

## Le vocabulaire caractéristique d'Harpagon

Roland Pelchat

Numéro 1, 1971

URI : <https://id.erudit.org/iderudit/800001ar>

DOI : <https://doi.org/10.7202/800001ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université du Québec

ISSN

0315-4025 (imprimé)

1920-1346 (numérique)

[Découvrir la revue](#)

Citer cet article

Pelchat, R. (1971). Le vocabulaire caractéristique d'Harpagon. *Cahier de linguistique*, (1), 9-31. <https://doi.org/10.7202/800001ar>

## LE VOCABULAIRE CARACTÉRISTIQUE D'HARPAGON

Robert Garapon, dans un article qu'il consacrait aux personnages du *Bourgeois Gentilhomme*, résumait en quelque sorte une tradition littéraire qui remontait jusqu'à Boileau lorsqu'il écrivait : "On répète communément que Molière n'a pas de style propre, mais qu'il prend le style de chacun de ses personnages<sup>1</sup>."

Qu'est-ce à dire sinon qu'on reconnaissait à l'auteur, même de son vivant, une capacité d'individualisation qui donnait à chacun de ses personnages une personnalité cohérente et naturelle. Cette caractérisation se traduirait-elle, entre autres moyens<sup>2</sup>, par des écarts lexicaux ? En d'autres termes, la caractérisation des personnages a-t-elle un répondant statistique ?

Poser comme hypothèse que le lexique d'une pièce de théâtre est l'urne dans laquelle est puisé le vocabulaire des personnages, c'est postuler que le choix des mots relève des lois du hasard.

La norme est alors la structure de la pièce dont les rôles ne sont que des échantillons. Et le style lui-même est défini à son tour comme un écart par rapport à cette norme.

- 
1. Robert Garapon, "la Langue et le style des différents personnages du *Bourgeois Gentilhomme*", dans *le Français moderne*, avril 1958, p. 103-113.
  2. Il serait possible de faire intervenir d'autres articulations comme les caractères d'ordre phonique, morpho-syntaxique, sémantique.

Puisque les échantillons formés par les rôles sont censés refléter la structure de l'ensemble, il s'agit de rechercher s'ils s'écartent de la norme. Le but du calcul statistique est d'étudier, à partir de cette norme, une structure théorique pour chacune des parties. Si le modèle concorde avec la réalité, nous dirons qu'il n'y a pas d'écart et que le caractère étudié est l'effet du seul hasard ; si, au contraire, des écarts sont signalés, il faudra demander aux lois du hasard à quel degré de probabilité ils correspondent. Il restera à préciser en quoi ces écarts correspondent à des faits de style.

La caractérisation lexicale semble se présenter de la façon suivante : des vocables se retrouvent exclusivement ou de façon anormale chez certains personnages et en particulier chez Harpagon. Mais comment vérifier cette impression par des moyens traditionnels ? Comme le fait remarquer G. Herdan, "ce qui apparaît intuitivement comme une relation de cause à effet n'est peut-être, considéré sous l'angle quantitatif, que le hasard"<sup>3</sup>. En d'autres termes, la répétition prise dans l'absolu peut n'être qu'illusion : elle ne prend un sens que par comparaison avec les répartitions possibles du vocable. On peut prévoir que cette répartition est fonction de la longueur relative des rôles, des sujets traités, mais aussi du degré d'individualisation qu'inconsciemment ou non l'auteur a prêté à ses personnages. Quand on a un index des fréquences, il n'est pas nécessaire de faire appel aux calculs des probabilités pour affirmer tant pour des raisons thématiques que numériques (il occupe 30% du texte de *L'Avare*), qu'Harpagon fait un emploi exagéré des termes relatifs à l'argent mais ce n'est pas toujours aussi simple. On a l'impression que des vocables s'iden-

---

3. Gustav Herdan, *Language as Choice and Chance*, Croningen, P. Noordhoff, 1956, p. 5.

tifient à certains personnages sans qu'on puisse justifier leur présence anormale ni par le sujet traité, ni par la longueur relative des textes.

De même, un examen minutieux du texte ne nous livre généralement pas les vocables qui sont absents dans un rôle. Est-ce un pur hasard, par exemple, que le mot *coeur*, qui apparaît pourtant 20 fois dans la pièce, n'est jamais utilisé par Harpagon ? Comment confirmer ou infirmer ce que l'intuition nous permet déjà en grande partie de sentir ?

#### Richesse du vocabulaire

Il n'est pas indifférent tout d'abord de comparer le vocabulaire d'Harpagon par rapport au lexique total de la pièce. La richesse lexicale est une notion tout à fait relative : elle signifie que deux échantillons étant de même taille, c'est celui qui aura le plus de mots différents qu'on considérera comme riche. Mais comment comparer le vocabulaire de fragments de longueurs différentes ? On sait que le nombre de mots différents ( $V$ ) est fonction du nombre total de mots ( $N$ ), mais nous ignorons selon quelle loi  $V$  progresse en fonction de  $N$ . Ce rapport n'est pas constant si l'on compare plusieurs fragments issus d'un même texte ; il n'est pas constant à l'intérieur d'un échantillon, l'auteur pouvant faire appel à un lexique plus étendu pour des raisons thématiques.

