

Troisième partie :

La phrase leclézienne et son rythme.

3.1. Introduction

“Le style est bien évidemment un produit naturel (même dans son inauthenticité). Les mots, les rythmes, les images ne viennent point au hasard. La signification n’est pas une idée qui cherche à se faire jour à travers le lexique ; la profondeur et la forme que prend ce sens sont déjà unies au moment où elles se cherchent, comme s’il y avait là effet du destin divin, et la phrase avance pareille à un automate, propulsée à la fois par elle-même et par ce qu’elle doit accomplir.”

L’extase matérielle, page 52.

“La phrase est le passage d’un point de pensée à un autre point de pensée. Le passage est pris dans un manchon pensant, écrit Henri Michaux dans *Ecuador*.”²⁴³ Pourtant, cette belle définition ne va pas de soi. Il y a, en effet, de très nombreuses définitions de la phrase, aussi par les théoriciens du langage, celle-ci, tout comme la définition du mot, n’étant évidente pour aucun linguiste. L’une des plus souvent citées est celle de Bloomfield²⁴⁴ :

“Chaque phrase [d’un énoncé] est une forme linguistique indépendante, qui n’est pas incluse dans une forme linguistique plus large en vertu d’une construction grammaticale quelconque.”

Sur le plan typographique, on a souvent considéré que cette indépendance de la phrase, cette clôture et cette auto-suffisance s’exprimait par la ponctuation : une phrase serait comprise entre deux ponctuations fortes. Il s’agit naturellement d’une approximation qui ne prend en compte ni la réalité syntaxique ni la réalité énonciatrice de la phrase, comme l’ont montré, notamment, les études sur l’oral. Néanmoins, travaillant sur de la prose écrite et de facture assez classique, nous nous contenterons dans un premier temps de ce critère formel qui a le mérite

²⁴³ *Ecuador* p. 41. Michaux et Le Clézio se connaissaient et Le Clézio n’a jamais caché son admiration pour Michaux et sa fascination pour son œuvre. Dans *Vers les icebergs* il en fait un commentaire détaillé et plein de louanges.

²⁴⁴ L. Bloomfield (1965) : p. 161-162.

d'être stable et aisément manipulable ; lui seul en tous cas peut donner lieu à des traitements quantitatifs réguliers.

En effet, maintes études ont été réalisées sur l'ordre des mots, le rythme des phrases ou bien sur la prosodie, mais souvent les méthodes font plus appel aux impressions qu'à la statistique. La nature même de l'étude statistique de la construction phrastique exclut de l'analyse le contenu lexical et la thématique qui préoccupent souvent les critiques littéraires. Selon Étienne Brunet²⁴⁵ :

“Il faut avouer aussi que la phrase est faite de rapports entre les mots et que ces rapports échappent à la statistique lexicale qui n'étudie que les mots individuels, détachés de leur contexte immédiat.”

Or, les méthodes statistiques peuvent s'appliquer sans difficulté insurmontable à l'étude de la phrase et les résultats complètent souvent les analyses purement stylistiques. Des études lexicométriques antérieures, notamment celles de Charles Muller²⁴⁶, celles d'Étienne Brunet²⁴⁷ et celles de Dominique Labbé²⁴⁸, ont montré l'utilité de cette étude. L'analyse de la manière dont sont agencés les éléments minimaux pour construire des phrases, des textes, des discours, ainsi que l'examen de la distribution quantitative des mots à l'intérieur de segments, peuvent donner des informations intéressantes sur un auteur et permettre de révéler certaines caractéristiques de son style.

“La phrase est hiérarchique, écrit Roland Barthes²⁴⁹ : elle implique des sujétions, des subordinations, des rections internes. De là son achèvement : comment une hiérarchie pourrait-elle rester ouverte ? La phrase est achevée ; elle est même précisément : ce langage-là qui est achevé. La pratique en cela, diffère bien de la théorie. La théorie

²⁴⁵ É. Brunet, “La phrase de Zola”, *Actes du Colloque de Victoria*, in *La critique littéraire et l'ordinateur*, B. Derval, M. Lenoble (éds.), Montréal, 1987, p. 113-114.

²⁴⁶ Ch. Muller (1964) et (1967).

²⁴⁷ É. Brunet (1978), (1983) et (1988).

²⁴⁸ D. Labbé (1990).

²⁴⁹ R. Barthes (1973) : p. 80.

(Chomsky) dit que la phrase est en droit infinie (infiniment catalysable), mais la pratique oblige à toujours finir la phrase. [...] C'est en effet le pouvoir d'achèvement qui définit la maîtrise phrastique et marque, comme d'un savoir-faire suprême, chèrement acquis, conquis, les agents de la phrase. [...] Valéry disait : "On ne pense pas ses mots, on ne pense que ses phrases." Il le disait parce qu'il était écrivain. Est dit écrivain, non pas celui qui exprime sa pensée, sa passion ou son imagination par des phrases, mais celui qui pense des phrases : un *Pense-Phrase* (c'est-à-dire : pas tout à fait un penseur, et pas tout à fait un phraseur)."

Afin de compléter les nombreuses études existantes sur le style de notre écrivain, nous nous proposons d'étudier la phrase leclézienne avec l'application des méthodes classiques de la lexicométrie. Ces méthodes donnent, en fait, un accès direct à plusieurs caractères de la phrase qui sont, grâce à l'outil informatique, simples à isoler²⁵⁰.

Nous explorerons d'abord la longueur du mot, facteur déterminant quant au rythme du récit. Un autre facteur révélateur du style d'un auteur est la longueur de la phrase. Celle-ci est généralement mesurée par le nombre de mots graphiques relevés dans l'intervalle de deux ponctuations fortes, qui sont également des jalons dans notre étude. Avant d'aborder les résultats de cette analyse, il convient toutefois de bien définir et d'étudier les signes que nous définissons comme marquant une ponctuation forte : le point, le point d'exclamation, le point d'interrogation et les points de suspension. Nous nous intéresserons par la suite à la segmentation à l'intérieur de la phrase en analysant, au préalable, les signes qui la représentent ; la virgule, le point-virgule, les deux-points ainsi que les signes de parenthèse.

²⁵⁰ Dans cette étude nous exploiterons notre corpus principal de 31 œuvres, dans sa version qui s'appuie sur les formes (corpus A) et dans sa version lemmatisée (corpus B). Pour affiner l'analyse, nous avons également jugé utile d'exploiter, à certaines occasions, le corpus qui ne s'intéresse qu'à l'œuvre romanesque de Le Clézio (corpus H).

La longueur de la phrase et les caractéristiques de ses segments dépendent étroitement de sa complexité. Le lecteur de Le Clézio n'est pas seulement frappé par les longues énumérations et les phrases-listes que nous avons déjà évoquées, mais aussi par des phrases incomplètes où les signes de ponctuation semblent avoir une autre fonction que celle de limiter la phrase. L'étude de ces phénomènes permet une réflexion plus approfondie sur la phrase de Le Clézio, qui parfois donne au lecteur une impression insolite. L'analyse approfondie des subordonnants et des coordonnants permet de faire des conclusions quant à la complexité de la phrase. L'étude de la relation entre subordination et coordination est également intéressante car elle permet d'observer les traits de style très particuliers de Le Clézio et le rythme de sa prose²⁵¹.

²⁵¹ On associe souvent le rythme à la rime et à la poésie, mais étant donné que notre auteur écrit essentiellement sous forme de prose et que la poésie est absente de notre corpus, notre étude ne s'intéressera pas à ce phénomène.

3.2. La longueur du mot.

“Il ne doit pas y avoir de mot pour l'espoir, ou pour le désespoir. Seulement un mot lent pour l'attente, un mot qui va durer des siècles.”

Le livre des fuites, page 232-233.

La longueur du mot contribue à conférer au texte son rythme. Ainsi l'étude de la longueur du mot permet-elle de distinguer les traits stylistiques particuliers d'un écrivain, d'une œuvre ou bien d'un genre littéraire qui peuvent influencer la longueur moyenne du mot dans un texte.

“En général, écrit Étienne Brunet²⁵², le volume d'un mot n'est qu'un critère secondaire, auquel l'écrivain ne prête guère attention. Si certains s'en préoccupent, c'est pour trouver le mot court, celui qui donne à l'écriture concision et nervosité.”

Il y a plusieurs façons de mesurer la longueur des mots – on peut choisir de compter par exemple le nombre de phonèmes, de lettres ou de syllabes. Dans notre étude sur la longueur du mot dans l'écriture de J.M.G. Le Clézio, comme dans la plupart des travaux lexicométriques, nous nous appuierons sur le nombre de lettres par mot.

Le tableau suivant regroupe les classes de mots du corpus A, établies sur le critère de la longueur (ou nombre de caractères desdits mots). La première colonne (lg1) dénombre les formes qui n'ont qu'une lettre, la deuxième colonne (lg2) celles qui en ont deux etc.... Les 3 dernières colonnes du tableau comportent les regroupements suivants : la classe 9 (lg9) rassemble les mots de 9 et de 10 lettres, la classe 10 (lg10) ceux de 11, 12 et 13 lettres et la classe 11 (lg11) ceux qui contiennent plus de 13 lettres :

²⁵² É. Brunet (1988) : p. 99.

Texte	lg1	lg2	lg3	lg4	lg5	lg6	lg7	lg8	lg9	lg10	lg11
Pv	2354	22569	11123	10533	8862	6823	5360	3588	3938	1391	209
Fi	2221	23294	12123	11135	9224	7817	6508	4380	4678	1682	232
De	2574	28985	15161	12846	11516	10191	8130	5264	5891	1993	259
Ex	2088	21725	12379	10598	7783	6405	5561	3492	3974	1533	258
Fu	2721	23070	13734	11390	10006	8386	6463	4011	4224	1371	183
Gu	3275	24344	15260	13191	10184	8707	6700	4670	4909	1482	190
My	294	2223	1416	1272	774	792	637	404	392	121	15
Vo	3354	25853	16222	15896	10009	8579	7031	4249	4297	1153	146
Pr	178	1907	1154	839	962	772	561	350	486	206	28
Mo	2037	22822	13040	11783	9888	8588	6577	4000	4020	1061	103
In	2556	27911	18852	15156	10645	10187	8257	4504	4717	1373	159
Ic	47	584	410	363	202	201	168	112	124	32	2
Az	79	639	503	361	308	254	151	110	136	32	3
Ds	3909	34625	20816	19761	15542	12528	9449	5374	5671	1299	114
Vi	322	3419	2241	1688	1553	1531	932	518	663	150	3
Ro	2374	19534	11063	10850	8309	6577	5052	3089	3065	857	64
Ch	3539	32550	17811	17433	12039	11070	8288	5248	5082	1191	112
An	548	5879	2876	2741	2483	2107	1853	994	1022	301	66
Rd	1811	8938	5159	3859	3540	3019	2384	1552	1606	494	70
Re	1557	19866	11916	7810	7391	6781	5876	4472	5643	2445	399
Pi	1870	16850	8696	9116	7195	5231	4309	2608	2441	744	55
Si	52	539	398	253	167	210	169	106	146	45	5
On	1910	17026	9333	8695	7972	7103	5004	3052	3028	792	81
Et	3232	26249	15533	14761	11878	10218	7747	4441	5089	1313	114
Pa	206	2599	1363	1171	1033	833	712	505	505	135	14
Di	1945	17470	9096	7450	7044	5540	4689	3646	4224	1748	340
Qu	5077	38915	21719	20276	15691	13369	10957	6166	6515	1838	158
Po	2328	22028	10686	10314	8918	6115	5217	3273	3076	839	83
Fe	1397	16090	9674	6596	6017	5689	4777	3710	4669	1830	230
Nu	534	5140	2918	2209	2204	2065	1435	802	934	322	33
Ha	1953	15578	8029	8291	7599	5681	4637	2750	3011	836	73
total	58342	529221	300704	268637	216938	183369	145591	91440	98176	30609	3801

Tableau n°24 : La longueur des mots, effectifs des classes de mots (valeurs absolues) du corpus A.

