULLER, *Essai de statistique lexicale. L'illusion comique de P. Corneille*, Paris, Klincksieck, 1964 et *Etude de statistique lexicale. Le vocabulaire du Théâtre de P. Corneille*, Paris, Larousse, 1967.
- (2) Un programme de fréquence par partie avait déjà été fait par le Professeur Et. EVRARD, mais il ne tient pas compte de la référence des mots. Cf. *Revue de l'Organisation internationale pour l'Etude des Langues anciennes par Ordinateur*, 1967, n° 3, pp. 81-95. Ce programme a été amélioré depuis.
- (3) J. DENOOZ, *Les tragédies de Sénèque. Documents, Etudes, Recherches*, Doctorat en Philosophie et Lettres, Université de Liège, 1972. L'édition est celle de R. PEIPER, G. RICHTER, *L'Annaei Senecae tragaediae*, Leipzig, Teubner, 1867.
- (4) Nous n'avons pas tenu un compte spécial des conjectures ni des variantes retenues par l'éditeur.
- (5) Pour ne pas compliquer davantage, car ces instructions demandent beaucoup de tests et de boucles vu le type de l'ordinateur, nous n'avons pas perforé la catégorie grammaticale. C'est peut-être une erreur, mais elle est facile à réparer.

- (6) Les fichiers à traiter sont assez longs (plus de 11.000 cartes pour *Hercule sur l'Oeta*). Aussi un dispositif de cartes d'introduction permet-il d'interrompre et de reprendre l'exécution du travail.
- (7) Cf. *Tables of the Cumulative Binomial Probability Distribution*, dans *Annals of the Computation Laboratory of Harvard University*, 35, 1955.
- (8) B. SEIDENSTICKER, *Die Gesprächverdichtung in der Tragödien Senecas*, Heidelberg, 1969.
- (9) J. MAROUZEAU, *Traité de Stylistique latine*, 2e éd., Paris, 1946, p. 224.

		TRIMESTRE									
MEVS		1	2	3	4	5	6	7	8	9	
		0539	0527								
		1088	0194								
		1001	0252								
		0514	0541								
		0535	0918	0070							
		0448	0325								
		1043	0906								
		0483	1065								
			0247								
		1001	0523								
		0463	0243						0398		
		1076									
		1111	1096								
		0458	0281								
			0522								
			0237								
			0289								
			0912								
		0999	0971								
MEVS		0322	4	2	6	2					
MEVS		0526	4	2	6	2					
		0464	0322								
			0526								
37	1	15	20	1	0	0	0	0	1	0	
MICO		1	2	3	4	5	6	7	8	9	
			0184	0099	0049			0674	0825		
5	1	0	1	1	1	0	0	1	1	0	
MILES		1	2	3	4	5	6	7	8	9	
									0363		
1	1	0	0	0	0	0	0	0	1	0	
MINAE		1	2	3	4	5	6	7	8	9	
MINAE		0338	1	1	7	8					
			0291						0573		
2	1	0	1	0	0	0	0	0	1	0	
MINISTER	1	1	2	3	4	5	6	7	8	9	
			0326								
			0333			0308					
3	1	0	2	0	0	1	0	0	0	0	
MINOR		1	2	3	4	5	6	7	8	9	
			0290	0097							
								0736	0611		
								0705	0603		
6	1	0	1	1	0	0	0	2	2	0	
6	3	0	1	1	0	0	0	2	2	0	
MINOS	N	1	2	3	4	5	6	7	8	9	
				0023							
1	1	0	0	1	0	0	0	0	0	0	