Nous avons mis à l'essai la méthode expérimentée par Charles Muller<sup>4</sup>. Le problème se résume à ceci : connaissant  $V$ , le nombre de mots différents de la pièce, peut-on, à partir de cette norme,

---

4. Charles Muller, *Essai de statistique lexicale*, Paris, Klincksieck, 1964, p. 63-76.

calculer une valeur théorique  $V'$  pour un rôle déterminé qui nous permettrait d'estimer l'étendue de son vocabulaire par rapport à l'ensemble de la pièce. Il s'agit en somme de calculer un effectif théorique et de le comparer avec la réalité. La prévision statistique sera le nombre de mots différents ( $V'$ ) que nous donnerait un tirage aléatoire des mots-occurrences du fragment ( $N'$ ) parmi tous les mots de la pièce ( $N$ ). Il suffira ensuite de mettre en regard la valeur réelle de  $V$ , de calculer l'écart et de l'interpréter par rapport à la norme fournie par l'ensemble des vocables de *l'Avare*.

Voici les estimations que les calculs nous fournissent. Nous donnons les résultats pour les principaux rôles afin de faciliter les comparaisons.

	Théorique	Réel	Écart
HARPAGON	1003	963	- 40
CLÉANTE	705	642	- 63
VALÈRE	627	576	- 51
FROSINE	568	557	- 11
M. JACQUES	458	396	- 62
LA FLÈCHE	346	330	- 16
ÉLISE	345	306	- 39
MARIANE	312	292	- 20

On sera peut-être surpris de constater que toutes les valeurs réelles soient inférieures aux effectifs théoriques. La tendance en moins observée chez tous les personnages ne vient que confirmer un fait relevé à plusieurs reprises par C. Muller : "Un mot employé une première fois dans une partie du texte a plus de chances d'y apparaître une seconde fois que d'apparaître ailleurs<sup>5</sup>". Dans

---

5. Charles Muller, *op. cit.*, p. 70.

l'établissement des valeurs théoriques, l'expérience confirme qu'on trouve plus fréquemment des écarts négatifs que l'inverse. Mais que toutes les valeurs réelles soient inférieures aux valeurs calculées mérite attention.

Envisageons l'hypothèse suivante : tous nos échantillons sont écrits dans une même langue, mais ils sont thématiquement ou stylistiquement différents au point de n'avoir en commun que les mots outils et un nombre restreint de mots de signification. Le vocabulaire de l'ensemble serait alors  $V_a + V_b \dots + V_n$  + le vocabulaire commun. Il est bien évident alors que toutes les valeurs réelles seraient inférieures à celles que laisserait prévoir le calcul. Considérant nos résultats, on peut sans hésitation parler d'indépendance lexicale des rôles, c'est-à-dire que les personnages, soit pour des raisons thématiques ou stylistiques, ou les deux à la fois, ont un vocabulaire bien à eux. Les calculs que nous venons d'effectuer reposent sur une distribution homogène des éléments du lexique, ils postulent l'hypothèse que le nombre de vocables est proportionnel à la longueur des échantillons en supposant que tous les vocables d'une fréquence donnée ont la même probabilité d'emploi dans tous les fragments. Il suffit que les fragments, par suite de servitudes lexicales relevant du sujet à traiter ou par choix personnel de l'écrivain, aient un vocabulaire spécifique pour que l'écart en moins soit possible.

Si l'on revient à notre dernier tableau, il n'est pas facile d'évaluer les écarts entre le modèle et l'observation. Sans prétendre à un classement, nous pouvons quand même nous faire une bonne idée de l'étendue du vocabulaire d'Harpagon si nous gardons à l'esprit que les rôles sont placés par ordre décroissant d'importance numérique. Malgré un déficit de 40, Harpagon nous semble avoir une langue assez riche, car des rôles plus courts comme ceux

de Cléante, Valère, Maître Jacques et même d'Élise ont des écarts plus considérables.

La même méthode qui nous a permis de calculer un vocabulaire théorique pour un rôle donné va nous permettre de préciser cette notion de pauvreté ou de richesse lexicale en l'appliquant aux catégories grammaticales.

Ainsi, sachant que le texte d'Harpagon occupe 0,301 de la pièce, on peut calculer le nombre de verbes qui devraient échapper à un tirage au sort. Le calcul en fait voir 246. Si les 536 verbes de *L'Avare* étaient répartis au hasard, il faudrait donc prévoir  $536 - 246 = 290$  verbes différents. Il s'en trouve en réalité 285. L'écart est très faible et ne saurait expliquer à lui seul la richesse lexicale du rôle.