Le tableau permet de faire la même observation que dans toutes les autres études de même caractère : les plus grands effectifs se trouvent dans les premières classes, celles des mots courts. Les classes les plus fournies sont en effet celles des mots de deux et de trois lettres ; il s'agit souvent des nombreux mots-outils qui, en règle générale, sont des mots très courts.

“Il n’est plus nécessaire, écrit Charles Muller²⁵³, de démontrer que la longueur du mot n’est pas un caractère contingent et dépourvu de signification. Ses relations avec la fréquence sont assez évidentes : les mots les plus employés sont les plus courts ; dans les flexions, les formes les plus fréquentes sont aussi les plus courtes.”

“On prévoit assez facilement, continue-t-il²⁵⁴, que si l’on mesure cette longueur moyenne dans des fragments assez courts, on obtiendra des valeurs très différentes ; que si on augmente leur étendue, les résultats deviendront plus stables ; qu’enfin, à la limite, en dépouillant un corpus très vaste, on devrait tendre vers une longueur moyenne qui serait caractéristique de l’état de langue auquel appartiennent les échantillons.”

C’est pourquoi nous avons recours aux valeurs pondérées selon la loi normale dans les histogrammes qui tendent à pallier la différence de longueurs des textes de notre corpus.

Nous avons regroupé les mots courts, ceux qui contiennent de 1 à 4 lettres, et leur inventaire compte 587.563 occurrences. L’inventaire des mots longs, de plus de 8 lettres, a un effectif de 132.586 mots. Les histogrammes ci-dessous rendent compte de leur distribution respective dans le corpus :

²⁵³ Ch. Muller (1979) : p.151.

²⁵⁴ *Ibid.*, p. 153.

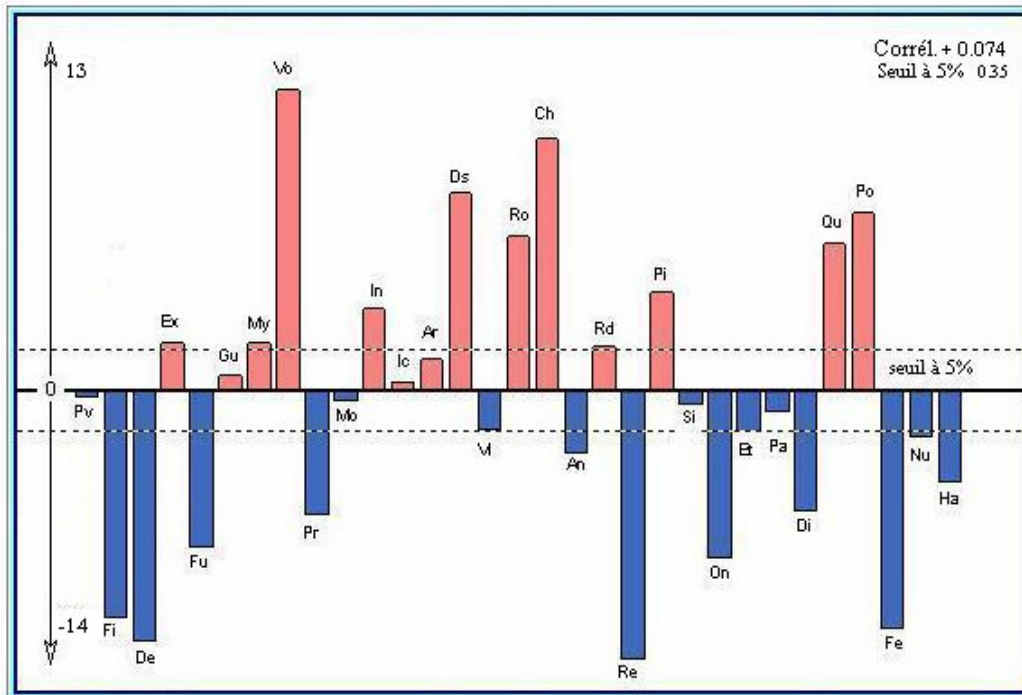


Figure n°43 : La distribution relative des mots d'une longueur de 1 à 4 lettres (écarts réduits).

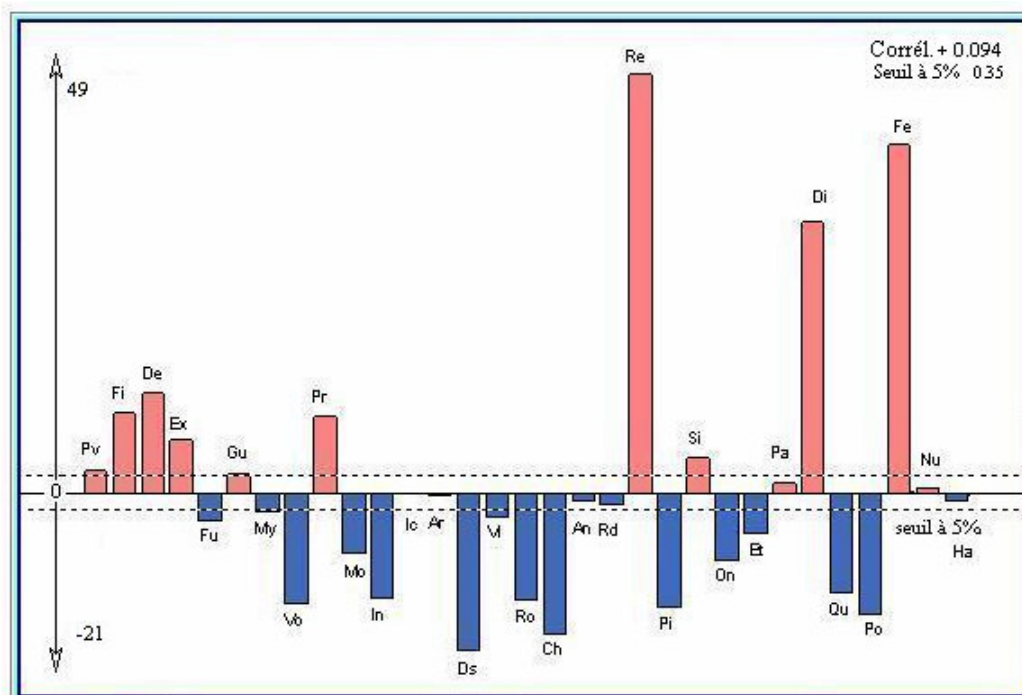


Figure n°44 : La distribution relative des mots d'une longueur de plus de 8 lettres (écarts réduits).

Ces deux histogrammes permettent d'observer la symétrie en miroir des deux graphiques. L'excédent des mots courts dans un texte est compensé par le déficit en mots longs et vice versa²⁵⁵. Les histogrammes permettent également, encore une fois, d'observer l'opposition générique du corpus. En effet, les mots longs se trouvent surtout dans les ouvrages ethnologiques et dans la biographie *Diego et Frida* ainsi que dans les premiers livres du corpus, tandis que les mots courts dominent le genre romanesque, à partir de *Voyages de l'autre côté* au détriment des mots longs.

Cette opposition générique n'est pas du tout propre à Le Clézio. Charles Muller l'a constaté depuis bien longtemps²⁵⁶ :

“Dans le cas du français moderne, le recours à un vocabulaire technique, savant ou simplement recherché augmente la proportion des mots longs, faisant ainsi monter notre indice ; un style familier ou relâché, sur le plan du lexique, agit évidemment en sens inverse. D'autre part une syntaxe soignée tend à réduire la densité du texte en mots grammaticaux, donc en mots courts, éléments que la langue courante multiplie.”

Ecartons, pour un instant, les autres genres et regardons de près celui des romans et des nouvelles, afin d'étudier l'évolution dans l'écriture romanesque de Le Clézio quant à la longueur du mot. Les histogrammes ci-dessous illustrent, comme dans les histogrammes précédents, la distribution des mots courts et celle des mots longs du corpus Roman :

²⁵⁵ Étienne Brunet a constaté la même opposition dans son étude du vocabulaire de Hugo. Cf. É. Brunet (1988) : p. 105.

²⁵⁶ Ch. Muller (1979) : 153.

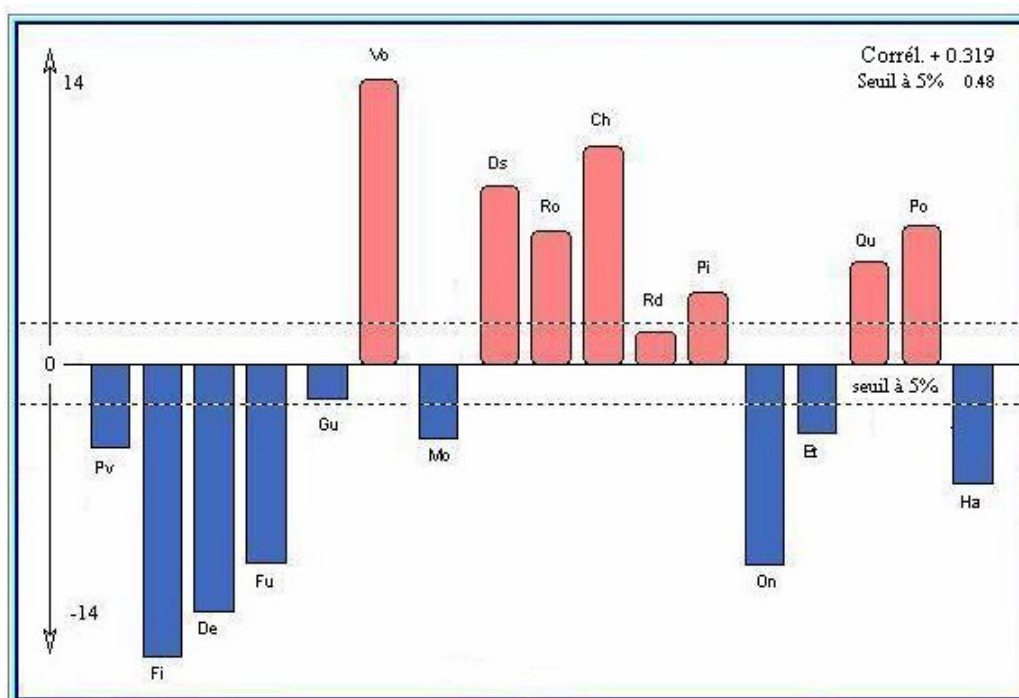


Figure n°45 : La distribution relative des mots d'une longueur de 1 à 4 lettres (écarts réduits) dans le corpus Roman (D).

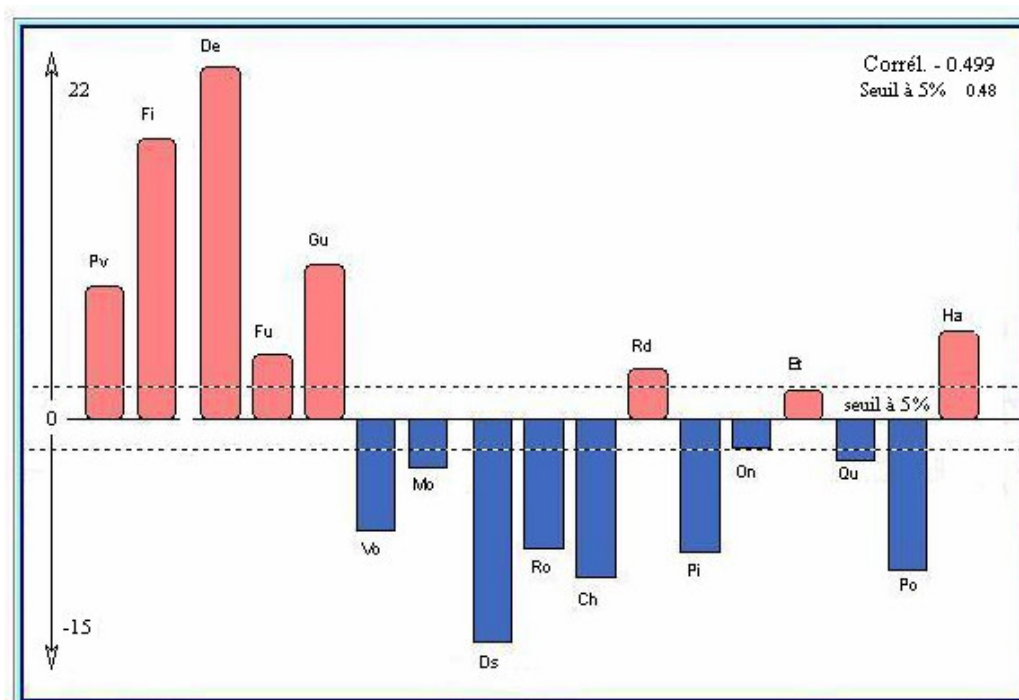


Figure n°46 : La distribution relative des mots d'une longueur de plus de 8 lettres (écarts réduits) dans le corpus Roman (D).

