

Quatrième partie :

Les parties du discours ; analyse “grammatico-métrique”.

4.1. Introduction.

“Ces mots que j’aime ainsi : quand, dépouillés de toute fantaisie, ils n’ont de signification que dans le système lexical, par rapport à d’autres mots...”

Dans cet état pur, les mots sont semblables à des animalcules primitifs, à des protozoaires. Le substantif, noyau. L’adjectif, extension du substantif comme une membrane adductive. Le verbe, flagelle. Les struments, corps inanimés, véhicules nutritifs. C’est ce langage-là, élémentaire, véridique, que je voudrais parler. C’est quand ils sont si près de la mort que les mots sont profondément dans la vie.”

L’extase matérielle, page 42.

Les parties du discours, que l’on appelle aussi catégories grammaticales, sont une classification des mots qui nous a été transmise par la syntaxe traditionnelle, basée sur trois critères qui sont la nature, la fonction et la position.

Quel est l’intérêt d’analyser la distribution des parties du discours dans un corpus littéraire informatisé, et de quelle façon cette analyse peut-elle contribuer à la compréhension de l’utilisation et de la fonction des catégories grammaticales ?

“Investigating this question, écrit Douglas Biber³⁴², can help to understand how different varieties exploit the grammatical categories of words available to them... Using the corpus, we can analyze the distribution and function of these different categories of words and study the part that they play in fulfilling the communicative function of different registers.”

En effet, la distribution des parties du discours n’est pas constante, elle varie selon les époques, les auteurs et les genres. Déjà l’étude de Guiraud³⁴³ montre que la

³⁴² D. Biber, S. Conrad, R. Reppen (1998) : p.57.

³⁴³ P. Guiraud (1954).

distribution des catégories grammaticales dans les ouvrages littéraires est fortement influencée par le genre et l'époque où ils s'inscrivent.

Étienne Brunet estime que l'emploi des catégories grammaticales dans un texte donné peut être un indice très révélateur³⁴⁴ :

“C'est bien en effet dans la distribution des catégories grammaticales et principalement des classes nominale et verbale, que se manifeste la distinction des styles, des genres et des écrivains.”

En effet cette distribution, qui est bien un critère de distinction des œuvres, manifeste peut-être des choix plus subtils que celle du vocabulaire – en tout cas moins liés à la thématique de chaque ouvrage, ce qui peut apporter à l'analyse des éléments nouveaux.

Il s'agit en réalité de choix inconscients faits par l'auteur lors de la création et de l'élaboration d'un texte qui permettent au chercheur de distinguer des divisions grammaticales caractéristiques et personnelles. Selon Étienne Brunet³⁴⁵:

“...on croit saisir là une division pertinente, où l'on ne peut pas ne pas rejoindre le choix conscient de l'écrivain. Mais il serait difficile de prétendre que le choix de l'écrivain s'exerce délibérément dans ces domaines. De la même façon un homme peut se reconnaître à ses empreintes digitales, mais on n'a pas conscience de ses empreintes et on ne les sent pas au bout des doigts.”

Peut-on à partir de l'étude des parties du discours mieux comprendre l'écriture leclézienne ? Quels sont les particularités grammaticales qui contribuent à son style d'écriture ? Dans cette étude, qui ne prétend pas être exhaustive, nous nous attarderons, après avoir étudié la distribution des catégories grammaticales du

³⁴⁴É. Brunet (1983) : p. 836.

³⁴⁵É. Brunet (1981) : p. 295.

corpus, sur des aspects morphologiques et syntaxiques qui nous ont paru intéressants et révélateurs dans la catégorie nominale aussi bien que dans la catégorie verbale.

Malheureusement les parties du discours ne sont pas des catégories entièrement stables et étanches les unes des autres. Si dans chaque contexte l'appartenance d'un mot à une catégorie précise s'impose sans grande difficulté, en revanche le même mot rentrera dans une autre catégorie pour peu que change le contexte, même si le contenu sémantique reste constant. Les cas de *Manger pour vivre* et *Le manger fait vivre* ou bien *Tout est bon* et *Le tout est bon* sont des exemples de ce phénomène relativement courant dans la langue française.

En effet, les fondements et les méthodes sur lesquels s'appuient les études des catégories grammaticales varient d'un ouvrage à l'autre. Il convient donc, avant de nous lancer dans l'étude de catégories grammaticales dans l'œuvre leclézienne³⁴⁶ - qui s'appuiera principalement sur le grand corpus lemmatisé (le corpus B) et sur le corpus exclusivement romanesque également lemmatisé (le corpus E) -, de nous attarder un peu sur les principes et les techniques qui guideront cette étude.

³⁴⁶ Cette étude n'a pas la prétention d'être exhaustive et nous ne traiterons pas, faute de place, systématiquement toutes les catégories grammaticales comme celle notamment des mots-outil.

4.2. Les méthodes d'analyse.

“Il invente aussi une langue, une véritable langue avec ses mots, ses règles de grammaire, son alphabet, sa symbolique, une langue pour rêver plus que pour parler, une langue pour s’adresser au monde étrange dans lequel il a choisi de vivre.”

Voyage à Rodrigues, page 103.

Désormais la quantification et la lemmatisation des corpus ouvrent la voie à cette composante essentielle de l’écriture qu’est l’étude de la distribution des parties du discours. Cette analyse, qui demande l’accès à la forme canonique du mot, au lemme, ne peut guère se fonder sur la distribution des effectifs d’un corpus s’appuyant sur la forme graphique. C’est la lemmatisation qui permet d’étiqueter le corpus selon les catégories grammaticales et de classer les éléments du vocabulaire selon leur appartenance à une catégorie spécifique.

Les codes grammaticaux fournis par l’étiqueteur morphosyntaxique au cours de l’opération de lemmatisation “automatique” constituent ici un outil indispensable.

Toutefois, la lemmatisation automatique repose sur des critères préétablis selon des méthodes d’analyse différentes³⁴⁷ et les résultats peuvent varier selon le choix de l’étiqueteur et de ses principes. Il faut également savoir que la lemmatisation automatique n’est pas sans erreurs : erreurs d’analyse contextuelles, mauvaise définition de certaines classes grammaticales, etc. Il est donc recommandé de tenir compte de ces imperfections lors de l’exploitation des résultats.

Les parties du discours distinguées par les logiciels d’analyse morpho-syntaxique sont toutefois nettement plus fiables que les fonctions grammaticales car leur reconnaissance est plus facile. Pour beaucoup de mots qui ne sont pas sujets à l’homographie, le codage est automatique et indépendant du contexte : la

³⁴⁷ Cf. chapitre 1.5. “La lemmatisation”.

proposition du dictionnaire, unique, est immédiatement acceptée. Et quand il s'agit de diviser les homographes en catégories différentes (par exemple les cas très nombreux, du type *la marche/il marche*, où le substantif peut se confondre avec un verbe), une analyse syntaxique simple, désormais quasi-automatique (selon les différents logiciels), emporte la décision. D'autres cas peuvent être plus difficiles, comme la reconnaissance de l'adverbe *fort* dans le syntagme *un homme fort bon*. Néanmoins l'étiquetage morphologique est généralement plus fiable que le "parsing" syntaxique qui recourt à des modèles et des procédures complexes.

Dans cette étude nous exploitons différentes versions de notre corpus ; d'une part, nous appuyant sur le corpus des 31 œuvres, une première version lemmatisée avec Winbrill (corpus K)³⁴⁸ et une deuxième version lemmatisée avec Cordial 7 (corpus B) ; d'autre part, nous appuyant sur un corpus réduit de douze œuvres, une version lemmatisée avec Cordial 7 (corpus G) et une autre lemmatisée selon la méthode Labbé (corpus I)³⁴⁹. Nous avons aussi souvent recours au corpus exclusivement romanesque dans sa version lemmatisée.

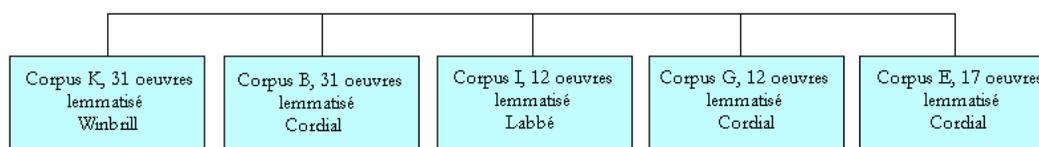


Figure n°77 : Différentes versions du corpus exploitées dans l'étude "grammatico-métrique".

Winbrill, qui était le seul lemmatiseur dont nous disposions au début de notre recherche, a été abandonné à cause des nombreuses difficultés rencontrées (erreurs d'analyse et confusions entre classes grammaticales). En effet, dès l'accès à un autre analyseur intégré au logiciel Hyperbase, nous avons opté pour ce nouvel outil et c'est au bout du compte le corpus B lemmatisé avec Cordial 7 qui sera notre référence fondamentale, étant donné qu'il couvre la totalité de notre grand corpus.

³⁴⁸ Cf. chapitre 1.5.4. "Winbrill".