En poursuivant ce raisonnement, on peut calculer un effectif théorique des catégories grammaticales que nous comparerons avec la réalité.

VERBES			SUBSTANTIFS			ADJECTIFS			AUTRES MOTS		
th.	réel	écart	th.	réel	écart	th.	réel	écart	th.	réel	écart
290	285	-5	398	368	-30	94	86	-8	220	224	+4

Ce qui frappe au premier abord et ce qui a déjà été constaté, c'est la prédominance des écarts négatifs, que nous interprétons comme un signe de spécialisation lexicale : un vocable employé une fois par un personnage a plus de chances de se retrouver chez ce personnage que n'importe où ailleurs, entraînant comme conséquence une spécialisation catégorielle.

Avant d'interpréter ces écarts, il serait intéressant de les comparer avec ceux que nous donnerait la répartition des mots-occur-

rences. Le calcul se ramène à une série de multiplications. Dans la construction du modèle, l'hypothèse nulle consiste à supposer que les mots-occurrences sont répartis proportionnellement à la taille des échantillons à l'intérieur des catégories grammaticales. Prenons un exemple. *L'Avare* comprend 4144 verbes. Si le texte était strictement homogène, Harpagon, qui occupe 0,301 de la pièce devrait en utiliser  $0,301 \times 4144 = 1247$ . Or, il s'en trouve 1272, ce qui lui donne un excédent de 25.

On mettra donc en regard les écarts relevés au niveau du vocabulaire et ceux qu'on a constatés dans le texte. Les nombres que nous utilisons ci-après sont ceux dont on se servirait pour le calcul d'un  $X^2$ , c'est-à-dire les carrés des écarts rapportés aux effectifs théoriques. Ils nous serviront à mesurer en quelque sorte l'importance des écarts absolus donnés au tableau. Nous indiquerons par un + ou un - le sens de ces écarts.

	Verbes	Substantifs	Adjectifs	Autres mots	$X^2$
Vocabulaire	-0,09	-2,26	-0,68	+0,07	3,10
Texte	+0,59	-0,04	-1,48	-0,01	2,12

Il est à noter que les écarts ne vont pas nécessairement dans le même sens et nous aident à mieux comprendre la langue du personnage. Un déficit sur le plan du vocabulaire qui se traduit par un excédent dans le texte est l'indice d'une plus grande redondance du vocabulaire. C'est par là que s'exprime en grande partie la spécialisation lexicale. Il faudrait une étude des fréquences et sous-fréquences pour la préciser et la localiser. Au contraire, un excédent dans le vocabulaire qui n'a pas son correspondant dans le texte signifierait un renouvellement excessif de certains éléments du lexique.

Compte tenu de ces principes et de la prédominance des écarts négatifs, on peut en induire que la richesse du vocabulaire d'Harpagon s'exprimera à un faible degré par les catégories du verbe et de l'adjectif, mais surtout par les autres mots (probablement à cause du grand nombre de noms propres). On remarque un déficit important de substantifs.

Par contre, au niveau du texte, c'est-à-dire de la répétition de ces vocables, de la caractérisation, on peut soupçonner que la spécialisation lexicale s'exprimera principalement par la catégorie du verbe et du substantif.

#### Le vocabulaire caractéristique d'Harpagon

La recherche d'un vocabulaire théorique s'était traduite, on s'en souvient, par des écarts négatifs, ce que nous avons interprété comme un phénomène de spécialisation lexicale. Il avait suffi de calculer la sous-fréquence 0 en multipliant les puissances successives de  $q$  par les effectifs correspondants de  $V_i$ .

Il nous a semblé intéressant de voir comment se produisait le déficit global attaché à chacune des valeurs de  $V_i$ . Pour ce faire, nous n'avons qu'à reprendre les développements du binôme et de calculer pour la fréquence  $f$  dans la pièce, les effectifs théoriques des sous-fréquences 0, 1, 2...  $f$  dans le rôle. Il nous restera ensuite à mettre en regard les valeurs réelles et à calculer les écarts, que nous noterons par + ou -. Voici les résultats obtenus pour les fréquences 1 à 9 (1610 vocables sur 1831, soit 88% du vocabulaire total) et appliqués au rôle d'Harpagon.