Cette fois-ci, l'effet miroir est encore plus net, les bâtons déficitaires d'un graphique sont compensés par des bâtons excédentaires dans l'autre. Les histogrammes permettent aussi d'isoler trois périodes distinctes dans le corpus : la première caractérisée par l'excédent des mots longs ; la deuxième, à partir de *Voyages de l'autre côté*, où les mots courts sont nettement excédentaires au détriment des mots longs ; enfin une troisième période, avec une progression, légèrement perturbée des mots longs et une régression des mots courts.

Il est intéressant de comparer ces histogrammes avec ceux que nous avons observés auparavant concernant la richesse du vocabulaire et l'accroissement lexical²⁵⁷, car ils sont, en effet, presque superposables. Nous pouvons distinguer pratiquement les trois mêmes périodes dans l'œuvre avec des valeurs excédentaires au début, une rupture nette à la fin des années 1970, suivi d'une période déficitaire et de la même progression hésitante vers la fin de l'œuvre. Cela met en évidence le fait que, lorsqu'il y a un apport de vocabulaire, ce sont surtout des mots longs qui sont ajoutés au corpus et lorsque l'apport lexical est négatif, ce sont les mots courts qui dominent. L'explication réside, en grande partie, dans le fait que les mots très courts sont souvent des mots-outils, d'emploi constant, alors que le lexique neuf est généralement constitué par des mots pleins, plus longs.

Il convient également d'étudier les mots de longueur moyenne, ceux entre 5 et 8 lettres. Ce classement de mots en longueur moyenne n'est toutefois pas évident. Étienne Brunet qui a fait leur inventaire chez Hugo et écrit à leur propos²⁵⁸ :

“... ils restent dépendants d'une décision subjective qui a fixé les barrières des trois groupes... On voit que les genres grignotent les marges du territoire voisin.”

²⁵⁷ Cf. chapitre 2.5. “L'accroissement lexical.”

²⁵⁸ É. Brunet (1988) : p. 107.

Chez Le Clézio, ce n'est pas le cas, l'histogramme ci-dessous permet de constater une distribution tout à fait indépendante des graphiques précédents :

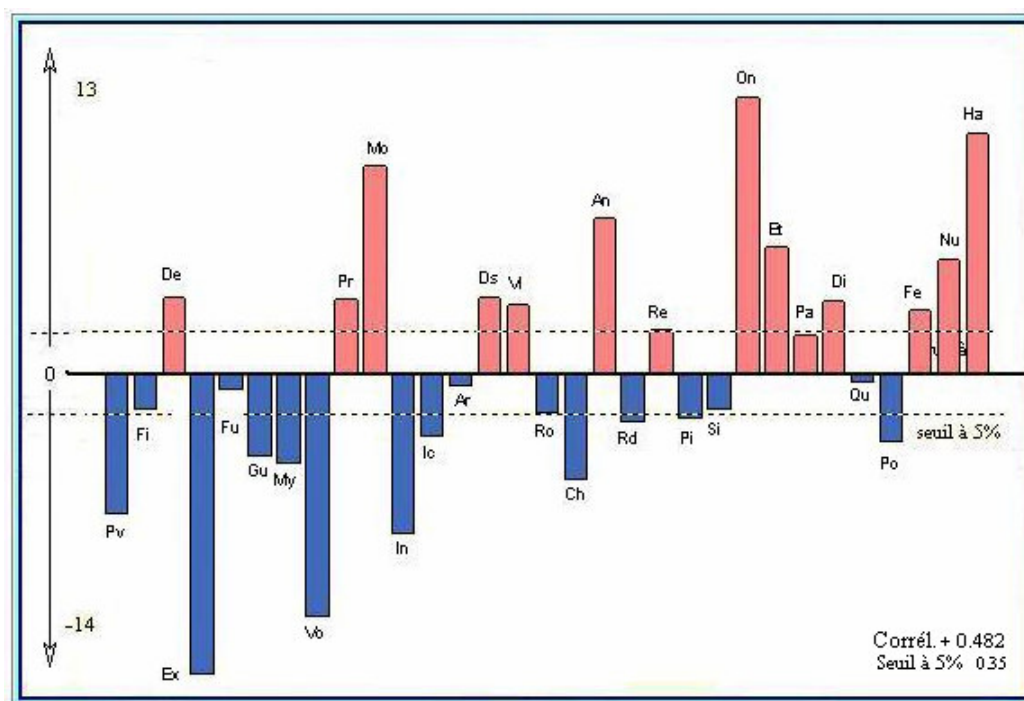


Figure n°47 : La distribution relative des mots d'une longueur entre 4 à 8 lettres (écarts réduits) dans le corpus A.

Nous ne pouvons pas, dans ce graphique, observer l'opposition des différents genres comme nous en avons maintenant l'habitude dans nos divers histogrammes. Il semble qu'il n'y ait pas de "grignotage", ni de la part des mots longs, ni de la part des courts.

En revanche, ce graphique rend compte d'un autre phénomène, celui de l'évolution temporelle de l'œuvre. Il permet d'observer un déficit initial des mots de longueur moyenne, suivi par une augmentation tout au long de l'œuvre, aboutissant à des valeurs excédentaires des mots moyens de plus en plus importantes. Le coefficient de corrélation significativement positif confirme cette tendance.

Le graphique ci-dessous, prenant en compte uniquement les œuvres romanesques, permet de constater le même mouvement :

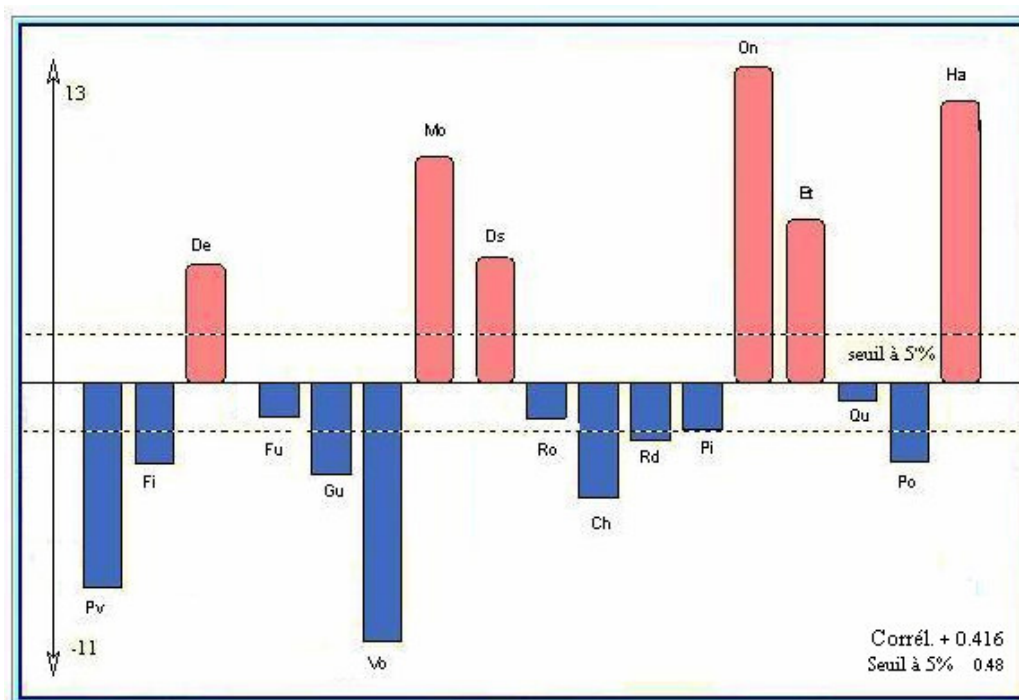


Figure n°48 : La distribution relative des mots d'une longueur entre 4 à 8 lettres (écarts réduits) dans le corpus Roman (D).

L'histogramme illustre la même évolution que nous avons observée dans le grand corpus (toutefois sans un coefficient de corrélation franchement positif comme dans l'histogramme précédant), mais en soulignant les exceptions à la règle de façon plus visible. *Le déluge* semble ici “grignoter” sur les mots longs. *Mondo*, qui était le seul livre à être déficitaire dans les deux graphiques précédents, puise l'essentiel de son vocabulaire parmi les mots de longueur moyenne. *Désert*, le livre le plus déficitaire en mots longs du corpus, paraît compenser dans le groupe des mots de longueur moyenne. Quant à *La quarantaine* et à *Poisson d'or*, qui sont déficitaires en mots de longueur moyenne, ils compensent en puisant ce déficit parmi les mots courts.

Afin de mieux illustrer cette dynamique, nous avons recours à l'analyse factorielle du corpus dans son ensemble. Les chiffres entourés d'un cercle correspondent aux différentes classes de longueur :

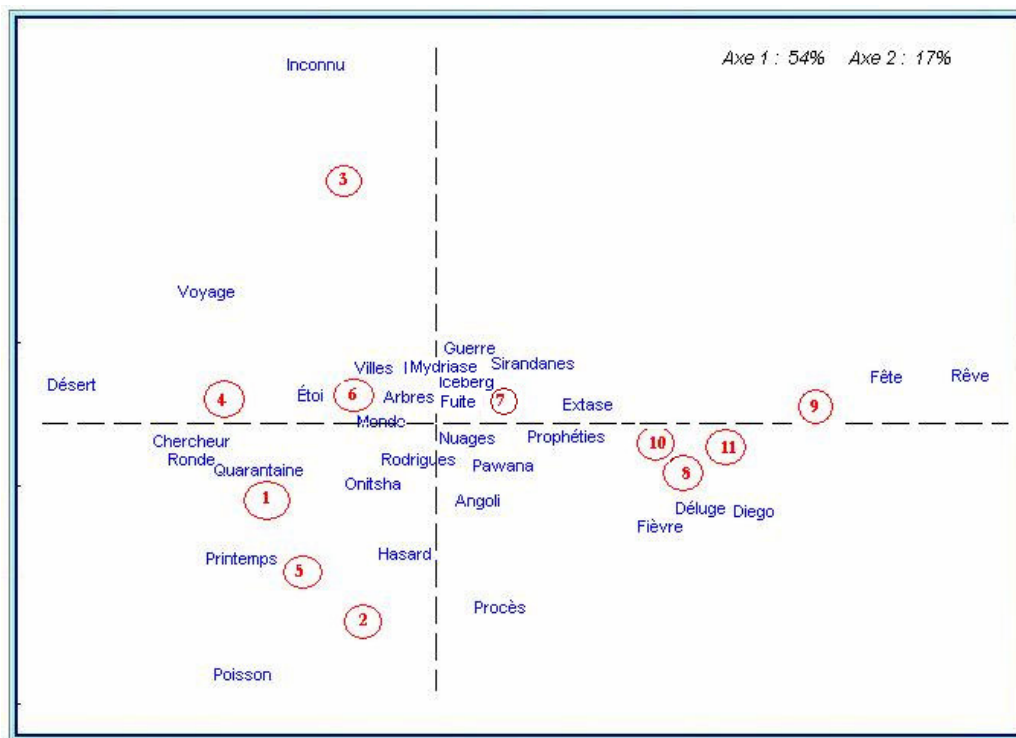


Figure n°49 : Analyse factorielle de la longueur du mot.