³⁴⁹ Pour le détail sur les normes de saisie et de dépouillement voir D. Labbé : (1990).

4.3. La distribution des catégories grammaticales du corpus.

“Le langage n’est pas une “expression” ni même un choix ; c’est être soi. De même, être vêtu de telle ou telle manière, porter une cravate, se coiffer longuement ou rapidement, laver plus consciencieusement telle ou telle partie de son corps sont des actes qui révèlent la direction du regard intérieur. C’est dire l’importance de cette coalition entre la forme et le fond.”

L’extase matérielle, page 51.

La version lemmatisée d’Hyperbase selon le programme Cordial 7 aboutit au bout du traitement à quelque 200 codes grammaticaux différents en utilisant toutes les combinaisons possibles. Le logiciel Hyperbase regroupe par la suite les codes et fournit la liste des fréquences.

Nous en avons extrait les 11 catégories fondamentales parmi celles que propose le programme Cordial ; verbes, substantifs, adjectifs, déterminants, pronoms, numéraux, interjections, prépositions, adverbes, conjonctions et délimiteurs (signes de ponctuations). Le tableau ci-dessous présente la distribution des parties du discours du corpus de 31 œuvres selon les effectifs :

N°	TITRE	VERB	SUBST	ADJ	DETER	PRONOM	NUME	INTERJ	PREP	ADV	CONJ	PONCT
1	PROCES	13104	17200	4468	11340	10422	631	135	9962	5350	4364	14043
2	FIÈVRE	13619	19137	5615	13525	10147	403	132	10570	5883	4424	16464
3	DÉLUGE	17084	24333	6930	17343	12074	578	152	12717	6452	5402	18916
4	EXTASE	13156	15588	5328	11989	11061	160	6	7984	6231	4295	14446
5	FUITE	13962	21700	5425	14026	10973	394	74	10090	5525	3652	16873
6	GUERR	15207	22147	5712	15418	12161	568	111	10928	6072	4792	15973
7	MYDRIASE	1353	1979	546	1476	1028	18	3	901	639	385	1498
8	VOYAGES	16667	21311	5795	14866	14184	226	74	10621	7633	5444	16201
9	PROPHÉTIES	965	2295	498	1708	513	86	10	801	344	277	1237
10	MONDO	14616	19135	4592	14305	10417	133	54	9601	5937	5142	13243
11	INCONNU	16162	23217	7213	19072	14284	239	25	10930	7343	5872	19509
12	ICEBERGS	396	459	118	330	347	6	0	242	224	123	463
13	ARBRES	463	509	182	419	376	5	0	226	216	180	397
14	DÉSERT	20729	30210	7344	23106	15627	170	60	15477	8469	7903	20308
15	VILLE	1895	3530	844	2757	1210	54	11	1420	699	604	2268
16	RONDE	12453	15354	3936	11500	9606	177	35	8416	5130	4223	11756
17	CHERCHEUR	19178	27104	5475	20522	15281	342	24	14576	6507	5411	18763
18	ANGOLI	3905	4849	988	3647	2545	47	0	2593	1180	1122	3286
19	RODRIGUES	4481	9062	1941	6264	3266	169	8	4217	1713	1465	5408
20	RÊVE	9411	21701	5129	15692	5951	294	20	9317	2845	3933	11356
21	PRINTEMPS	11606	12207	2892	9004	9773	124	13	6552	4052	2901	10439
22	SIRANDANES	283	528	155	420	225	0	1	226	144	106	380
23	ONITSHA	10887	17267	3201	11719	7021	145	15	7941	3335	2507	11174
24	ETOILE	18729	23005	4623	17313	13331	165	20	12125	6018	5267	17019

25	PAWANA	1558	2191	437	1784	1095	16	2	1092	507	399	1488
26	DIEGO	8151	17942	4660	11667	5703	298	3	8966	2994	3063	8901
27	QUARANTAI	25398	33810	7232	24427	18432	372	31	17653	7420	5995	23173
28	POISSON	14640	15026	3438	10233	12384	152	29	8101	4728	4156	12199
29	FÊTE	7627	17858	4091	12606	4784	253	13	7903	2599	3038	9258
30	NUAGES	2539	5470	1055	3783	1675	81	3	2385	820	797	2927
31	HASARD	10884	13833	3124	9661	7227	147	21	7531	3340	2690	9195
TOTAL		321108	459957	112987	331922	243150	6453	1085	232064	120349	99932	328561
TOTAL OCCURENCES		2257568										

Tableau n°33 : La distribution des parties du discours (effectifs).

Avant de commenter les variations importantes entre les catégories différentes dans les divers ouvrages de notre corpus dont témoigne ce tableau, nous nous proposons de regarder le corpus dans son ensemble. La figure ci-dessous rend compte du classement hiérarchique des catégories grammaticales du corpus B, exprimé en pourcentage :

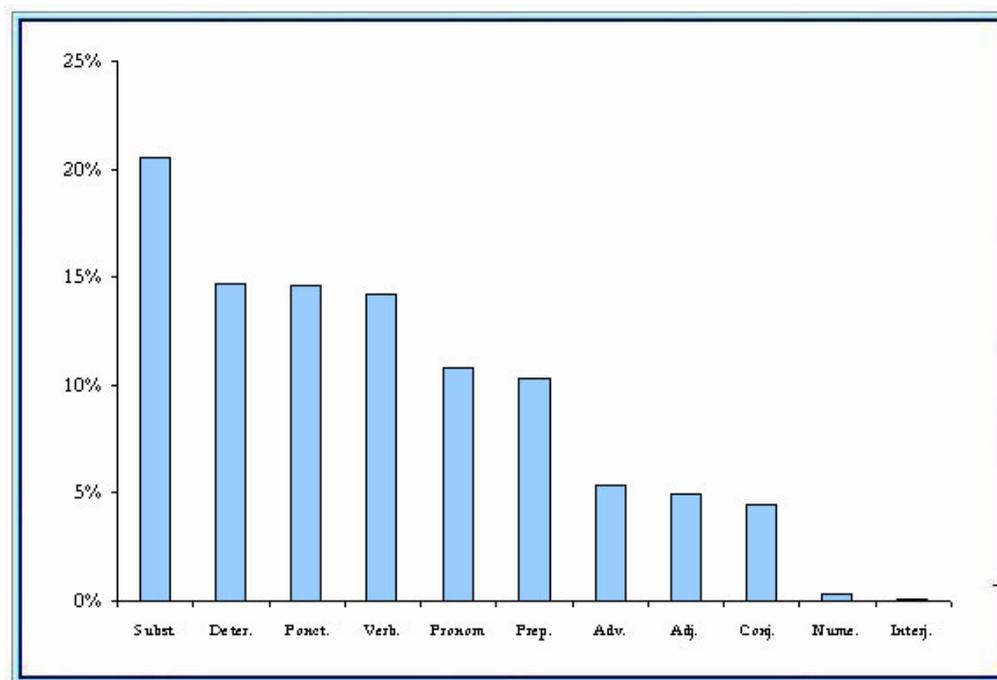


Figure n°78 : Classement hiérarchique des catégories grammaticales du corpus B exprimé en pourcentages (fréquences absolues).

Ce graphique permet de constater la richesse en substantifs (20,5%) de notre corpus, une des caractéristiques des corpus de langue française. Les catégories des déterminants, des signes de ponctuation et des verbes occupent chacune entre 14 et 15% du corpus. Quant aux pronoms et aux prépositions, ils se regroupent autour de 10% chacun tandis que les adverbes, les adjectifs et les conjonctions se rassemblent autour de 5% chacun. Les catégories de numéraux et d'interjections sont dérisoires par leur nombre.

Pierre Guiraud³⁵⁰ avait déjà observé que dans la littérature le nombre de substantifs et celui des verbes varient en proportion inverse, le substantif étant dominant dans la prose abstraite et le verbe dans les récits. La richesse du substantif dans l'écriture leclézienne, caractéristique partagée avec les grands écrivains du XIX^{ème} siècle, est étonnante car on aurait pu croire que les romans et les nouvelles de Le Clézio étaient plus proches des récits et donc devaient favoriser le verbe.

4.3.1. Comparaison externe.

En effet, l'utilisation de ces différentes catégories grammaticales varie d'un écrivain à l'autre. Étienne Brunet fait une étude très détaillée de la distribution des catégories grammaticales dans ses ouvrages sur les vocabulaires de Proust, de Zola et de Hugo³⁵¹. Il a pu observer que parmi les écrivains du XIX^{ème} siècle³⁵², Chateaubriand était favorable à la catégorie nominale (adjectifs et substantifs), Zola cultivait plutôt les participes et les verbes, Proust accumulait les mots de relation, Hugo avait un penchant pour le substantif et le verbe et enfin Giraudoux fuyait les adjectifs, en cherchant, tout comme Hugo, les noms et les verbes. Le Clézio s'approche dans son profil de ces deux derniers, avec un goût prononcé

³⁵⁰ P. Guiraud (1954): p 104.