f \ f'	0	1	2	3	4	5	6	7	8	9
1	th. réel 574 588 + 14	th. réel 247 233 - 14								
2	th. réel 151 154 + 3	th. réel 130 120 - 10	th. réel 28 35 + 7							
3	th. réel 55 66 + 11	th. réel 71 61 - 10	th. réel 31 27 - 4	th. réel 4 7 + 3						
4	th. réel 23 23 0	th. réel 40 35 - 5	th. réel 23 16 - 7	th. réel 7 10 + 3	th. réel 1 10 + 9					
5	th. réel 14 17 + 3	th. réel 30 29 - 1	th. réel 26 21 - 5	th. réel 11 12 + 1	th. réel 3 3 0	th. réel 0 2 + 2				
6	th. réel 5 10 + 5	th. réel 13 13 0	th. réel 14 12 - 2	th. réel 8 3 - 5	th. réel 2 4 + 2	th. réel 0 0 0	th. réel 0 0 0			
7	th. réel 3 5 + 2	th. réel 9 7 - 2	th. réel 12 9 - 3	th. réel 8 9 + 1	th. réel 4 4 0	th. réel 1 2 + 1	th. réel 0 1 + 1	th. réel 0 0 0		
8	th. réel 2 3 + 1	th. réel 7 5 - 2	th. réel 10 6 - 4	th. réel 8 4 - 4	th. réel 4 6 + 2	th. réel 2 4 + 2	th. réel 0 2 + 2	th. réel 0 0 0	th. réel 0 3 + 3	
9	th. réel 1 1 0	th. réel 4 5 + 1	th. réel 7 7 0	th. réel 7 6 - 1	th. réel 5 4 - 1	th. réel 2 1 - 1	th. réel 0 1 + 1	th. réel 0 0 0	th. réel 0 0 0	th. réel 0 1 + 1

On constate tout d'abord un net excédent de la sous-fréquence 0. Pour des raisons thématiques et stylistiques, Harpagon évite 39 vocables qui devraient normalement y apparaître si le vocabulaire était également réparti et proportionnel à la longueur du fragment. L'écart est positif également pour les sous-fréquences  $f$  ou du voisinage de  $f$ . La zone comprise entre ces deux extrémités et qui correspond à la répartition la plus normale présente des déficits presque constants.

Ces données numériques ne font que traduire sur un autre plan la distinction entre vocabulaire commun et vocabulaire spécifique. Comme on peut le vérifier, le vocabulaire spécifique se caractérise soit par l'absence, soit par la répétition de certains vocables ou encore, comme c'est le cas la plupart du temps, par les deux à la fois. C'est exactement le phénomène que l'on rencontre chez Harpagon : il "refuse" de façon significative un certain nombre de mots du vocabulaire de la pièce, mais en même temps il *adopte* certains vocables qu'il répète anormalement. Molière caractérise son personnage principal, non pas tant en le dotant d'un vocabulaire d'une richesse extraordinaire, mais en lui donnant du relief, une originalité qui a son expression dans l'indépendance du lexique.

Le même raisonnement ne s'applique pas aux mots qui n'apparaissent qu'une fois dans la pièce et dont l'absence ou la présence sont étrangères à la notion de répartition au sens strict du terme. D'ailleurs, les écarts peuvent aller en sens inverse. Frosine, dont la spécialisation du lexique ne fait pas de doute (déficit constant pour la sous-fréquence 1), présente un excédent de 29 unités pour les mots employés une (1) fois.

Le développement de la loi binomiale nous a permis d'établir le comportement des effectifs d'une même fréquence. Si on reprend le tableau d'Harpagon, par exemple, on constate que pour les mots

de fréquence 4, il y a un excédent de 3 pour les mots employés 3 fois dans le rôle, et un excédent de 9 pour les mots employés 4 fois. De même pour les mots employés 9 fois dans la pièce, on en trouve un qui apparaît 9 fois dans le rôle alors que le calcul n'en prévoyait aucun. On se doute bien que certaines sous-fréquences sont très significatives, d'autres moins et la majorité tout à fait normale. Mais comment le savoir ?

Après avoir observé comme mouvement d'ensemble une répartition anormale des effectifs, que nous attribuons à la spécialisation lexicale, est-il possible de déterminer d'avance la zone des sous-fréquences qui est responsable de ce vocabulaire particulier et, par cette démarche, d'identifier les vocables qui en font partie ? Entre le vocabulaire spécifique et le vocabulaire commun, il n'y a pas de démarcation bien nette : tout se ramène à une question de degré que le calcul des probabilités fait bien ressortir. Le verbe *être* a une probabilité voisine de 1 d'être présent dans tous les rôles tandis que le mot *coquin* a une probabilité infime de retrouver ses 8 occurrences chez Harpagon. Entre ces deux extrêmes, on ne sait où situer la frontière.

Il serait possible de dresser un lexique de tous les mots utilisés par Harpagon, mais ce serait mettre sur un pied d'égalité des vocables qui ont des probabilités très différentes. Ce que nous avons voulu faire ressortir surtout, c'est le vocabulaire spécifique, celui qui émerge du vocabulaire commun d'une façon significative et que C. Muller appelle le vocabulaire caractéristique<sup>6</sup>. Autrement dit, quels sont les vocables de la pièce, qui se distinguent par une présence ou une absence significative dans un rôle ?