Le premier facteur de l'analyse est celui des genres. Il oppose les ouvrages ethnologiques et la biographie aux romans et aux nouvelles classiques. Nous constatons également que les premiers romans de l'école "nouveau roman" se regroupent entre eux et, à proximité de la biographie ; ceux de la deuxième et de la troisième période forment aussi deux groupes assez homogènes. L'analyse permet en outre d'observer le regroupement des classes de longueur différentes. Les mots courts restent ensemble à gauche du graphique, les mots longs à droite, tandis que ceux de longueur moyenne sont regroupés au centre.

En conclusion, l'étude de la longueur du mot chez Le Clézio nous a permis de confirmer l'opposition des genres littéraires, les mots longs se trouvant plutôt dans les ouvrages ethnologiques et les mots courts dans le genre romanesque.

Toutefois, à l'intérieur de ce dernier genre, les premiers livres, ceux de l'école du "nouveau roman", sont excédentaires en mots longs. Les mots de longueur moyenne, en revanche, semblent figurer dans le corpus indépendamment du genre étant à leur tour conditionnés par le temps, en augmentation régulière au fur et à mesure que l'œuvre progresse. Cette analyse a également permis de distinguer la même dynamique que dans les analyses de la structure lexicale, celle des trois périodes : une première excédentaire en mots longs, une rupture nette suivie d'une deuxième période déficitaire, dominée par les mots courts, et enfin, une troisième période caractérisée par une progression modérée de la longueur du mot.

Un autre facteur, plus souvent étudié que celui de la longueur du mot, est la longueur de la phrase, son unité supérieure, considérée comme un critère déterminant dans l'analyse stylistique et dans l'étude du rythme d'un écrivain.

3.3. La ponctuation forte.

“ , ”
“ ! ”
“ ! ! ”
“ ... ”
« ? »

Rien que des signes, provisoires, rapides, rapides, et on va plus vite qu’eux. On est déjà de l’autre côté de ce qu’ils ont dit, on court comme la lumière.

Voyages de l’autre côté, page 29-30.

La ponctuation est essentiellement d’ordre syntaxique et doit être comprise comme l’écrit Nina Catach²⁵⁹ :

“... associant à la fois : une *suite de mots* (aspect constructif), un *message* (aspect actuel), une *substance* et une *forme* intonatives (mélodie expressive et aspect intonatif) et un *sens* (contenu du message, résultant de l’ensemble des données précédentes). La ponctuation tient dignement son rôle à ces différents niveaux de la syntaxe.”

Dans le langage écrit, la fin de la phrase est généralement représentée par le point, que Littré a défini dans son dictionnaire comme une “petite marque que l’on met dans l’écriture pour indiquer la fin des phrases.” La première fonction du point semble donc de marquer la fin de toute phrase ; toutefois, le point a bien d’autres fonctions et d’autres signes peuvent également marquer la fin de phrase : les points de suspension, le point d’interrogation, le point d’exclamation, le point-virgule et les deux points. Quant aux deux derniers, ils se trouvent également assez souvent à l’intérieur de la phrase, ou bien, lorsqu’ils coïncident avec la fin d’une phrase grammaticale, ils marquent un lien logique avec la phrase suivante. C’est pour cette raison que nous ne les définissons pas comme des marqueurs de fin de phrase dans cette étude et que nous les considérerons dans l’analyse de la segmentation interne de la phrase.

²⁵⁹ N. Catach (1994) : p. 48.

La ponctuation n'est pas simplement l'utilisation de divers signes. Jacques Drillon²⁶⁰ donne à ces signes une bien plus grande importance :

“Tous les signes de ponctuation sont des raccourcis ; tous, sans exception, sont la marque d'une ellipse. Une chose était à dire, si constante qu'on l'a symbolisée. On peut, *al rovescio*, développer ces symboles, les écrire « en extension », comme en dit en mathématique. [...] Voilà l'idée maîtresse : les signes ont un sens.”

L'emploi des signes de ponctuation varie selon le genre de discours, l'époque et l'auteur. La ponctuation peut également connaître de grandes variations chez un même écrivain. Nous nous intéresserons dans ce chapitre à l'emploi des quatre signes de ponctuation forte – le point, le point d'exclamation, le point d'interrogation et les points de suspension – dans l'œuvre de Le Clézio afin d'en mesurer les variations. Toutefois, avant de pouvoir effectuer cette étude, il convient d'éclaircir certaines ambiguïtés.

3.3.1. La désambiguïsation des signes et l'établissement de l'inventaire.

Le fait que les points soient utilisés non seulement pour marquer la fin d'une phrase, mais qu'ils puissent aussi se trouver à l'intérieur de la phrase, marquant une abréviation ou une siglaison quelconque, peut évidemment fausser l'étude statistique de la ponctuation forte.

La lecture de l'œuvre de Le Clézio, surtout du début de la production, donne souvent l'impression, avec ses nombreux noms propres abrégés et ses longues énumérations de produits ou d'articles divers, de compter de nombreux points à l'intérieur de la phrase qui pourraient troubler l'étude de l'emploi du point final.

²⁶⁰ J. Drillon (1991) : p. 19.

Mais il est évident que le recensement de toutes les abréviations et de tous les sigles dans un grand corpus est très coûteux et la valeur d'un tel travail discutable. Nous nous sommes donc limitée, avec l'aide du logiciel Hyperbase et de la technique d'Étienne Brunet²⁶¹, à filtrer les points qui terminent les mots abrégés²⁶² pour ainsi pouvoir les déduire de l'effectif total des points finaux. La même technique a été utilisée pour recenser l'effectif de l'abréviation "etc." Concernant cette abréviation, il est à noter que, dans notre corpus, elle est surtout utilisée en fin de phrase dans notre corpus, et le même point qui termine l'abréviation clôt également la phrase (des 67 occurrences, 47 coïncident avec le point final de la phrase).

Nous avons ainsi filtré 873 points marquant des abréviations dans notre corpus, en les repérant lettre par lettre. Le tableau ci-dessous rend compte de leur distribution :

²⁶¹ Cf. É. Brunet (1988) : p. 64-65.

²⁶² Le traitement initial qu'effectue le logiciel Hyperbase, lors de la création d'une base de données, insère un espacement entre la dernière lettre d'un syntagme et le point. Par conséquent, le filtrage tient également compte de sigles de type H.L.M. et P.M.E.

Tableau de l'abréviation																												
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	etc	total
Procès	3	0	4	1	1	0	1	0	0	3	2	1	20	3	2	2	0	14	4	0	4	0	4	0	0	1	4	74
Fièvre	0	1	2	0	2	12	0	5	1	12	0	3	7	0	1	0	0	0	3	2	0	1	0	1	0	1	4	58
Déluge	7	1	3	3	2	1	0	0	1	1	0	2	2	1	0	3	0	1	8	4	1	1	1	1	0	0	2	46
Extase	8	1	2	3	1	1	1	0	0	2	0	2	6	2	1	2	2	1	3	1	0	0	2	4	2	0	2	40
Fuite	0	0	1	1	3	4	1	53	0	59	0	1	2	4	0	1	0	0	2	2	0	1	3	0	1	0	1	140
Guerre	3	90	3	1	2	0	1	0	0	0	0	0	0	0	1	2	0	0	6	0	0	3	1	0	0	0	2	115
Mydriase	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Voyage	0	0	0	0	2	0	0	0	0	0	1	0	31	0	0	1	0	0	1	0	0	0	0	0	0	0	0	36
Prophéties	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Mondo	0	0	1	0	0	0	0	0	0	0	0	0	20	0	1	1	0	1	0	0	0	0	0	0	0	0	0	24
Inconnu	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Icebergs	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arbres	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Désert	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	2	0	0	0	0	6
Villes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Ronde	5	0	2	0	0	0	0	1	0	0	0	3	3	0	0	2	0	0	2	0	3	0	1	0	0	0	0	22
Chercheur	2	0	2	2	0	4	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	10	0	0	0	1	39
Angoli	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rodrigues	0	0	2	2	2	0	0	0	0	0	0	0	9	2	2	0	0	0	3	0	0	0	0	0	0	0	0	22
Rêve	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	78	0	1	0	0	0	0	0	1	0	0	0	81
Printemps	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
Sirandanes	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Onitsha	1	0	1	12	1	0	0	0	0	0	1	0	31	0	13	0	0	0	0	1	0	0	0	0	0	0	0	61
Etoile	1	0	4	0	0	1	0	0	0	0	0	0	46	1	0	0	0	0	2	0	0	0	3	0	0	0	0	58
Pawana	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Diego	2	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	2	11
Quarantaine	0	0	1	1	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6
Poisson	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	8
Fête	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	4
Nuage	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Hasard	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3
total	34	93	35	28	17	24	6	60	5	78	5	15	197	13	22	93	2	20	34	14	8	6	32	7	3	2	20	873

Tableau n°25 : Abréviations.

Cette distribution n'est pas régulière et comporte d'assez grandes variations. Un effectif important d'une lettre suivie par un point dans un ouvrage particulier tient souvent au fait qu'il s'agit des initiales des noms propres ou des titres des personnages. La présence de Bea B. dans *La guerre* ou bien du D.O. – le District Officer – dans *Onitsha* se manifeste dans notre tableau.

Le lecteur de Le Clézio se souvient également des “jeux de noms” auxquels se livre parfois l'auteur. Le personnage principal de *Le livre des fuites*, Hogan, est également appelé Jeune Homme Hogan, Jeune Homme, J. Hombre Hogan, Juanito Holgazán, J.H. Hogan, J.H.H. ou bien J.H. dans les différents chapitres du

roman²⁶³, ce que reflète le nombre relativement important de “H.” et de “J.” de notre tableau.

Dans le cas de “M.”, qui fournit l’effectif global le plus important, il s’agit évidemment de l’abréviation de “Monsieur”. Parmi les 8 livres du corpus qui en contiennent, nous trouvons surtout des livres où les personnages principaux sont des enfants ou des jeunes confrontés au monde des adultes.

Le taux important de “P.” dans *Le rêve mexicain* témoigne du genre littéraire. Dans cet essai, Le Clézio fait de nombreux renvois à d’autres pages du livre, dans le texte même, ce qui explique ce grand effectif.

La comparaison avec le nombre d’abréviations qu’a relevé Étienne Brunet dans la littérature du XIX^{ème} siècle, permet de constater que, par exemple, le corpus Victor Hugo de 2,6 millions d’occurrences contenait 4654 abréviations tandis que celui de 2,2 millions d’occurrences de Le Clézio n’en compte que 873. Ce faible effectif chez notre écrivain semble contredire les remarques que fait Étienne Brunet sur l’apparition relativement récente des sigles : “La mode n’étant pas encore à la siglaison, écrit-il²⁶⁴, [...] cette habitude ne s’impose guère qu’au XX^{ème} siècle...” Le faible taux d’abréviations chez Le Clézio pourrait refléter une autre évolution générale à la fois linguistique et culturelle, celle vers un style moins relevé où le tutoiement prime sur le vouvoiement bien qu’il n’y ait pas de rapport direct entre l’emploi de M. et de Mme et le vouvoiement. L’inventaire d’abréviations du corpus Victor Hugo contient, en effet, un effectif beaucoup plus important de “M.” ainsi que de “MM.”, cette dernière abréviation étant totalement absente de notre corpus.

Le recensement de ces points qui ne sont pas des marqueurs de fin de phrase permet de déduire les 873 occurrences et d’établir l’inventaire des signes de ponctuations fortes de notre corpus dont la distribution est la suivante :

²⁶³ Le Clézio donne aussi d’autres noms à ce personnage dans le roman.

²⁶⁴ É. Brunet (1988) : p. 64.