³⁵¹ Cf. É. Brunet (1988) : p. 171-172.

³⁵² La norme utilisée dans ses études est celle de la fréquence des catégories grammaticales dans le corpus du T.L.F.

pour le nom, et comme Giraudoux, une aversion pour l'adjectif, nous y reviendrons plus en détail dans ce chapitre.

Toutefois, les variations que l'on peut observer à l'intérieur d'un corpus peuvent parfois être aussi importantes qu'entre différents écrivains. Il convient donc, avant de se prononcer sur la distribution des catégories grammaticales d'un corpus aussi varié que le nôtre, de prendre en considération la distribution à l'intérieur du corpus.

4.3.2. Étude interne.

En effet, si nous regardons d'un peu plus près les catégories grammaticales dans notre corpus, nous constatons que leur distribution n'est pas régulière. Pour une vision synthétique des accords qui lient les codes grammaticaux et les différents sous-corpus, nous avons recours à l'analyse factorielle de la liste de fréquences de ces différentes classes du corpus :

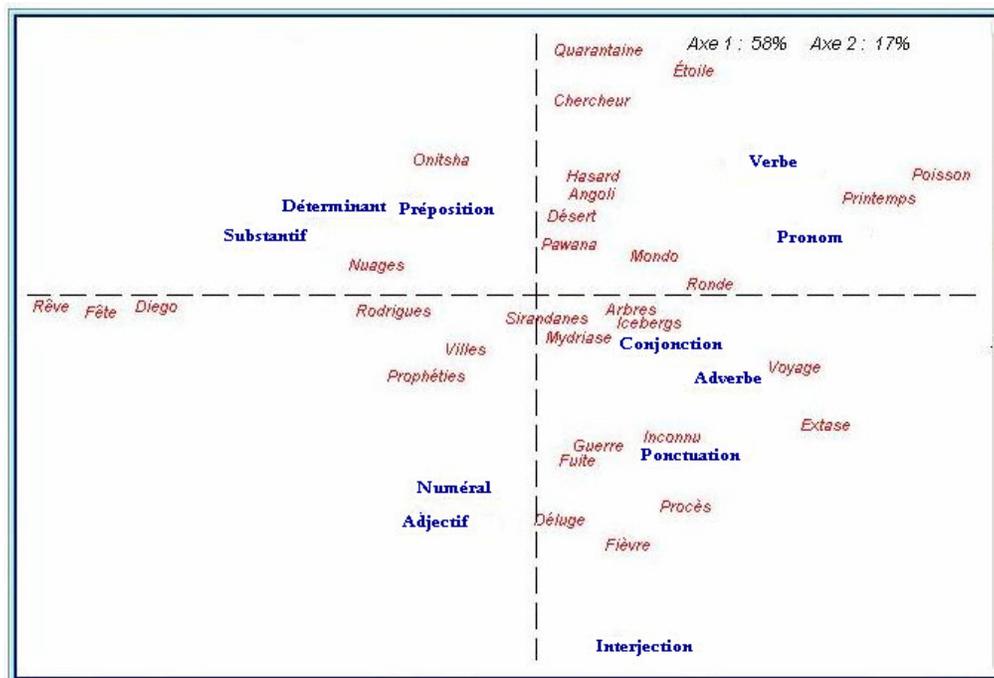


Figure n°79 : Analyse factorielle de la distribution grammaticale selon la lemmatisation par Cordial 7 du corpus B.

Nous voyons que le premier facteur oppose la catégorie verbale à la catégorie nominale. Le substantif à gauche attire les prépositions, les déterminants et les adjectifs tandis que le verbe en haut à droite attire les pronoms et les adverbes.

Le second facteur parcourt la chronologie de l'écrivain du bas vers le haut du graphique. Les premiers ouvrages, *Le procès-verbal*, *La fièvre*, *Le déluge*, *La guerre* et *Le livre des fuites* se trouvent en bas du graphique autour des catégories secondaires qui témoignent d'une écriture foisonnante (adjectifs, adverbes et interjections). Les derniers romans, *La quarantaine*, *Poisson d'or*, *Etoile errante*, *Le chercheur d'or* et *Hasard*, se situent en haut du tableau autour des catégories fondamentales, témoignant peut-être d'un assagissement de l'écriture, d'un travail de simplification de style. Nous y reviendrons dans l'étude détaillée des différentes catégories grammaticales.

L'analyse factorielle rend également compte de l'opposition générique. Les ouvrages ethnologiques se regroupent à l'extrême gauche du graphique, les premiers romans appartenant à l'école du "nouveau roman" en bas à droite, tandis que les œuvres fictionnelles se trouvent au centre supérieur du tableau. Les ouvrages qui se trouvent au milieu sont les plus courts, tous genres confondus.

Il est intéressant de comparer ce tableau avec une analyse basée sur un même corpus traité par deux lemmatiseurs différents. Nous nous proposons d'exploiter, dans cette étude, le corpus de 12 œuvres.

La version du corpus lemmatisée selon la méthode de Dominique Labbé trie les occurrences en une vingtaine de catégories fondamentales :

1. Verbe :

11 forme fléchie

12 forme au participe passé

13 forme au participe présent

14 forme à l'infinitif

- 2. Substantif :
 - 20 “nom propre” (mot à majuscule initiale)
 - 21 substantif masculin
 - 22 substantif féminin
 - 3. Adjectif :
 - 30 Adjectif “pur”
 - 31 Participe dans un emploi “adjectivé”
 - 5. Pronom :
 - 51 Personnel
 - 52 Relatifs, réfléchis, interrogatifs, possessifs, etc.
 - 6. Adverbes
 - 7. Déterminant:
 - 71 Articles (défini et indéfini)
 - 72 Numéraux et cardinaux
 - 73 Possessifs
 - 74 Démonstratifs
 - 75 Adjectifs indéfinis
 - 81. Préposition
 - 82. Conjonction
 - 91, 92, 93. Locution, expression, interjection
- Ponctuation :
- “p” : ponctuation mineure (interne à la phrase)
 - “P” : ponctuation majeure (délimitant la phrase)

Le logiciel Hyperbase regroupe par la suite les codes et fournit la liste des fréquences. Cette liste permet de voir la distribution des catégories grammaticales principales dans le corpus. Pour l’illustration graphique des relations multilatérales, nous avons une fois de plus recours à l’analyse factorielle :

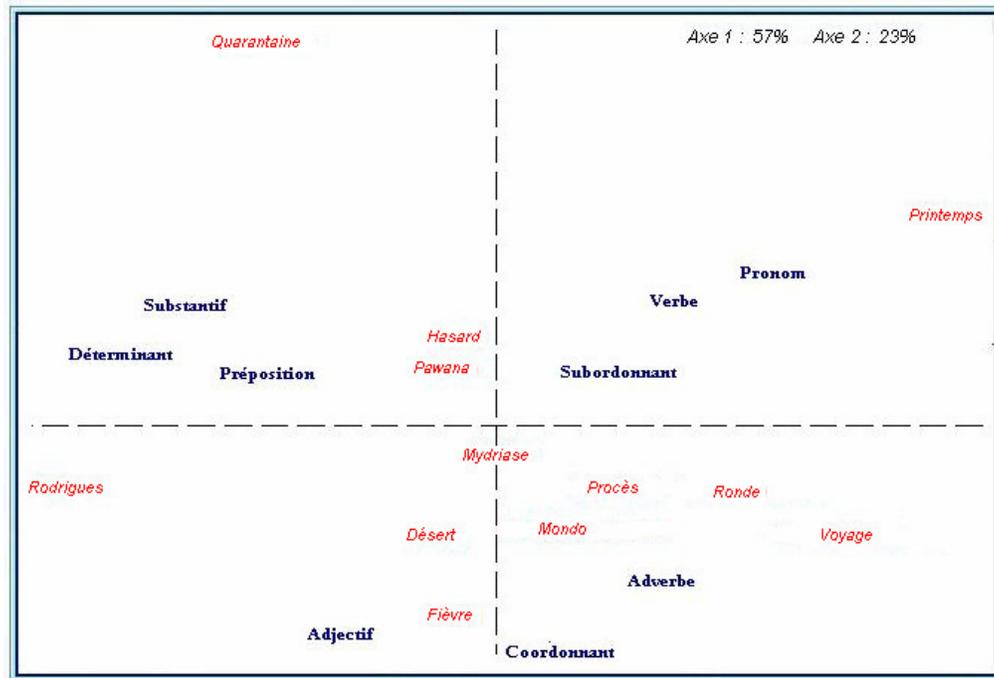


Figure n°80 : Analyse factorielle de la distribution grammaticale selon la lemmatisation Labbé (corpus I).

Afin de connaître le rôle que pourraient jouer les différentes techniques de lemmatisation sur les résultats d'une analyse de ce caractère, nous comparons le graphique ci-dessus avec celui de l'analyse factorielle de la distribution des catégories grammaticales du même corpus mais traité avec une autre méthode de lemmatisation, Cordial 7, que nous avons déjà exploitée dans ce chapitre.