---

6. Charles Muller, *Initiation à la statistique linguistique*, Paris, Larousse, 1968, p. 204.

Prenons comme exemple un mot qui apparaît 8 fois dans la pièce. Son effectif pour chacune des parties peut varier de 0 à 8. La probabilité étant de 0,3 pour Harpagon et de 0,7 pour le reste de la pièce, on peut prévoir, si le texte était homogène, une moyenne  $p \times f$ , c'est-à-dire  $8 \times 0,3 = 2,4$  occurrences pour le rôle et  $0,7 \times 8 = 5,6$  pour tous les autres personnages réunis.

Dans l'hypothèse d'une distribution gaussienne on peut calculer de part et d'autre de la moyenne, les écarts qui devront être considérés comme significatifs. On prendra comme seuil celui de 5%, qui est obtenu par  $\bar{x} \pm 2 \sigma$ . Dans notre exemple, l'écart type est le résultat de  $\sqrt{n p q}$  où  $n$  est donné par la fréquence étudiée,  $p$ , la probabilité du fragment,  $q$ , la probabilité complémentaire.

$$\sigma = \sqrt{8 \times 0,3 \times 0,7} = \sqrt{1,68} = 1,3$$

$$2 \sigma = 2,6$$

On pourra considérer comme significatif dans le rôle les mots qui ont une fréquence inférieure à  $2,40 - 2,60$  ou supérieure à  $2,40 + 2,60$ . Les écarts négatifs ne sont même pas significatifs pour la fréquence 0, mais on devra tenir compte des excédents pour les fréquences 5, 6, 7 et 8. Le calcul fait ressortir des mots comme *Anselme, ô, pendre, peur, aise, coquin, justice, voleur*.

Nous avons l'intention de généraliser le procédé et de l'appliquer fréquence par fréquence au rôle d'Harpagon. Voici à titre d'exemples le début du calcul qui permet de fixer les bornes significatives pour le rôle d'Harpagon.

f	np	npq	$\sigma = \sqrt{npq}$	$2 \sigma$	limites
2	0,6	0,42	0,65	1,30	néant 2*
3	0,9	0,63	0,79	1,58	" 3
4	1,2	0,84	0,92	1,84	" 4
5	1,5	1,05	1,02	2,04	" 4
15	4,50	3,15	1,78	3,56	0 9
16	4,80	3,36	1,82	3,64	1 9
17	5,10	3,57	1,88	3,76	1 9
etc.					

\* Ce calcul inclut donc les mots de fréquence 2 qui n'apparaissent que dans le vocabulaire d'Harpagon. Si on invoque la loi binomiale, la probabilité pour cette catégorie de mots n'est plus de 0,05 mais bien  $p^2$  ou  $0,3 \times 0,3 = 0,09$ . Cette distinction faite, nous conservons les limites telles qu'elles apparaissent ci-dessus.

Contentons-nous d'observer dans le tableau suivant que les écarts en moins deviennent significatifs à la fréquence 10. Si on considère les 79 vocables (75 + 4) dont la répartition est significative, on constate qu'ils représentent 8% des mots étudiés, soit un peu plus des 5% attendus. On doit admettre malgré l'imprécision de la méthode que dans cette zone du vocabulaire d'autres facteurs que le hasard sont responsables de la répartition normale des mots.

Le tableau suivant donne l'effectif des mots de fréquence 2 à 24 dans la pièce et leur répartition dans le rôle d'Harpagon.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	...	24	
2	154	120	35																		
3	66	61	27	7																	
4	23	35	10	10	10																
5	17	29	21	12	3	2															
6	10	13	12	3	4	0	0														
7	5	7	9	9	4	2	1	0													
8	3	5	6	4	6	4	2	0	3												
9	1	5	7	6	4	1	1	0	0	1											
10	0	1	3	2	2	3	1	0	0	0	0										
11	0	2	3	3	1	2	2	0	1	1	0	0									
12	0	1	3	4	3	1	1	0	0	0	0	0	0								
13	1	4	1	2	2	0	2	0	2	0	0	0	0	0							
14	0	2	3	3	1	2	1	0	1	0	1	1	0	0	0						
15	0	0	2	1	1	1	0	0	0	0	0	0	0	0	1	0					
16	0	0	1	1	1	2	2	1	0	1	0	0	0	0	0	0	0				
17	0	0	0	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0			
18	1	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	...	0
19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	...	0
20	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	...	0
21	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	...	0
22	0	0	0	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	...	0
23	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	...	0
24	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	...	0

éc. nég.

signif.