Texte	.	...	!	?	Texte	.	.	!	?
Pv	3510	132	123	482	Ch	5655	5616	164	408
Fi	4164	217	150	409	An	1123	1123	24	58
De	5474	313	196	337	Rd	981	959	8	135
Ex	4468	0	117	416	Re	2446	2365	39	66
Fu	5307	45	834	422	Pi	3496	3494	92	137
Gu	5260	50	283	264	Si	71	69	5	18
My	489	2	15	26	On	3939	3878	124	131
Vo	5686	15	169	139	Et	5346	5287	134	195
Pr	336	1	2	7	Pa	462	462	2	24
Mo	4758	79	237	251	Di	2020	2009	24	25
In	5389	21	158	318	Qu	7896	7890	223	374
Ic	111	3	1	14	Po	4189	4181	108	171
Ar	149	1	3	8	Fe	2118	2114	14	69
Ds	5374	94	185	228	Nu	803	802	7	31
Vi	633	7	7	17	Ha	3061	3058	34	149
Ro	3267	38	88	191					
Total					.	...	!	?	
					108152	97682	1380	3570	5520

Tableau n°26 : Les signes de ponctuations fortes chez Le Clézio, effectifs absolus.

L'addition de ces quatre signes fournit l'effectif de la ponctuation forte dans les différents sous-corpus, qui ne sont réellement comparables qu'en tenant compte de la taille respective de chaque ouvrage. L'histogramme suivant illustre la distribution des écarts réduits après pondération :

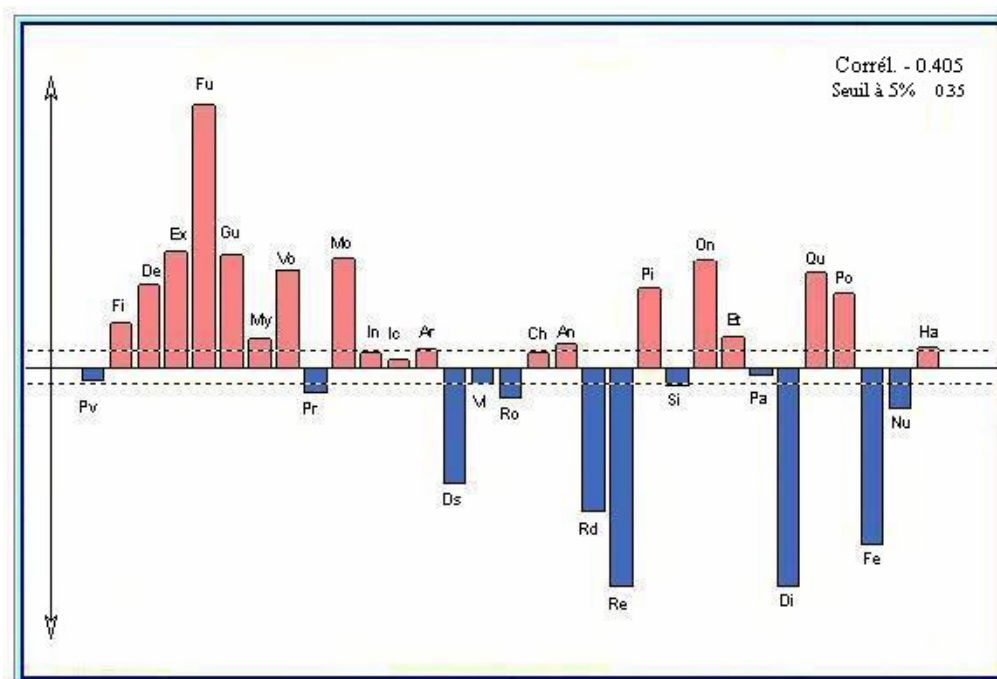


Figure n°50 : Distribution relative des quatre signes de ponctuation forte dans le corpus A.

Il est aisé de constater que la distribution n'est pas régulière, variant selon les périodes et selon les différents genres. Mais avant de commenter ces résultats en détail et d'en tirer des conclusions, nous nous intéresserons à la distribution de chaque signe pris séparément à l'intérieur du corpus, en tenant compte des tailles différentes des sous-corpus, et en commençant par le plus fréquent : le point final.

3.3.2. L'emploi du point final.

Le point final, principal marqueur de fin de phrase, compte un effectif de 97682, après la déduction des 873 points filtrés des 98555 occurrences relevées dans le corpus entier. Leur distribution relative (valeurs des écarts réduits) est illustrée par la figure suivante :

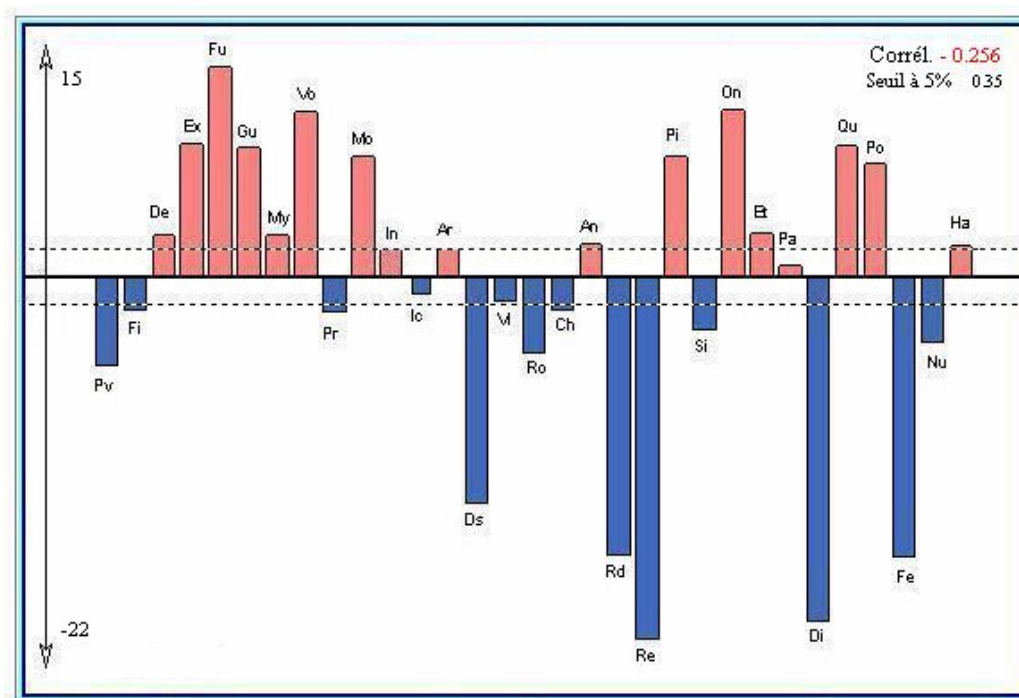


Figure n°51 : Distribution des points dans le corpus A.

Que cet histogramme ressemble fort à l'histogramme illustrant la distribution de l'ensemble de la ponctuation forte n'a rien d'étonnant, le point final étant son

composant principal. Nous observons que la fréquence varie d'une œuvre à une autre et que les fluctuations sont même d'une importance considérable.

La fréquence du point est déficitaire dans les deux premiers livres, puis élevée dans ceux qui suivent sur le premier tiers du graphique. Après *Mondo et autres histoires*, le point se fait plus rare, les romans *Désert* et *Voyage à Rodrigues* sont nettement déficitaires ainsi que les ouvrages ethnologiques et la biographie *Diego et Frida*. Dans les livres de fiction, la fréquence du point final augmente de nouveau vers la fin de l'œuvre pour diminuer légèrement dans les deux derniers romans *Poisson d'or* et *Hasard*.

En revanche, lorsque le point est utilisé trois fois de suite, comme signe de suspension, la situation est tout à fait différente.

3.3.3. Les points de suspension.

“Les trois points terminateurs me font hausser les épaules de pitié.
A-t-on besoin de cela pour prouver que l'on est un homme d'esprit, c'est-à-dire un imbécile ? Comme si la clarté ne valait pas le vague, à propos de points !”

Lautréamont, *Poésies*, II, page 364.²⁶⁵

Le Clézio n'emploie pas souvent les points de suspension. Nous n'en relevons que 1380 occurrences dans notre corpus. L'histogramme ci-dessous permet de visualiser leur distribution à travers le corpus :

²⁶⁵ Proust partage cette même méfiance à l'égard des points de suspension.

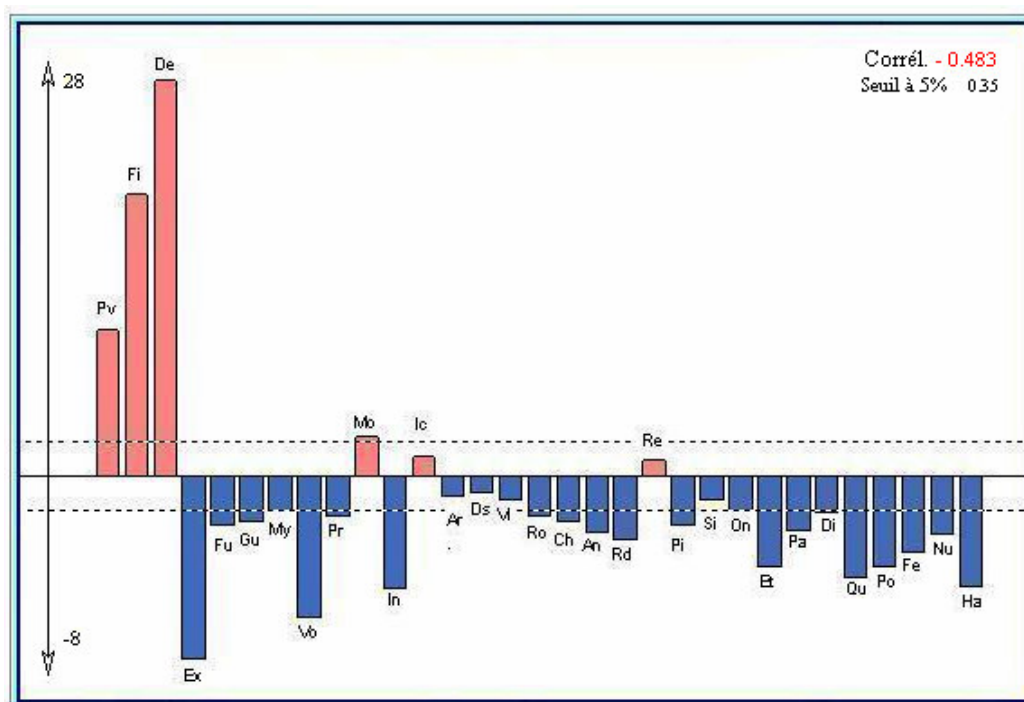


Figure n°52 : La distribution des points de suspension dans le corpus A (écarts réduits).

En effet, Le Clézio n'en fait usage que dans ses trois premiers livres : *Le procès-verbal*, *La fièvre* et *Le déluge*, pour ensuite les abandonner. L'histogramme n'indique aucune valeur positive significative après ces trois premiers ouvrages.

Les points de suspension servent, nous le savons, à remplacer une partie de l'énoncé ou à l'interrompre. "Ils tiennent, selon Bachelard²⁶⁶, en suspens ce qui ne doit pas être dit clairement." L'écriture leclézienne, tout au contraire, veut tout dire et tout rendre vivant par les mots et ne cherche pratiquement jamais à interrompre les phrases pour produire des effets, comme par exemple le suspens. Nous saisissons ici un trait fondamental du style de l'auteur ; Jean Onimus l'exprime ainsi²⁶⁷ :

"En écrivant, Le Clézio voudrait rejoindre la substance des choses. Non pas seulement les dire mais les appeler à exister par la parole. « Ce qui

²⁶⁶ G. Bachelard, in *Le Nouveau Petit Robert* (1995) : p. 2186.

²⁶⁷ J. Onimus (1994) : p. 160.

est beau, quand on ne vit pas seulement les choses, mais qu'on les dessine et les écrit, c'est qu'elles sont deux ou trois fois »²⁶⁸

Le Clézio n'est pas le seul à éviter les points de suspension, les études d'Étienne Brunet²⁶⁹ montrent qu'ils ne sont guère prisés par Hugo ou par Proust. En revanche, Zola et Giraudoux en font un usage beaucoup plus important.

La distribution du troisième signe de ponctuation, le point d'exclamation, est quelque peu différente du point final et des points de suspension.