L'analyse factorielle de la liste de fréquences de ces catégories permet de voir la distribution des catégories grammaticales dans le corpus :

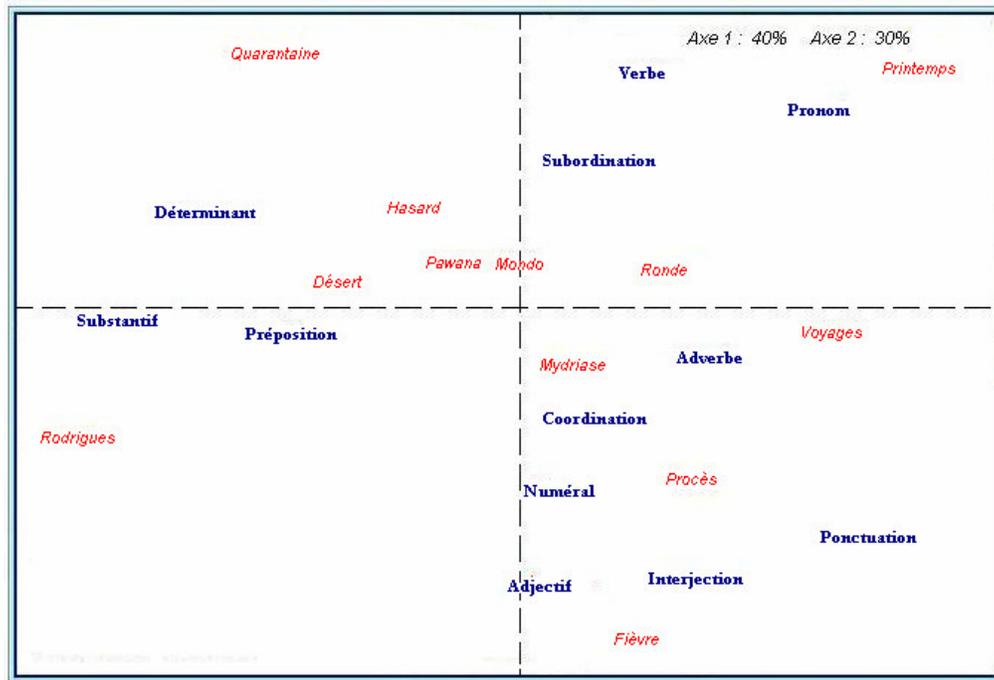


Figure n°81 : Analyse factorielle de la distribution grammaticale selon la lemmatisation de Cordial 7 (corpus G).

Bien que les critères soient un peu différents³⁵³, les deux tableaux sont très semblables : nous constatons la même opposition entre le groupe du verbe et celui du substantif avec les autres catégories qui gravitent entre ces deux pôles, ainsi que l'évolution dans le temps de l'auteur.

Ces deux projections permettent également de constater le regroupement selon les genres littéraires. Les ouvrages "nouveau roman" se regroupent en bas à droite avec les adjectifs, les interjections, les numéraux et les signes de ponctuations. L'absence de ces trois dernières catégories dans l'étiquetage Labbé modifie cependant sensiblement le positionnement de plusieurs œuvres le long de l'axe 2. Le substantif sollicite les romans plus traditionnels tandis que le verbe attire les recueils de nouvelles. Le seul ouvrage sous forme de journal, *Voyage à Rodrigues*, se trouve isolé des autres.

³⁵³ La division en catégories des deux lemmatiseurs est légèrement différente.

Examinons pour l'instant le phénomène dont témoigne le premier facteur de l'analyse factorielle, l'opposition du groupe verbal et du groupe nominal, et revenons à notre corpus principal, celui de 31 livres.

4.4. L'opposition des catégories grammaticales.

On observe souvent dans un corpus clos, comme nous venons de le faire, que deux camps, la catégorie nominale et la catégorie verbale, s'affrontent : la classe du verbe et les catégories qui lui sont proches (subordonnants, relatifs, pronoms et adverbes) s'opposent à la classe nominale qui réunit autour du substantif les adjectifs, les déterminants, les prépositions et souvent les coordinations.

Le graphique ci-dessous illustre comment les substantifs et les verbes varient les uns par rapport aux autres dans les différents livres du corpus. Les courbes s'appuient sur les écarts réduits³⁵⁴ des deux catégories grammaticales³⁵⁵.

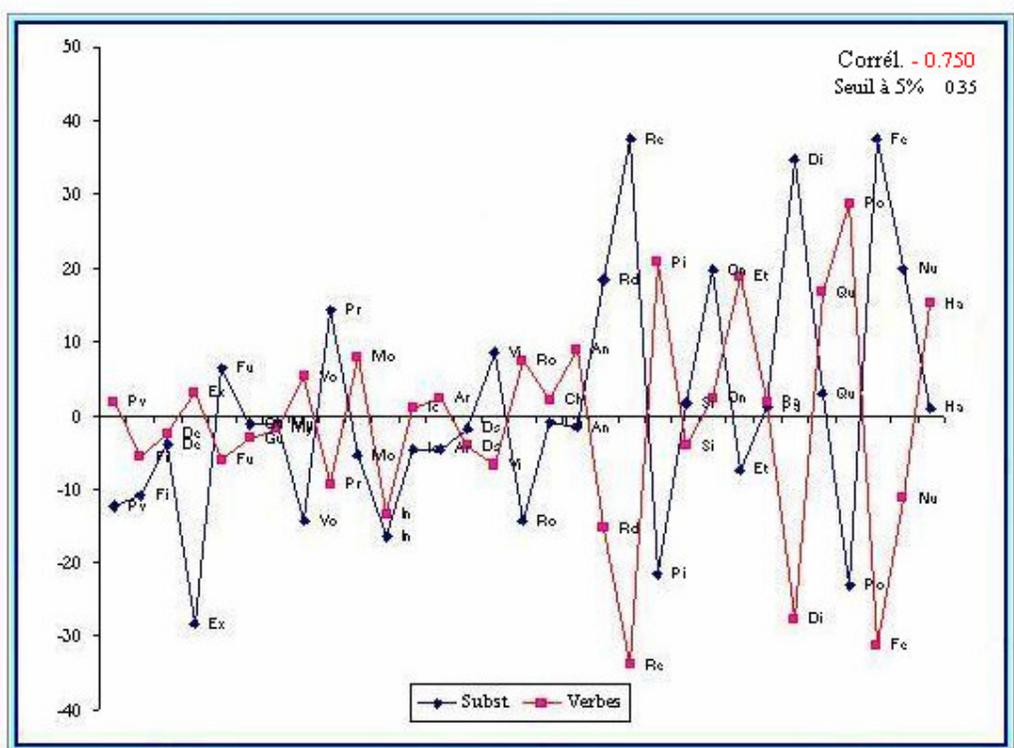


Figure n°82 : Comparaison de la distribution interne des substantifs et des verbes (écarts réduits) dans le corpus B.

³⁵⁴ Pour le calcul de l'écart réduit, voir chapitre 1.4.1. "Les principes statistiques", p. 67.

³⁵⁵ Les écarts réduits sont reportés sur le graphique n°78 dont l'axe horizontal est formé par la moyenne sur tout le corpus, du moins comprise dans un intervalle de fluctuation "normale" de ± 2 écarts autour de l'axe. S'ils n'avaient de fluctuation stylistique significative (du point de vue probabiliste), tous les points seraient donc confondus avec l'axe horizontal.

Le graphique rend compte de l'opposition des deux catégories, qui semble s'accroître, avec des écarts de plus en plus importants et des mouvements de plus en plus amples, au fur et à mesure que l'œuvre progresse. Ces courbes permettent en effet de voir de près ce que l'analyse factorielle nous a déjà indiqué : lorsqu'une œuvre est riche en substantifs, elle est pauvre en verbes et vice-versa. Le coefficient de corrélation significativement négatif (-0,75) nous apporte la preuve formelle de ce phénomène que souligne encore, ci-dessous, l'histogramme du quotient entre les 459.957 substantifs et les 321108 verbes, qui se révèle très sensible au genre³⁵⁶ :

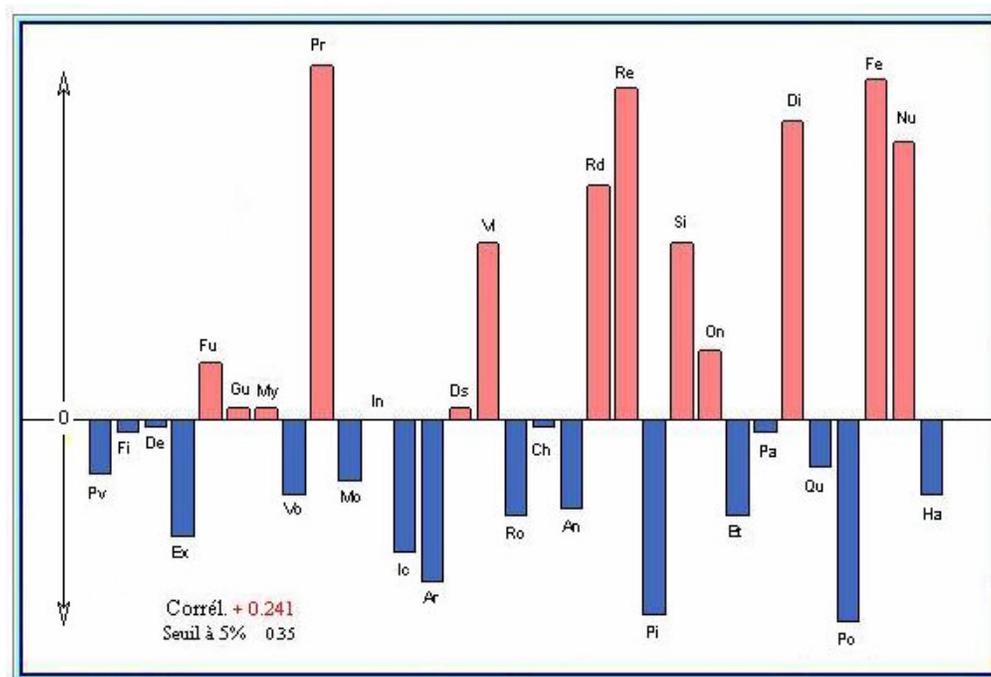


Figure n°83 : Histogramme du quotient substantifs/verbes (corpus B).