(4)

écarts non

significatifs

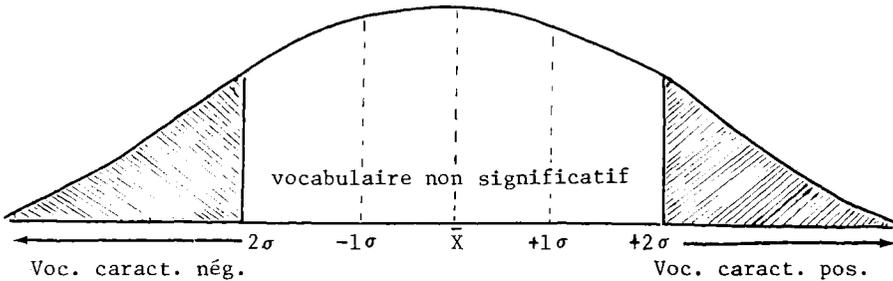
(819)

écarts positifs

significatifs

(75)

Une fois que les limites (inférieure et supérieure) ont été fixées fréquence par fréquence, il suffit de parcourir l'index par fréquence et de dégager les vocables dont la sous-fréquence s'écarte de  $2\sigma$  de la fréquence théorique. Le nouveau lexique, ainsi expurgé des éléments jugés non significatifs, représente le vocabulaire caractéristique d'Harpagon qu'on pourrait représenter de la façon suivante:



On aurait pu se contenter peut-être de dresser un inventaire à double entrée : l'une pour le vocabulaire positif et l'autre pour le vocabulaire négatif : mais nous avons voulu pousser d'un cran notre recherche. Il n'est pas indifférent, croyons-nous, à considérer la liste relativement longue de ces vocables, de mesurer les écarts qui les séparent de la moyenne et de les présenter selon cet ordre. Problème très simple au stade où nous sommes : il suffit pour chacun des mots de calculer l'écart entre la moyenne et la fréquence réelle dans le rôle, rapporté à l'écart type :

$$z = \frac{\text{fr. th.} - \text{fr. réelle}}{\text{écart type}}$$

L'écart réduit ( $z$ ) exprime cette déviation par rapport à la moyenne, en prenant l'écart type comme mesure étalon.

Prenons deux vocables tirés du vocabulaire caractéristique d'Harpagon : *fils* et *père*. Si on calcule l'écart réduit du premier,

on obtient 
$$z = \frac{14 - 7,5}{2,3} = 2,1$$

et pour le 2<sup>e</sup> 
$$z = \frac{6 - 27,6}{4,4} = - 5,1$$

Un écart réduit commence à être significatif à 2. On voit que le premier est à ranger dans le vocabulaire positif et le second dans le vocabulaire négatif ; mais alors que l'écart du premier est tout juste significatif, l'écart du deuxième l'est à un très haut point.

On présente donc le *vocabulaire caractéristique d'Harpagon* en donnant en tête de liste le vocabulaire positif en commençant par les vocables qui ont le plus grand écart réduit et nous terminerons par le vocabulaire négatif. Les chiffres entre parenthèse indiquent la fréquence du vocable dans la pièce.

tu	+ 13,2 (104)	toi	+ 4,6 (21)
te	+ 8,7 (62)	pendard	+ 4,5 (9)
ton (poss.)	+ 5,2 (15)	coquin	+ 4,3 (8)
justice	+ 4,3 (8)	soi	+ 2,6 (3)
voleur	+ 4,3 (8)	taire	+ 2,6 (3)
moi	+ 4,1 (134)	cassette	+ 2,5 (13)
dix	+ 4,0 (14)	prétendre	+ 2,5 (13)
fille	+ 4,0 (43)	belle s.f.	+ 2,4 (5)
traître	+ 3,8 (11)	dépense	+ 2,4 (5)
comment	+ 3,5 (35)	euh	+ 2,4 (5)

astre	+ 3,4 (5)	sembler	+ 2,4 (9)
chez	+ 3,4 (14)	cela	+ 2,3 (98)
sortir	+ 3,4 (5)	épouser	+ 2,3 (16)
approcher	+ 3,3 (4)	non	+ 2,3 (36)
bruit	+ 3,3 (4)	on	+ 2,3 (141)
ça	+ 3,3 (4)	écu	+ 2,2 (14)
certes	+ 3,3 (4)	mon	+ 2,2 (288)
commissaire	+ 3,3 (4)	abord	+ 2,1 (2)
or	+ 3,3 (4)	adieu	+ 2,1 (2)
pistole	+ 3,3 (4)	bienheureux	+ 2,1 (2)
potence	+ 3,3 (4)	charge s.f.	+ 2,1 (2)
souper	+ 3,3 (4)	charité	+ 2,1 (2)
ville	+ 3,3 (4)	cheveu	+ 2,1 (2)
voilà	+ 3,3 (48)	constituer	+ 2,1 (2)
aller	+ 3,2 (69)	couper	+ 2,1 (2)
carosse	+ 3,2 (7)	coutume	+ 2,1 (2)
argent	+ 3,1 (60)	damoiseau	+ 2,1 (2)
bon	+ 3,1 (33)	défaire	+ 2,1 (2)
heure	+ 3,1 (11)	déposition	+ 2,1 (2)
qui int.	+ 3,1 (33)	destiner	+ 2,1 (2)
aise	+ 2,8 (8)	devoir s.m.	+ 2,1 (2)
garde s.f.	+ 2,8 (8)	enterrer	+ 2,1 (2)
boire	+ 2,6 (3)	fier v.	+ 2,1 (2)
compter	+ 2,6 (3)	fils	+ 2,1 (29)
insolence	+ 2,6 (3)	ha	+ 2,1 (2)
nettoyer	+ 2,6 (3)	haut adv.	+ 2,1 (2)
perruque	+ 2,6 (3)	hors	+ 2,1 (2)
jurer	+ 2,1 (2)	partie	+ 2,1 (2)
larron	+ 2,1 (2)	partir	+ 2,1 (2)
louis	+ 2,1 (2)	poule	+ 2,1 (2)
maîtresse	+ 2,1 (2)	que int.	+ 2,1 (37)