3.3.4. Le point d'exclamation.

Notre corpus compte 3570 points d'exclamation. L'histogramme ci-dessous illustre la distribution relative de ce signe dans notre corpus :

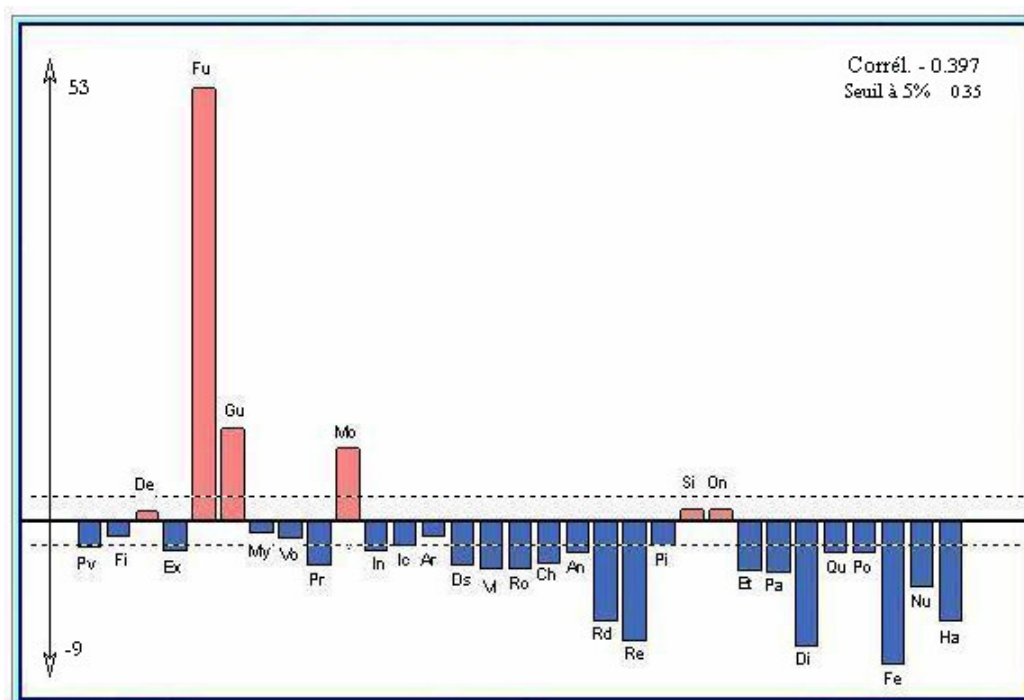


Figure n°53 : Distribution des points d'exclamation dans le corpus A (écarts réduits).

²⁶⁸ *L'inconnu sur la terre*, p. 78.

²⁶⁹ É. Brunet (1988) : p.74-75.

Le Clézio utilise le point d'exclamation de façon significative dans *La guerre* et dans *Mondo et autres histoires*, mais c'est surtout dans *Le livre des fuites* qu'il exploite cette ponctuation, le livre contenant 23 % des effectifs à lui seul. Dans ce roman l'auteur a en effet souvent recours à l'exclamation pour exprimer la hantise du personnage principal dans sa découverte du monde extérieur.²⁷⁰

“Les choses entraient, ou peut-être même naissaient sur place, sans qu'il ait à s'inquiéter. Tout ce qu'il pouvait penser, c'était une suite d'exclamations maladroites, comme :

- Le tabouret, oh ?
- La statue de porcelaine, ah, oh !
- Ha ! Le tapis ! - La peau de léopard !
- Oh ! La calebasse...
- Oh ! Ah ! Le gros lézard empaillé ! Oh !”

Toutefois, *Le livre des fuites* n'est pas le seul livre qui exploite le point d'exclamation de façon significative. Nous relevons également un emploi intéressant, bien que plus restreint, dans *La guerre* et dans *Mondo*. En revanche, le point d'exclamation est déficitaire – en dehors des valeurs positives mais non significatives car très faibles de *Le déluge*, de *Sirandanes* et d'*Onitsba* – dans les autres livres du corpus.

L'opposition générique se manifeste également dans cet histogramme ; les ouvrages ethnologiques et les essais sont nettement déficitaires, leur style descriptif n'étant pas favorable à l'emploi du point d'exclamation.

L'emploi de ce signe est souvent mis en relation avec l'emploi du point d'interrogation, étant tous deux des signes de ponctuation qui expriment le sentiment. Chez Hugo, par exemple²⁷¹, les deux signes se rencontrent dans les mêmes situations sauf dans la poésie, une écriture qui “atténue l'interrogation mais

²⁷⁰ *Le livre des fuites*, p. 38-39.

²⁷¹ É. Brunet (1988) : p.76.

accentue l'exclamation". Dans les textes de Zola cependant l'effectif des points d'exclamation est deux fois plus important que l'effectif des points d'interrogation²⁷².

Chez Le Clézio, en revanche, nous ne retrouvons pas le même parallélisme entre le point d'exclamation et le point d'interrogation.

3.3.5. Le point d'interrogation.

Constatons, premièrement, que Le Clézio fait un usage beaucoup plus important du point d'interrogation que du point d'exclamation - nous en comptons 5520 effectifs contre 3570. L'étude des signes de ponctuation chez Hugo a permis de constater la situation inverse tout comme chez Zola²⁷³. Cette constatation n'a rien d'étonnant pour celui qui connaît l'œuvre de Le Clézio, souvent caractérisée par une grande modestie dans l'expression et où l'auteur, à travers ses différents personnages, semble surtout [se] poser des questions et chercher des réponses que s'exclamer.

Par ailleurs, la distribution des deux signes à travers le corpus est très différente. Il semble qu'ils sont, contrairement aux constatations qu'a faites Étienne Brunet concernant la littérature du XIX^{ème} siècle, tout à fait indépendants l'un de l'autre. Il n'y a ni attirance ni répulsion à constater entre les deux, comme le montre l'histogramme ci-dessous qui mesure le quotient entre la distribution des deux signes :

²⁷² É. Brunet, "La phrase de Zola", *Actes du Colloque de Victoria*, in *La critique littéraire et l'ordinateur*, B. Derval, M. Lenoble (éds.), Montréal, 1987, p. 122.

²⁷³ É. Brunet (1985) : p. 104 (10664 points d'interrogation contre 24025 points d'exclamation). Il faut toutefois tenir compte du fait que ce corpus inclut également de la poésie et que lorsqu'on l'exclut, l'excédent des points d'exclamation tend à disparaître.

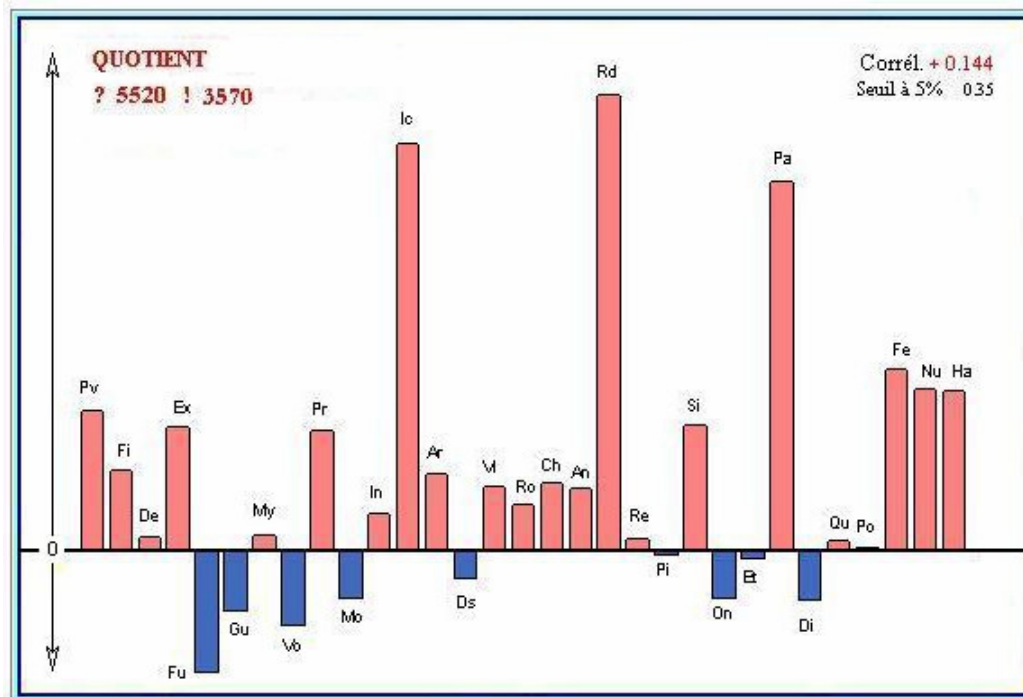


Figure n°54 : Le quotient entre points d'interrogation et points d'exclamation dans le corpus A.

L'histogramme ci-dessous permet l'observation de la seule distribution du point d'interrogation dans le corpus :

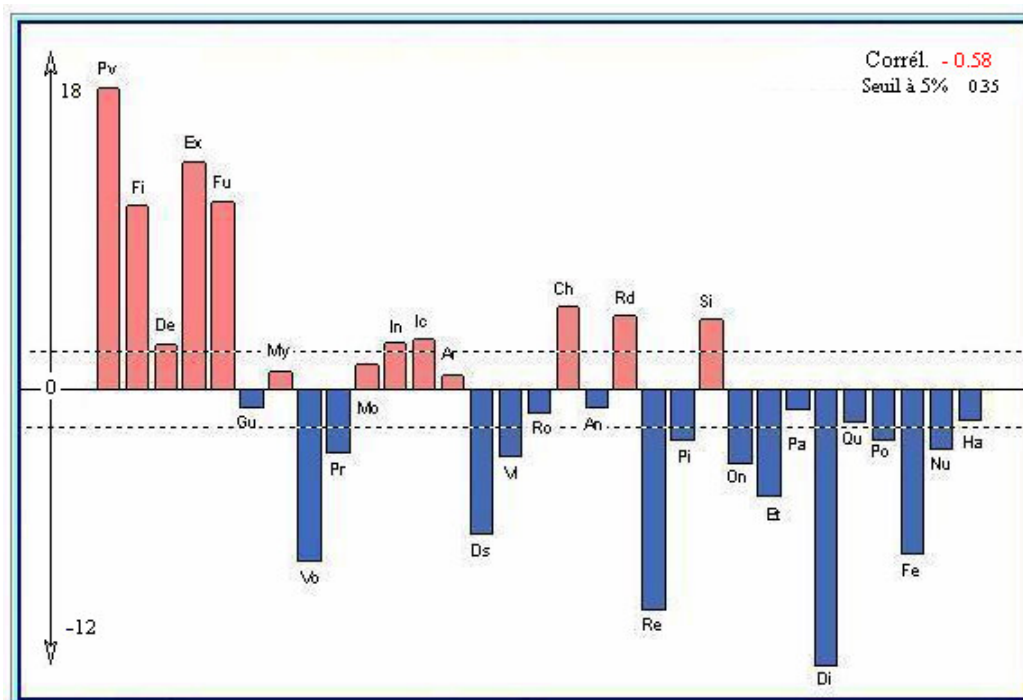


Figure n°55 : La distribution relative des points d'interrogation dans le corpus A.

La fréquence importante des points d'interrogation dans les premiers sous-corpus, correspondant aux débuts de Le Clézio, est remarquable. Ce n'est qu'à partir de *La guerre* que nous pouvons observer des valeurs négatives. Par la suite, le déficit n'est interrompu que pendant deux courtes périodes ; la première correspondant à l'année 1978 – l'année qui compte le plus de parutions, celle qui suit la rupture de l'écriture leclézienne²⁷⁴, et la deuxième à la fin des années 80, quand la fiction des œuvres se déroule aux îles de l'Océan indien. Le coefficient de corrélation significativement négatif témoigne de l'abandon successif du signe.

L'histogramme permet de faire la même observation à propos du point d'exclamation : il y a moins de points d'interrogation dans les ouvrages ethnologiques, la biographie et dans les essais littéraires – à l'exception de *L'extase matérielle* du début de l'œuvre - que dans les œuvres romanesques.