³⁵⁶ Le quotient est le rapport entre les deux séries. Il permet de voir comment se séparent les parallèles quand deux séries sont liées et parallèles. Comme les deux séries peuvent avoir un poids très inégal, la seconde est d'abord ramenée à la dimension de la première, proportionnellement, pour que le total des deux séries soit le même. Le quotient est calculé ensuite terme à terme, et s'équilibre nécessairement autour de la valeur 1.

Toutefois, dans le graphique n°79 nous constatons qu'au début de la production de l'écrivain, dans sa période "nouveau roman", les deux courbes ne s'écartent point, elles se suivent au contraire, les deux catégories étant déficitaires dans cette partie de l'œuvre. C'est à partir de l'essai *L'extase matérielle* que l'opposition se déclare. Les deux courbes deviennent parallèles dans *La guerre* et dans *Mydriase* pour se séparer de nouveau à partir de *Voyages de l'autre côté*. Dans les romans et dans les recueils de nouvelles qui paraissent entre 1975 et 1986 l'opposition des deux catégories est observable sans être très importante. Les écarts les plus importants - avec un déficit important de verbes et un grand excédent de substantifs - sont à trouver dans les ouvrages d'ethnologie et dans les essais qui traitent du nouveau monde, comme *Le rêve mexicain* ainsi que dans la biographie *Diego et Frida*. *Poisson d'or* est le seul roman de cette époque qui présente un écart d'une grande amplitude, mais l'écart cette fois-ci témoigne d'un déficit important de substantifs et d'un excès de verbes. Ces deux dernières remarques confirment donc les observations de Pierre Guiraud³⁵⁷.

Dans les œuvres non fictionnelles – les ouvrages ethnologiques, les essais, les récits de voyage et la biographie – l'évolution de l'opposition entre la catégorie du substantif et celle des verbes est en effet assez spectaculaire. Au début, les substantifs sont déficitaires et les verbes excédentaires, mais assez vite les rôles s'inversent et l'écart s'amplifie de façon importante. Il est difficile de fournir une explication précise, mais à un moment qui correspond à la découverte de la culture amérindienne et mexicaine, capitale pour notre écrivain, les substantifs commencent à abonder tandis que les verbes diminuent de façon considérable. Cette découverte capitale, Le Clézio veut en témoigner et il répète souvent : "Être vivant c'est savoir regarder". Peut-être, à partir de ce moment, n'y a-t-il plus besoin du mouvement, des dialogues et de verbes (d'action ou de parole), il suffit de regarder et Le Clézio observe, décrit et partage ce qu'il voit avec ses lecteurs en recourant à de nombreux substantifs.

³⁵⁷ Cf. p. 297.

Comme nous venons de le rappeler, la bipolarité que nous pouvons observer des catégories des substantifs et des verbes chez Le Clézio n'a rien d'originale : elle a été observée dans bien d'autres corpus : Étienne Brunet³⁵⁸ l'a bien remarquée dans ses diverses études et il souligne également le rôle important de l'opposition des genres littéraires. De ce point de vue l'œuvre de Le Clézio s'inscrit tout à fait dans la dynamique générale de la littérature française.

Le graphique ci-dessous³⁵⁹, qui isole l'œuvre romanesque, illustre encore plus nettement cette opposition qui s'amplifie vers la fin de l'œuvre, à l'exception du roman *La quarantaine* où les courbes tendent à converger :

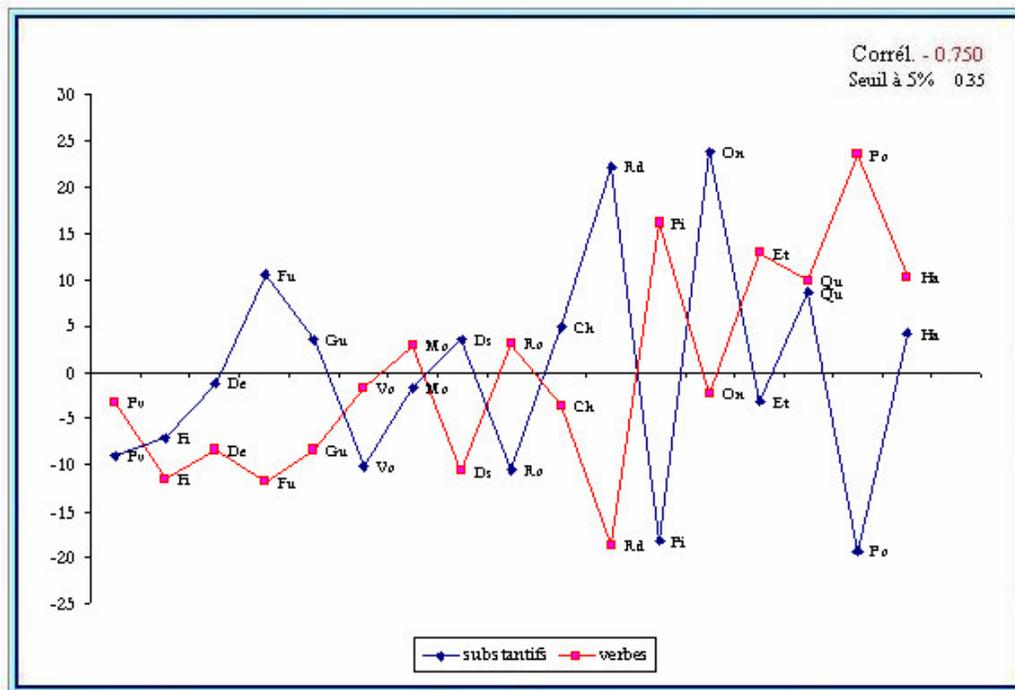


Figure n°84 : Comparaison de la distribution interne des substantifs et des verbes (écarts réduits) dans le corpus E.

Toutefois, le verbe et le substantif ne sont pas seuls dans les constellations opposées que nous avons vues, et il convient également de prendre en

³⁵⁸ É. Brunet (1985) : p. 155.

³⁵⁹ Pour l'interprétation de ce graphique, cf. p.304.

considération les éléments qui les entourent : les déterminants et les prépositions d'un côté et de l'autre les pronoms et les conjonctions. Nous nous proposons donc de regarder de plus près les diverses classes grammaticales constitutives des syntagmes nominaux et verbaux, en commençant par celle de la catégorie nominale.

4.5. Le syntagme nominal.

“Les mots n’étaient pas encore attachés entre eux, c’étaient de purs concepts, ils étaient libres, ils venaient en foule, selon une allure chaotique semblable à celle du rythme de la vie et de la matière ; les phrases n’avaient presque pas de structures grammaticales.”

La Fièvre, page 143.

Nous avons déjà pu constater l’ampleur de la catégorie nominale dans notre corpus. Rappelons que ce phénomène est constaté dans pratiquement toutes les études statistiques sur le vocabulaire français. Selon Marc Hug³⁶⁰ :

“Le syntagme nominal est l’unité qui s’impose d’emblée quand on commence l’étude statistique : non seulement c’est l’unité la plus abondamment représentée dans tout énoncé, et une unité très multiforme, mais c’est en même temps une unité fragmentaire, qui peut être étudiée indépendamment de la structure d’ensemble de la phrase...”

La catégorie nominale est également la catégorie qui domine la plupart des études sur le contenu lexical. Nous y reviendrons plus tard dans ce travail, pour l’instant nous nous intéressons aux caractéristiques quantitatives et formelles.

Nous nous pencherons dans ce chapitre premièrement sur son principal composant, le substantif, pour ensuite étudier son déterminant ainsi que le pronom et l’adjectif.

³⁶⁰ M. Hug (1989) : p. 5.