me	+ 2,1 (337)	ranger	+ 2,1 (2)
mignonne s.f.	+ 2,1 (2)	ravoir	+ 2,1 (2)
mot	+ 2,1 (10)	redire	+ 2,1 (2)
muraille	+ 2,1 (2)	ruban	+ 2,1 (2)
ouais	+ 2,1 (2)	verre	+ 2,1 (2)
paix	+ 2,1 (2)	là	+ 2,0 (46)
-----			
par	- 2,2	nous	- 3,4
pouvoir	- 2,3	notre	- 3,9
consentir	- 2,4	monsieur	- 4,1
madame	- 2,6	père	- 5,1
coeur	- 3,0	votre	- 5,1
que c.j.	- 3,2	vous	- 7,1

Nous avons là le vocabulaire qui identifie le mieux le personnage, qui le caractérise et qui lui donne son originalité. Il serait vain, croyons-nous, de vouloir à tout prix distinguer les vocables dont l'écart est attribuable au thème et ceux qu'il faut mettre au compte des variations du style. Pour certains, le problème ne se pose pas. Tous les vocables relatifs à l'argent, *pistole*, *cassette*, *écu*, etc., mettent bien en évidence les préoccupations de l'Avare<sup>7</sup>, mais nous croyons que *pendard* et *coquin*, dont les écarts sont considérables, sont des termes qui n'ont rien à voir avec le thème et qu'il faut les traiter comme des jurons : on en veut la preuve dans le fait que ces deux vocables n'apparaissent que chez Harpagon malgré leur très faible probabilité. Mais pour l'ensemble des vocables, la tâche est plus complexe et nous croyons sans intérêt de nous engager dans cette voie.

7. Et pourtant le mot *franc* se trouve 5 fois sur 7 chez Frosine et jamais dans le vocabulaire d'Harpagon.

On tentera quand même un essai de regroupement des vocables en ne retenant que ceux qui à notre avis caractérisent davantage le personnage.

Le vocabulaire de la puissance. Une compilation complète des répliques nous révèle qu'Harpagon est le seul à dialoguer avec tous les personnages de la distribution. En raison de sa situation de maître du logis, jaloux de son autorité, il n'est pas étonnant de constater des excédents pour des mots comme *qui ? que ? comment ?* trahissant par là sa manie de tout contrôler. *Approcher, aller, taire* sont parmi les verbes les plus fréquents, mais ce qui est plus important que leur fréquence, c'est qu'ils apparaissent surtout à l'impératif. Le *nom*, propre aux caractères autoritaires de même que l'exclamatif *paix !* pour réduire au silence, a une fréquence significative. Notons que le verbe *consentir* par contre est absent même s'il a une fréquence de 13 dans la pièce. Ses rapports sont marqués par la familiarité (*tu, te, toi, ton*). Précisons que l'indice pronominal — rapport entre les pronoms personnels des 1<sup>re</sup> et 2<sup>e</sup> personnes et les possessifs correspondants — fait de lui, avec un indice moyen de 5,07, le personnage le plus familier et le plus naturel de la pièce.

Le vocabulaire de la possession des biens. On ne pouvait en douter, Harpagon a le vocabulaire le plus diversifié pour parler d'argent. Il utilise tour à tour *or, pistole, argent, écu, louis*. Cependant, on a vu à la note précédente que le mot *franc* est totalement absent de son vocabulaire alors qu'il est utilisé 5 fois sur 7 par Frosine. Bien qu'il y ait dans la pièce une quantité assez considérable de nombres cardinaux et que le verbe *calculer* connaisse une fréquence anormale dans le rôle, seul le nombre *dix* est employé avec un écart significatif. Par contre, tout un vocabulaire relatif

à sa situation sociale trahit ses préoccupations d'avare ; *souper, boire, carrosse, dépense, pèrruque*. Faut-il voir une manifestation de son égocentrisme possessif dans l'excédent des *mon, moi, me* qui a son corollaire dans un déficit des *nous* et des *nôtres* ?