Nous pouvons en effet observer l'opposition générique dans l'étude de l'emploi des quatre signes de ponctuation forte ; ils sont toujours déficitaires dans les ouvrages ne relevant pas de la fiction. En revanche, dans les livres romanesques, la fréquence est très importante dans les premiers ouvrages pour décliner par la suite au fur et à mesure que l'œuvre progresse.

L'étude des quatre signes de ponctuation forte - le point, le point d'exclamation, le point d'interrogation et les points de suspension - et le recensement de leurs effectifs respectifs (97682 ., 1380 ..., 3570 ! et 5520 ?) que nous venons de faire, sont des éléments indispensables pour connaître la longueur de la phrase. La connaissance du nombre total de signes de ponctuation forte dans un corpus permet en effet de calculer et d'étudier la longueur de la phrase d'un écrivain.

²⁷⁴ Cf. Les résultats des analyses sur la structure du vocabulaire, chapitres 2.4. "La richesse du vocabulaire" et 2.5. "L'accroissement lexical."

3.4. La longueur de la phrase leclézienne.

“La signification n’est pas une idée qui cherche à se faire jour à travers le lexique; la profondeur et la forme que prend ce sens sont déjà unies au moment où elles se cherchent, comme s’il y avait là effet du destin divin, et la phrase avance pareille à un automate, propulsée à la fois par elle-même et par ce qu’elle doit accomplir.”

L’extase matérielle, page 52.

La longueur de la phrase est un critère important quand on veut décrire le style d’un auteur ou d’une œuvre littéraire. Ce trait particulier du discours, qui constitue un critère stylistique fortement caractéristique d’un auteur, est souvent indépendant de la volonté de celui-ci et permet ainsi d’effectuer des analyses intéressantes, notamment dans l’attribution des textes.

Michael Oakes²⁷⁵ écrit ceci à ce propos :

“One aspect of literary detective work on the computer concerns the statistical analysis of writing style. In order to perform a computer analysis of style, most studies concentrate on features which are easily identifiable and can be counted by machine. The features most frequently used are :

- word and sentence length
- the positions of words within sentences
- vocabulary studies, based on what Hockey²⁷⁶ calls the choice and frequency of words
- syntactic analysis, including the additional task of defining and classifying the syntactic features.”

²⁷⁵ M.P. Oakes (1998) : p. 199.

²⁷⁶ S. Hockey (1980).

Toutefois l'étude de la longueur de la phrase n'est pas seulement exploitée dans le travail de l'attribution de textes, elle permet avant tout de comparer différents auteurs, époques, genres ou formes du discours ainsi que d'étudier l'évolution de la longueur de la phrase, à l'intérieur d'une œuvre comme dans plusieurs œuvres successives d'un même auteur. Dans notre étude de la longueur de la phrase leclézienne, nous nous intéresserons aux deux aspects, la comparaison exogène aussi bien que l'étude endogène.

La longueur moyenne de la phrase s'obtient en divisant le nombre d'occurrences du corpus par le nombre de ponctuations fortes. La division des 2.281.659 occurrences de notre corpus par les 108.152 signes de ponctuation forte permet ainsi d'établir la longueur moyenne de la phrase de Le Clézio qui est de 21.09 mots par phrase.

Lorsque l'on compare cette longueur avec celle d'autres auteurs littéraires, il est également important de distinguer les différentes époques car la longueur de la phrase a évolué au fil de siècles.

En comparaison avec les études qu'a effectuées Étienne Brunet sur les auteurs du XIX^{ème} siècle²⁷⁷, nous constatons que la phrase leclézienne est plus longue que la phrase moyenne de Zola qui elle est de 15.82 mots, que celle de Hugo de 15 mots ou de la phrase du corpus T.L.F. XIX^{ème} – XX^{ème} siècles qui a une moyenne de 14.60 mots.

En revanche, la phrase de Le Clézio est plus courte que la phrase de Chateaubriand contenant 22.23 mots et de celle de Rousseau de 27.71 mots. La phrase la plus longue est évidemment celle de Proust avec 30.9 mots²⁷⁸, qui est exceptionnelle.

²⁷⁷ É. Brunet (1988) : p. 77.

²⁷⁸ É. Brunet (1983) : p. 124.

Étienne Brunet a pu observer qu'en règle générale "la phrase se raccourcit depuis 1789". La phrase de Le Clézio ne s'inscrit pas dans ce mouvement, la relative longueur de la phrase leclézienne reflète peut-être une tendance plus récente ; celle de l'allongement de la phrase dans la dernière moitié du XX^{ème} siècle ?

Dans ses études sur le vocabulaire du roman français contemporain, Gunnel Engwall²⁷⁹ a étudié la longueur moyenne de la phrase dans 25 romans publiés entre 1962 et 1968. Ces romans, écrits par 25 écrivains différents, présentent de très grandes variations avec une longueur de phrase qui varie de 7.8 à 23.4 mots.²⁸⁰ La phrase en moyenne la plus longue de son étude était celle de Georges Perec dans *Les choses*, qui contenait 23.4 mots. Dans son corpus se trouvait également *Le déluge* de J.M.G. Le Clézio dont les résultats se situent tout à fait au milieu de l'analyse, en dixième position parmi ses sous-corpus, avec une longueur moyenne de la phrase de 16.0 mots dans les dix sections aléatoirement distribuées dans le roman. Si Gunnel Engwall l'a choisi pour effectuer une étude approfondie "c'est parce qu'il ne se distingue pas particulièrement des autres romans du corpus". Comparée à la phrase de Giraudoux de 20.64 mots (compte tenu des romans exclusivement), la longueur moyenne de la phrase leclézienne de 21.09 mots semble tout à fait s'inscrire dans le mouvement littéraire de la dernière moitié du XX^{ème} siècle.

Dans sa comparaison des corpus de genres différents, Margareta Östman a pu constater des longueurs de phrases fort diverses²⁸¹. La longueur moyenne des phrases du corpus provenant des actes législatifs des Communautés européennes²⁸² était de non moins de 31 occurrences par phrase tandis que les

²⁷⁹ G. Engwall, "G.Perec : les Choses – sous quelques aspects quantitatifs" in *Actes du 6^e Congrès des Romanistes Scandinaves*, Uppsala 11-15 août 1975, p. 67-68.

²⁸⁰ Etant donné que cette étude ne comprend qu'un seul roman de chaque auteur (chaque roman entrecoupé en dix sections), il convient d'être prudent quant aux conclusions sur les habitudes caractéristiques des différents écrivains.

²⁸¹ M. Östman, "Quatre (un?) corpus – quatre genres, quelques résultats." in *JADT 1998, 4^{èmes} Journées Internationales d'Analyse Statistique des Données Textuelles*, S. Mellet (éd.), Nice, 1998, p. 489.

²⁸² Textes datant des années 1988-1993.

textes provenant des romans²⁸³ ou des manuels de français commercial²⁸⁴ ne comptaient que 14 mots par phrase.

Concernant l'étude sur le vocabulaire du discours politique, Jean-Marie Cotteret et René Moreau avait aussi observé chez le général de Gaulle un allongement de la phrase entre 1958 et 1969²⁸⁵. En effet, on voit dans les discours du général, pendant la période étudiée, une évolution très nette vers des phrases de plus en plus longues – elles ont une valeur moyenne d'environ 21 mots vers 1958 et d'environ 31 vers 1969. Dominique Labbé a fait le même calcul pour les discours de François Mitterrand²⁸⁶ où la longueur moyenne des phrases est de 33.2 vocables. Ce corpus ne semble pas montrer de tendance d'allongement avec le temps mais il est à noter que la phrase la plus brève se trouve dans le seul sous-corpus destiné à l'écrit, *Lettre à tous les Français*, avec la longueur moyenne de 23.9 mots, un taux plus proche des textes littéraires et de notre auteur. En effet, la valeur est souvent un peu plus courte dans les textes écrits que dans les discours oraux “ce qui prouve, selon Dominique Labbé²⁸⁷, que la rédaction permet d'écarter certaines aberrations de l'expression spontanée.” On peut penser aussi que les discours politiques (en principe oraux mais en réalité écrits) appartiennent au genre oratoire, qui aime la période longue et développée.²⁸⁸

L'étude de l'évolution de la longueur de la phrase à l'intérieur d'une œuvre ou d'un corpus apporte des informations intéressantes. Le tableau ci-dessous, illustrant la longueur moyenne des différents textes, permet de constater que la longueur de la phrase leclézienne n'est pas régulière :

²⁸³ Extraits de romans français publiés entre 1962 et 1968.

²⁸⁴ Manuels de français commercial publiés entre 1974 et 1977.

²⁸⁵ J.-M. Cotteret et R. Moreau (1969) : p. 42-50.

²⁸⁶ D. Labbé (1990) : p. 178-179.

²⁸⁷ D. Labbé (1991) : p. 178.

²⁸⁸ Cf. Cicéron, Bossuet ou Rousseau.

La longueur de la phrase							
Texte	occurrences	ponctuations	longueur	Texte	occurrences	ponctuations	longueur
Pv	92955	4247	21,89	Ch	134435	6251	21,51
Fi	101666	4940	20,58	An	24313	1209	20,11
De	122269	6320	19,35	Rd	38238	1110	34,45
Ex	90616	5001	18,12	Re	87414	2527	34,59
Fu	104451	6608	15,81	Pi	70230	3750	18,73
Gu	110602	5857	18,88	Si	2463	92	26,77
My	9771	532	18,37	On	76046	4169	18,24
Vo	113790	6009	18,94	Et	118756	5650	21,02
Pr	8690	346	25,12	Pa	10593	488	21,71
Mo	99114	5325	18,61	Di	73266	2092	35,02
In	124433	5886	21,14	Qu	165612	8537	19,40
Ic	2637	129	20,44	Po	86051	4480	19,21
Ar	2910	161	18,07	Fe	71025	2215	32,07
Ds	150843	5881	25,65	Nu	21563	843	25,58
Vi	15266	664	22,99	Ha	68261	3249	21,01
Ro	83380	3584	23,26				

Tableau n°27 : La longueur de la phrase leclézienne (ponctuations fortes).

La valeur moyenne est de 22.2²⁸⁹. Dans ses 31 ouvrages la moyenne varie entre 15.81 et 34.59. L'intervalle²⁹⁰ est donc de 18.78. La variation standard est, selon le calcul de l'écart type²⁹¹, de 4.8. L'histogramme de la page suivante permet de visualiser ces grandes variations dans le corpus :

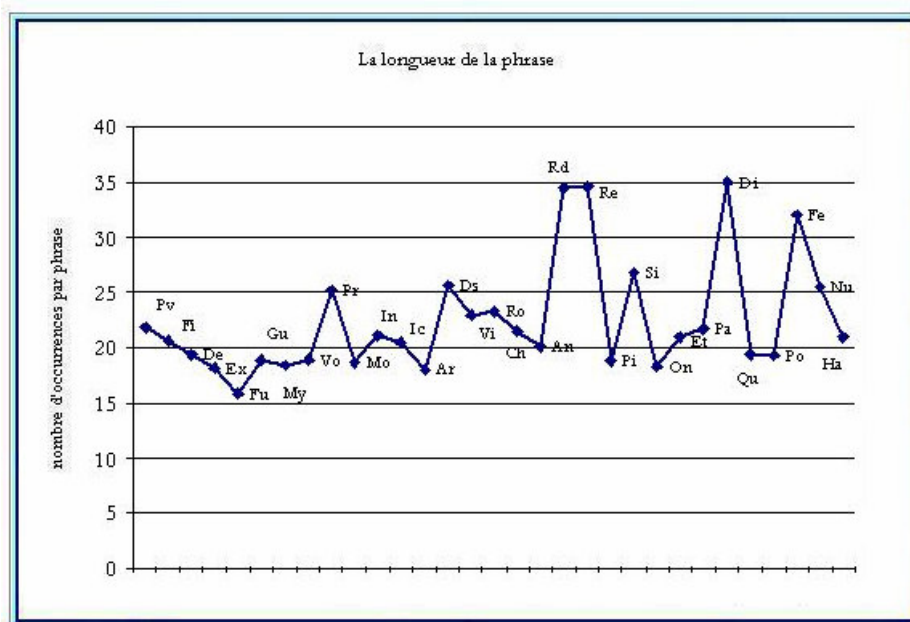


Figure n°56 : La longueur de la phrase leclézienne (ponctuations fortes).