4.5.1. Le substantif.

Le substantif est la catégorie grammaticale la plus importante (en nombre) de notre corpus³⁶¹. L'histogramme ci-dessous rend compte de sa distribution à travers l'œuvre³⁶² :

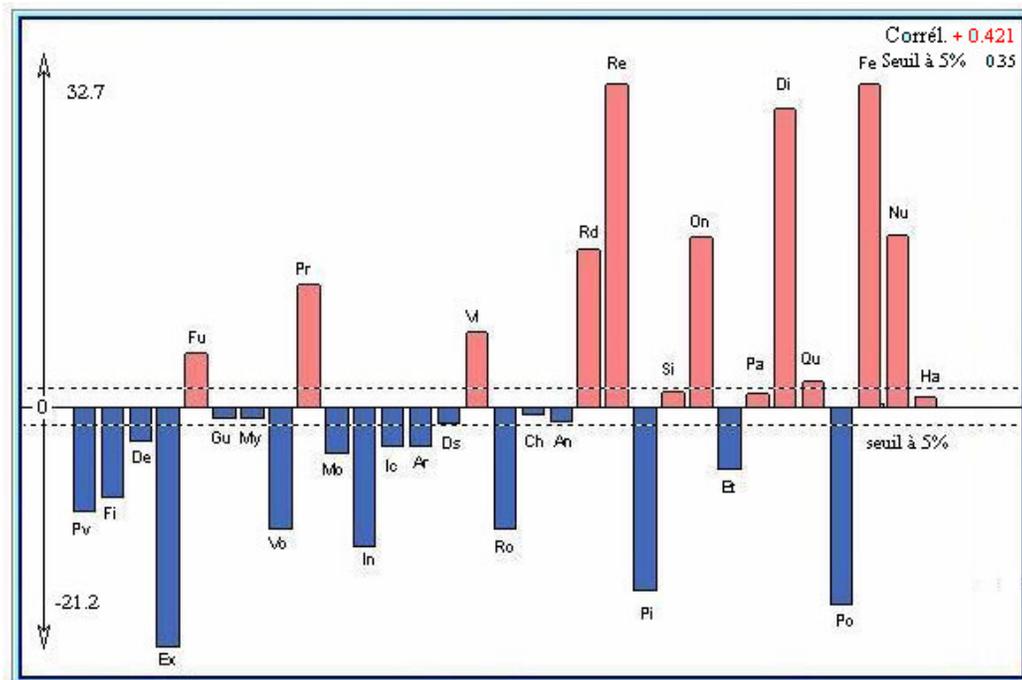


Figure n°85 : La distribution des substantifs dans le corpus B (écarts réduits).

Nous pouvons observer que la fréquence des substantifs à l'intérieur du corpus n'est pas constante. Le graphique rend compte de deux dynamiques différentes : la disparité générique et l'évolution dans le temps. Concernant l'opposition des genres littéraires, la tendance générale est celle d'un substantif déficitaire dans les romans tandis qu'il est excédentaire dans les ouvrages d'ethnologie, dans la biographie et dans les récits de voyages. Les valeurs positives extrêmes sont à trouver dans *Le rêve mexicain*, *La fête chantée* et dans *Diego et Frida*. Nous rappelons

³⁶¹ Cf. tableau n°33, p. 295-296.

³⁶² Pour l'interprétation du graphique, cf. p. 121.

que ce sont dans ces mêmes ouvrages que nous avons pu relever les phrases les plus longues³⁶³.

Les essais littéraires sont divisés en deux groupes. Les premiers, *L'extase matérielle* et *L'inconnu sur la terre*, dans lesquels Le Clézio dénonce la culture moderne occidentale, sont déficitaires en substantifs. A partir des années 1980 les essais sont nettement excédentaires en substantifs. Cela semble correspondre, comme nous l'avons dit, à la découverte de Le Clézio d'un autre monde et d'une autre culture, celle des Amérindiens, qui l'émerveille et qu'il décrit dans ces ouvrages.

Ces résultats sont confirmés par les constats qu'a faits Étienne Brunet concernant le vocabulaire français de 1789 à nos jours³⁶⁴ concernant "le style substantif". Le langage technique et le discours intellectuel (que l'on trouve dans les essais et dans les ouvrages ethnologiques de Le Clézio) utilisent plus volontiers les substantifs car ils sont plus adaptés aux définitions abstraites et "s'enchaînent avec une grande rigueur logique dans des phrases reflétant sous une forme très ramassée une pensée complexe."

Éliminons pour un instant l'effet de l'opposition générique et écartons les genres littéraires divers, pour nous concentrer sur l'écriture romanesque. Le graphique ci-dessous permet l'observation de la distribution relative des substantifs à l'intérieur du corpus E :

³⁶³ Cf. chapitre 3.4. "La longueur de la phrase".

³⁶⁴ É. Brunet (1981) : p. 304-304.

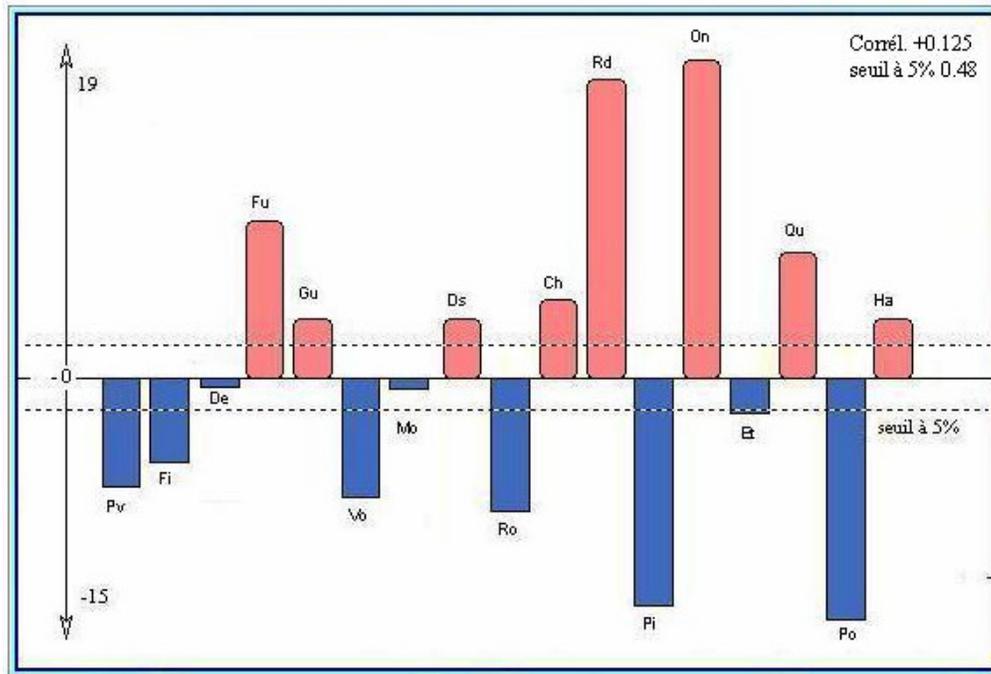


Figure n°86 : La distribution du substantif dans le corpus E (écarts réduits).

L'histogramme permet de constater de grandes variations à l'intérieur du genre. Après les déficits dans *Le procès-verbal* et dans *La fièvre*, la fréquence de substantifs “grimpe” et devient nettement positive dans *Le livre des fuites*. Dans ce roman le personnage principal entreprend “une déambulation à travers lui-même et à travers l'univers des villes monstrueuses, des autoroutes et des déserts et des grouillantes populations mourant de misère sur des sols pourris”³⁶⁵. La vocation de ce livre dépasse en fait celle du roman de fiction et Le Clézio critique le mythe moderne où l'image est transmise souvent par des cascades de substantifs dans de longues énumérations, se rapprochant plus d'un “discours intellectuel”.

Les romans qui ont pour thème les voyages et les découvertes sont également riches en substantifs. La découverte d'un nouvel univers, coloré d'exotisme et de pittoresque, donne lieu à une riche utilisation de noms, qu'il s'agisse de l'Océan Indien dans *Voyage à Rodrigues*, *Le Chercheur d'or* et dans *La quarantaine* ou bien de l'Afrique noire dans *Onitsba*. La traversée de l'Atlantique dans *Hasard* et la

³⁶⁵ La présentation sur la quatrième de couverture.

découverte du désert dans *Désert* semblent, à un moindre degré, également être favorables au substantif. On touche du doigt, à travers l'examen des simples catégories grammaticales, la proximité de ces œuvres de fiction romanesque avec certains aspects des ouvrages d'ethnologie.

En revanche, dans les livres qui favorisent le dialogue, et par conséquent une langue plus orale, le substantif est déficitaire. Des livres comme *Printemps et autres saisons* et *Poisson d'or* ne sont pas seulement riches en dialogues mais sont également des livres où l'action prime sur la description. Dans *Poisson d'or*, il faut aussi souligner l'effet de la technique du récit à la première personne qui “nous raconte son histoire”³⁶⁶ :

“Ce n'était pas très gentil de ma part, j'en suis d'accord avec vous, mais je ne pouvais plus. A l'heure où je devais rentrer chez nous, rue Jean-Bouton, je suis restée chez Madame. [...] Madame est rentrée un peu plus tôt. Quand elle m'a vue, elle n'a rien dit, elle a tout compris. Elle m'a embrassée, avant même d'ôter son imper et de lâcher ses clefs. Elle a dit : ça me fait bien plaisir, ma chérie, j'attendais ce jour-là, j'étais sûre qu'il arriverait.”