Le vocabulaire de l'agression redoutée. Harpagon est un personnage traqué. Il n'a à la bouche que les mots *pendard, coquin, traître, larron* pour interpellier ses serviteurs. Ses craintes tournent à l'obsession : il ne pense qu'à sa *cassette* qu'il a *enterrée* près de la *muraille*, il s'inquiète de savoir si son *bien* est sous bonne *garde* et il entend des *voleurs* au moindre *bruit*.

Le vocabulaire de la répression. Un certain nombre de vocables en appellent à la *justice*. Harpagon implore le *Commissaire* de trouver le coupable à qui il souhaite les pires châtiments (*potence, pendre*).

Le vocabulaire amoureux. Ce vocabulaire fait ressortir nettement son tête-à-tête avec Mariane. Molière a voulu peindre un amoureux ridicule. Le mot *astre* revient 5 fois dans une seule réplique. C'est par des termes comme *mignonne, belle, maîtresse* qu'il s'adresse à Mariane.

### Conclusion

S'il faut mettre au compte de l'individualisation des personnages des notions comme la richesse, la familiarité, le pittoresque il n'en est aucune qui caractérise davantage que la présence d'un vocabulaire spécifique. Dans un monde clos comme une pièce de théâtre, il est normal qu'un bon nombre de vocables (surtout les mots de haute fréquence) relèvent d'un vocabulaire commun et par

conséquent se répartissent à peu près selon l'importance relative des rôles. La spécialisation lexicale s'exprime par le rejet de certains mots, mais surtout par l'accaparement des sous-fréquences au voisinage de f. Dans ce jeu étrange des répétitions, c'est comme si un personnage manifestait une prédilection ou une aversion pour des vocables bien précis.

Il n'est pas étonnant que le phénomène se situe au niveau des basses fréquences parce que ce lexique englobe tout le vocabulaire de signification.

La fréquence absolue est une donnée simple qui fait assez bien ressortir l'univers mental d'un personnage, mais outre qu'elle risque de laisser dans l'ombre des vocables de faible fréquence (exemple : les 5 occurrences de *astre* qui n'apparaissent que chez Harpagon), elle fait surtout ressortir les préoccupations thématiques. Le vocabulaire caractéristique, tel que nous l'avons défini, est plus compatible avec l'idée de style et d'écart parce qu'il oppose un personnage avec tous les autres en ne retenant que les vocables qui ont une fréquence significative. Si on s'en tient aux substantifs, les mots les plus fréquents chez Harpagon sont *argent, fille, chose, monsieur* alors que le vocabulaire caractéristique fait sortir dans l'ordre *pendard, coquin, justice, voleur*. Le deuxième classement nous semble plus près de la vérité.

Roland Pelchat

BIBLIOGRAPHIE

- BOREL, Émile, *Probabilités et Certitudes*, Paris, P. U. F., "Que sais-je ?", 1967.
- EMORINE, Jacques, *Lexique et analyse lexicale de l' "Auto da Compadecida"*. *Étude statistique*, thèse de doctorat, Université de Toulouse, 1968.
- GARAPON, Robert, "la Langue et le style des différents personnages du *Bourgeois Gentilhomme*", *le Français moderne*, avril 1958, p. 103-113.
- GUILBERT, Louis, "De l'utilisation de la statistique en lexicologie appliquée", dans *Études de linguistique appliquée*, 1963, II, p. 12-24.
- GUIRAUD, Pierre, *Les Caractères statistiques du vocabulaire*, Paris, P. U. F., 1954.
- *Problèmes et méthodes de la statistique linguistique*, Paris, P. U. F., 1960.
- HERDAN, Gustav, *Type-Token Mathematics*, La Haye, Mouton, 1960.
- *Language as Choice and Chance*, Croninger, P. Noordhoff, 1956.
- MONSENEGO, Simone, *Étude stylo-statistique du vocabulaire des vers et de la prose dans la chantefable "Aucassin et Nicolette"*, Paris, Klincksieck, 1966.
- MOREAU, R., "l'Utilisation de la notion de fréquence en linguistique", dans *Cahiers de lexicologie*, 1962, III, p. 140-151.
- MOUNIN, Georges, "Essai sur la structuration du lexique de l'habitation", dans *Cahiers de lexicologie*, 1965, VI (1), p. 51.
- MULLER, Charles, *Essai de statistique lexicale : "L'Illusion comique" de Pierre Corneille*, Paris, Klincksieck, 1964.

- MULLER, Charles, *Étude de statistique lexicale : le vocabulaire du théâtre de Pierre Corneille*, Paris, Larousse, 1961.
- *Initiation à la statistique lexicale*, Paris, Larousse, 1968.
  - *Travaux de linguistique et de littérature*, 1964, II, 1, p. 235-244.
- PICOCHÉ, Jacqueline, " le Vocabulaire du mariage dans *Tartuffe*", dans *Cahiers de lexicologie*, 1965, VII (2).
- WAGNER, R. L. et P. GUIRAUD, "la Méthode statistique en lexicologie", dans *Revue de l'enseignement supérieur*, 1955, I, p. 154-159.