²⁸⁹ 22.2 est en réalité une moyenne des moyennes, en négligeant l'étendue inégale des œuvres (ce qui donne davantage d'importance aux œuvres courtes).

²⁹⁰ L'intervalle de variation est la différence entre les deux valeurs extrêmes d'un variable.

²⁹¹ Pour plus de détails sur ce calcul voir Körner S. et Wahlgren L. (1998) : p. 47.

Nous constatons que la phrase leclézienne ne change pas avec le temps, qu'il n'y a aucune tendance chronologique. La longueur de la phrase dépend, en réalité, surtout du genre. Le relatif allongement que nous pouvons constater vers la fin de l'œuvre provient des phrases très longues des biographies, des ouvrages ethnologiques et des essais de la dernière partie du corpus.

La phrase la plus longue se trouve dans la biographie *Diego et Frida* avec une moyenne de 35.02 mots par phrase. Ce texte très descriptif traite de la vie des deux peintres, leur art et leur milieu, le Mexique, ainsi que les pays qu'ils visitent, notamment les États Unis, et toutes ces longues descriptions et énumérations influencent évidemment la longueur des phrases, comme par exemple celle-ci, de 79 mots²⁹² :

“Depuis son retour au Mexique, Diego Rivera a commencé la plus extraordinaire collection d'objets d'art préhispanique, qu'il achète au cours de ses voyages en province ou au marché du Volador à Mexico : statuettes d'argile de Colima, masques de bébés souriants de la région olmèque, chiens nus de Tezcoco, figurines de fertilité du Nayarit - collection rassemblée aujourd'hui dans le musée que le peintre a fait construire pour elle, l'Anahuacalli (la maison de l'Anahuac).”

Les œuvres ethnologiques contiennent également des phrases d'une longueur importante, manifestant la richesse de descriptions détaillées consacrées à la civilisation amérindienne et l'Amérique précolombienne. La longueur de la phrase moyenne de la *Fête chantée* est de 32.07, celle de l'essai *Le rêve mexicain* est de 34.59 mots par phrase.

Ces différences de longueur de phrase dépendant étroitement du genre littéraire dans notre corpus reflètent une tendance générale de la langue. Étienne Brunet en

²⁹² *Diego et Frida*, p. 89.

souligne l'importance dans ses différents ouvrages²⁹³. La phrase est par exemple beaucoup plus longue dans les textes techniques que dans la prose littéraire, plus longue dans la prose poétique que dans les vers, plus longue lorsque domine la troisième personne que lorsqu'interviennent la première et la deuxième personnes.

Pour cette raison, nous avons trouvé intéressant, encore une fois, d'isoler le corpus romanesque de Le Clézio et d'étudier les variations à l'intérieur de ce sous-corpus pour tenter de neutraliser le facteur générique dans les variations observées.

La courbe ci-dessous illustre la variation de la longueur de la phrase dans le corpus "Roman" :

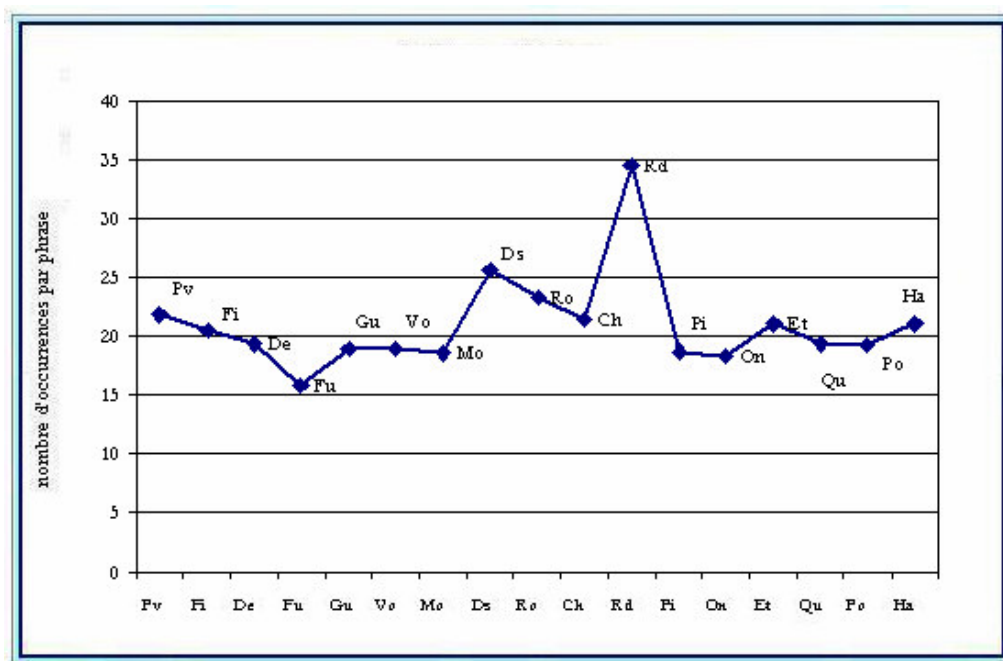


Figure n°57 : La longueur de la phrase dans le corpus romanesque, le corpus D.

²⁹³ Cf. notamment É. Brunet (1981) : p. 274-275.

La courbe de la longueur moyenne de la phrase romanesque de Le Clézio est, à part quelques exceptions, beaucoup plus uniforme avec une moyenne de 21.32 mots.

Dans ce sous-corpus romanesque, la phrase la plus longue se trouve dans *Voyage à Rodrigues*, avec la longueur exceptionnelle de 34.45. Dans ce livre, en partie autobiographique, Le Clézio raconte le voyage de son grand-père à l'île Rodrigues avec de nombreuses descriptions des paysages. Mais la raison principale de la longueur de la phrase réside toutefois ailleurs ; il s'agit plutôt d'un effet de la particularité du texte qui fausse les statistiques. Le livre contient en effet de nombreuses indications et annotations de plan qui guident le jeune chercheur du trésor à travers le roman. Ces séquences, comme l'exemple ci-dessous, sont souvent dépourvues de signes de ponctuation²⁹⁴ :

M / diagonale dans la direction du comble du commandeur

OM = _ 72° 30°

OM = _ 259 pieds français.

À l'ouest, vers le haut du plan, cette annotation à l'encre :

Décembre 1915

Solution Z (z)

1° 30'

209 pieds français

barrée d'un trait de crayon, avec ce rajout :

15 . 8 . 1918 : Solution (5)

À l'angle Nord - Ouest, au bas du plan, encore une date :

Nous y trouvons également et des pages entières de cryptogrammes²⁹⁵ :

“De même que toutes les traces du Corsaire se retrouvent dans la grille

²⁹⁴ *Voyage à Rodrigues*, p.74.

²⁹⁵ *Voyage à Rodrigues*, p. 113-114.

- le plan de la rivière et de la côte des documents Savy -, de même tous les secrets du langage de cette quête sont dans le carré formé par les vingt-cinq lettres de l'alphabet (en ce temps-là on ignore le w), plus une :

a b c d e f g h i j k l m n o p q r s t u v x y z a
b c d e f g h i j k l m n o p q r s t u v x y z a b
c d e f g h i j k l m n o p q r s t u v x y z a b c
d e f g h i j k l m n o p q r s t u v x y z a b c d
e f g h i j k l m n o p q r s t u v x y z a b c d e
f g h i j k l m n o p q r s t u v x y z a b c d e f
g h i j k l m n o p q r s t u v x y z a b c d e f g
h i j k l m n o p q r s t u v x y z a b c d e f g h
i j k l m n o p q r s t u v x y z a b c d e f g h i
j k l m n o p q r s t u v x y z a b c d e f g h i j
k l m n o p q r s t u v x y z a b c d e f g h i j k
l m n o p q r s t u v x y z a b c d e f g h i j k l
m n o p q r s t u v x y z a b c d e f g h i j k l m
n o p q r s t u v x y z a b c d e f g h i j k l m n
o p q r s t u v x y z a b c d e f g h i j k l m n o
p q r s t u v x y z a b c d e f g h i j k l m n o p
q r s t u v x y z a b c d e f g h i j k l m n o p q
r s t u v x y z a b c d e f g h i j k l m n o p q r
s t u v x y z a b c d e f g h i j k l m n o p q r s
t u v x y z a b c d e f g h i j k l m n o p q r s t
u v x y z a b c d e f g h i j k l m n o p q r s t u
v x y z a b c d e f g h i j k l m n o p q r s t u v
x y z a b c d e f g h i j k l m n o p q r s t u v x
y z a b c d e f g h i j k l m n o p q r s t u v x y
z a b c d e f g h i j k l m n o p q r s t u v w x y z
a b c d e f g h i j k l m n o p q r s t u v w x y z a

“Cette écriture n’est pas une appropriation. Elle est pareille au lieu qu’elle décrit : minutieuse, précise, secrète”, écrit Le Clézio²⁹⁶, dans ce roman très particulier de notre corpus.

Les phrases les plus courtes se trouvent dans *Le livre des fuites* avec une longueur moyenne de phrase de 15.3 mots. Les phrases courtes de ce roman, caractérisé par la hantise, l’angoisse et la fuite de notre monde moderne contribue à donner un rythme haletant à ce récit riche en dialogues et en monologues souvent saccadés²⁹⁷ :

“Tu as de l’argent ?” “Oui.” “Combien ?” “Tiens , regarde.” “OK.”
“Alors ?” “Coca, avec du rhum.” “Cuba Libre.” “Quoi ?” “Ça s’appelle Cuba Libre.” “Sais pas.” “C’est vrai.” “Comment tu t’appelles ?” “Jeune Homme.” “Quoi ?” “Jeune, Homme.” “Pas un nom, ça.” “C’est comme ça qu’on m’appelle.” “Hm-hm.” “Et toi ?” “Quoi ? ” “Comment tu t’appelles ?” “Ricky.” “Ricky ?” “Oui.” “D’où tu es ?” “Tobago.” “Longtemps que tu es ici ?” “Deux ans. Et toi ?” “Je suis arrivé ce soir.” “Tu viens ici ?” “Non.” “Où tu vas ?” “Peut-être à Tobago.” “Ah oui ? ” “Ou ailleurs, peut-être.” “Oui , ah oui.” “Et toi ?” “Quoi ?” “Tu vas rester ici ?” “Sais pas . J’irai peut-être à Londres.” “Oui ?” “Faire de la danse. Tu connais le Six Bells ?” “Non.” “C’est une boîte.” “Ah bon.”

Cet exemple illustre aussi, de manière extrême, l’effet du dialogue et du récit à la première et à la deuxième personne sur la longueur de la phrase. Les phrases sont plus courtes dans les livres où il y a beaucoup de dialogues comme dans *La guerre*, *Mondo et autres histoires*, *Printemps et autres saisons* et dans *La quarantaine*.

Les critiques ont parfois dit que les courtes phrases des nouvelles de *Mondo et autres histoires* semblent avoir été écrites par des enfants. “C’est le plus bel éloge que l’on

²⁹⁶ *Voyage à Rodrigues*, p. 83.

²⁹⁷ *Le livre des fuites*, p. 76-77.

puisse faire à Le Clézio”, affirme Yves Sinturel²⁹⁸. C’est une telle écriture que prône Le Clézio dans *L’inconnu sur la terre* : “ On est tout petit [...], si petit qu’aucune pensée ne pourrait rester longtemps, qu’aucune phrase très longue ne pourrait se faire. ”²⁹⁹

Toutefois, la seule ponctuation forte ne suffit pas à mesurer les phrases, ce qui nous a amenée à prendre en compte aussi la ponctuation faible - comprenant la virgule, les deux points, les paires de parenthèses et le point-virgule - pour avoir une idée plus précise du rythme des pauses tout au long du texte.

²⁹⁸ Y. Sinturel (1988), in S. L. Becket (1997) p. 228.

²⁹⁹ *L’inconnu sur la terre*, p. 288.