Il est intéressant de noter, comme nous venons de le faire, que la figure qui illustre la longueur de la phrase est pratiquement superposable à celle de la distribution des substantifs. Cela confirme ce que nous avons pu constater auparavant, c'est-à-dire que la construction de la phrase longue chez Le Clézio repose souvent sur l'énumération et l'accumulation, et ce sont, bien évidemment, des substantifs qui prolifèrent dans ce système. Nous notons également chez Le Clézio la même corrélation entre la fréquence du substantif et l'accroissement lexical ainsi que la richesse du vocabulaire qu'a déjà trouvée Charles Muller³⁶⁷ dans ses études sur le vocabulaire de Corneille, où les pièces les plus riches en substantifs avaient en général aussi un vocabulaire étendu, bien qu'il y eût des exceptions à cette règle.

³⁶⁶ *Poisson d'or*, p. 111.

³⁶⁷ Ch. Muller (1967) : p. 111-115.

Les études antérieures ont également montré que souvent chez les écrivains la fréquence du substantif diminue avec le temps ; c'est le cas de Corneille, de Hugo et bien d'autres auteurs.

“On peut affirmer, écrit Étienne Brunet à propos de la dominance nominale³⁶⁸, que l'usage littéraire de la langue résiste à l'invasion du substantif et que depuis deux siècles les écrivains tendent à lui substituer le verbe. On en arrive à un curieux paradoxe : le verbe qui est caractéristique du langage parlé trouve refuge dans l'écriture littéraire, alors que l'abus du style substantif réputé noble, s'exerce dans l'information quotidienne que nous délivre la presse, l'administration, la publicité et les médias.”

Le Clézio, en revanche, ne s'inscrit pas dans cette tendance générale, au contraire, chez lui l'utilisation du substantif augmente au fur et à mesure que l'œuvre progresse. Les valeurs positives dans *La guerre* et *Le livre des fuites* avec ses nombreux patchworks de coupures de journaux etc. pourraient éventuellement témoigner de la tendance décrite par Étienne Brunet, mais l'augmentation progressive du substantif nous semble provenir du fait que la description prime plutôt que l'action vers la fin de l'œuvre.

Toutefois, il convient dans ces interprétations de rappeler ce qu'écrit Britt-Marie Kylander³⁶⁹ :

“Il faut éviter de toujours interpréter de la même manière les textes qui sont dominés par une catégorie grammaticale particulière. Il est très important de savoir quels types de substantifs ou de verbes dominant dans un texte étudié.”

³⁶⁸ É. Brunet (1981) : p. 304.

³⁶⁹ B.-M. Kylander (1995) : p. 93.

La division de la catégorie de substantifs en genres, c'est-à-dire en masculin et en féminin, pourrait nous apporter plus de précisions sur l'usage du nom dans l'œuvre de Le Clézio.

Dans les études statistiques s'appuyant sur des corpus non lemmatisés, l'analyse du genre masculin et féminin se faisait surtout à partir de l'étude des articles, c'est-à-dire *la*, *le*, ou *une*, les formes *l'* et *les* ne disant rien sur le genre. Or ces formes sont ambiguës et des confusions avec les numéraux aussi bien qu'avec des pronoms personnels³⁷⁰ rendaient la statistique moins fiable. Les formes contractées *au* et *du* étaient délicates à traiter et le phénomène de la neutralisation de l'opposition du genre masculin-féminin dans les formes *l'*, *les*, *aux* et *des* ne facilitait certes pas la tâche.

Désormais, le logiciel Hyperbase dans sa version lemmatisée permet d'extraire les noms féminins et les noms masculins distinctement grâce à l'analyseur incorporé. Si nous divisons la catégorie des substantifs en genres, nous pouvons constater que parmi les 462632 substantifs le masculin domine avec 56,5% des effectifs, le féminin n'en comportant que 43,5%.

Les courbes du graphique n°82 permettent d'observer leur distribution respective à travers le corpus :

³⁷⁰ É. Brunet (1988) : p. 192 a relevé que les formes attribuées au pronom personnel représentaient 14% de celles qui revenaient aux articles mais que la part du numéral était faible face à l'article indéfini.

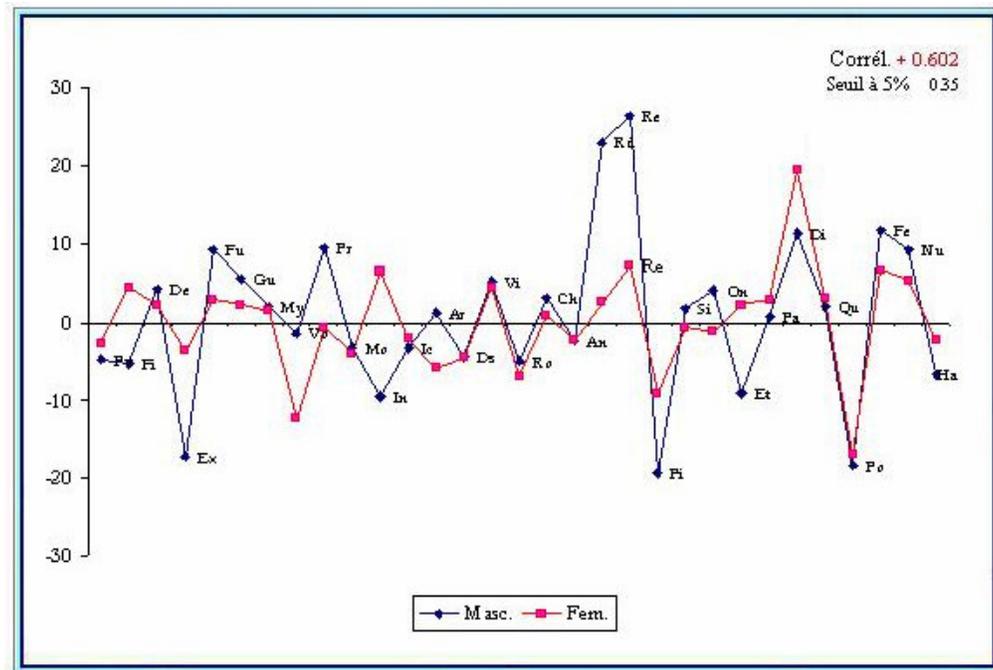


Figure n°87 : La relation entre le genre masculin et le genre féminin dans le corpus B (écarts réduits).

Dans *Voyage à Rodrigues* et dans *Le rêve mexicain* le genre masculin est très important. La biographie *Diego et Frida* est le seul livre à contenir des valeurs très élevées du côté féminin. Nous pouvons constater que la distribution des deux genres est relativement parallèle et qu'il n'y a pas d'opposition franche entre le masculin et le féminin, le coefficient de corrélation significativement positif (+0.60) en témoigne. C'est uniquement dans *L'inconnu sur la terre* et dans *Voyage au pays des arbres* que les courbes s'opposent de façon significative autour de l'axe des abscisses, comme le montre dans la figure 88, la courbe sur le quotient.

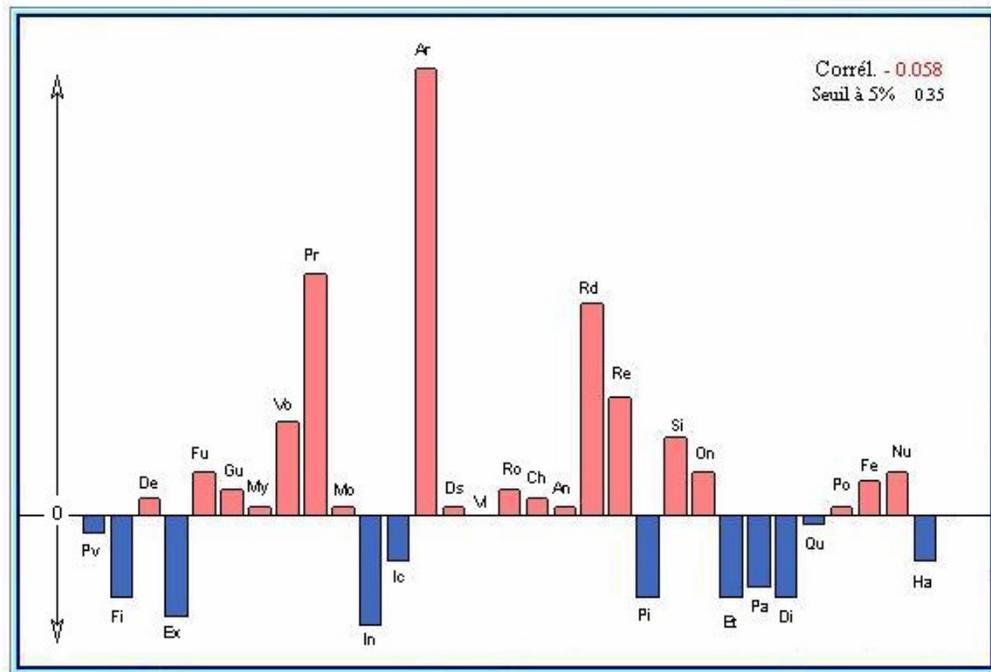


Figure n°88 : Courbe sur le quotient masculin/féminin (corpus B).

Cette courbe témoigne peut-être moins de l'opposition des genres littéraires que des différentes périodes dans l'écriture leclézienne. Après une première période avec peu d'opposition entre le masculin et le féminin, nous constatons de fortes oppositions dans la deuxième période, à l'exception de l'essai *L'inconnu sur la terre* et du récit poétique *Vers les icebergs*. L'opposition diminue par la suite, étant au plus bas dans *Etoile errante*, *Pawana* et *Diego et Frida*, pour augmenter légèrement vers la fin du corpus.

Dans l'analyse statistique du substantif, il convient également de prendre en compte les résultats de l'étude des articles qui contribuent en français à la distinction des deux genres.

4.5.2. Les articles.

A la comparaison avec les valeurs obtenues lors de l'analyse des deux articles *le* et *la*³⁷¹, nous pouvons constater que parmi les 104.009 occurrences l'article *le* emporte sur l'article *la* avec 57% contre 43%, c'est-à-dire un résultat quasi identique à celui du masculin et du féminin. La distribution interne toutefois diffère quelque peu :

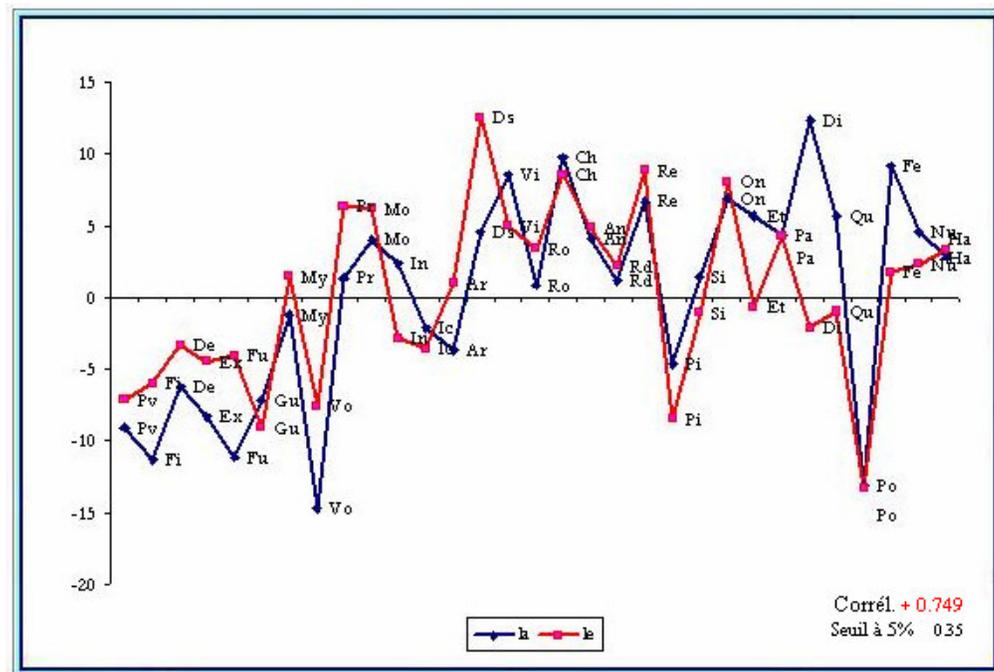


Figure n°89 : La distribution relative des articles *la* et *le* dans le corpus B (écarts réduits).

Le mouvement général, en effet, ne se distingue que légèrement des courbes précédentes, les variations sont un peu plus importantes et le début de l'œuvre manifeste un déficit dans les deux camps.

Nous pouvons constater que ces courbes font preuve d'un parallélisme encore plus fidèle que celles du masculin et du féminin, avec un coefficient de corrélation

³⁷¹ La forme unique du pluriel est exclue de l'analyse, ne permettant pas la distinction du genre en français.

de +0,749 ; c'est uniquement dans *Diego et Frida* que les articles s'opposent, l'article féminin étant excédentaire et l'article masculin déficitaire.

Dans son étude sur le vocabulaire de Hugo³⁷² Étienne Brunet a pu observer que l'article féminin l'emportait sur le masculin avec une proportion de 53,8% contre 46,2%. Ces résultats équivalent ceux obtenus par Gunnel Engwall dans son analyse du roman contemporain³⁷³ : 53,5% et 46,5%. La préférence pour l'article féminin est encore plus marquée dans *Frantext* avec 56% contre 44% pour le masculin. La prépondérance de l'article *le* par rapport à *la* semble vraiment être une caractéristique de notre écrivain.

Comment interpréter ces résultats ? Nous savons qu'en français l'opposition des genres n'a pas de motivation précise. Toutefois dans certains substantifs on peut observer une régularité en genre due à la présence de suffixes. Les suffixes *-tion*, *-té* ainsi que *-ade* marquant une action imposent le féminin, tandis que *-ment* et *-isme* exigent l'article masculin. Étienne Brunet écrit ceci à propos de la prépondérance habituelle de l'article féminin³⁷⁴ :

“L'abstraction intéresse plus l'article défini que l'indéfini (on dit moins souvent une justice que la justice) et qui concerne davantage le féminin que le masculin. [...] Le rapport *la/le* serait donc une mesure indirecte du degré d'abstraction d'un texte ou d'un corpus.”

Les valeurs excédentaires du suffixe *-tion* dans notre corpus se trouvent dans les ouvrages ethnologiques et dans les essais où effectivement plus de place semble être donnée à l'abstraction et au discours intellectuel :

³⁷² É. Brunet (1988) : p. 193.

³⁷³ G. Engwall (1982) Rappelons qu'il s'agit dans son étude de formes graphiques et qu'il faut en tenir compte lors de la comparaison.

³⁷⁴ É. Brunet (1988) : p. 193.

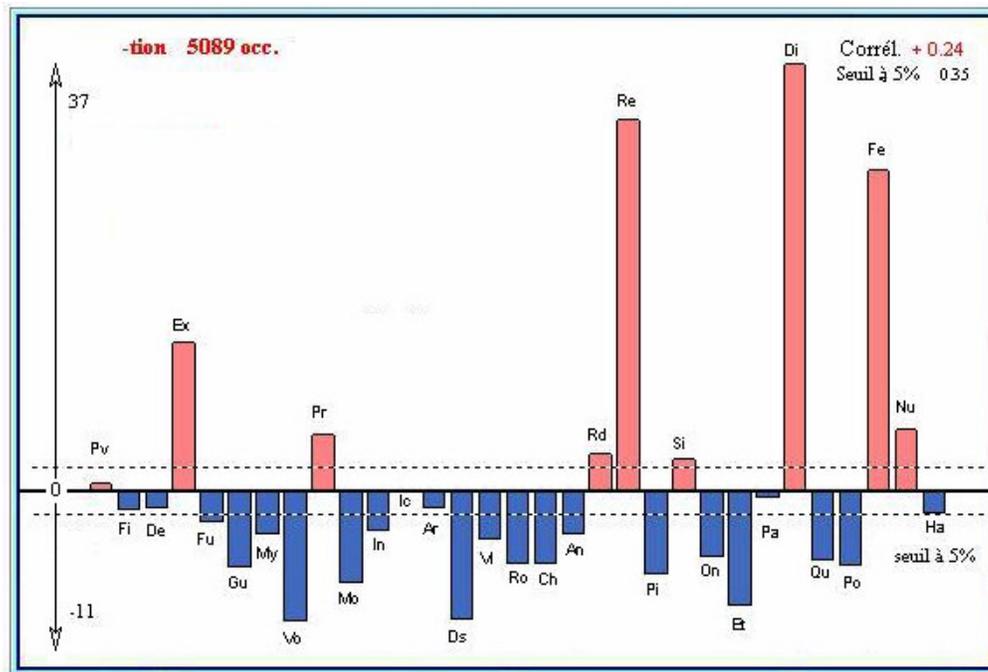


Figure n°90 : La distribution relative du suffixe -tion dans le corpus B (écarts réduits).

Dire à partir de ces analyses que Le Clézio fuit l'abstraction dans son œuvre romanesque serait sans doute une surinterprétation. Mais il y a chez notre auteur une volonté manifeste de rester dans le concret et de garder un langage simple et "terre à terre", il y revient souvent dans les divers entretiens et même dans ses textes. Toutefois, l'histogramme de l'article *la* seul témoigne d'une tendance croissante vers l'abstraction, avec un coefficient de corrélation significativement positif :

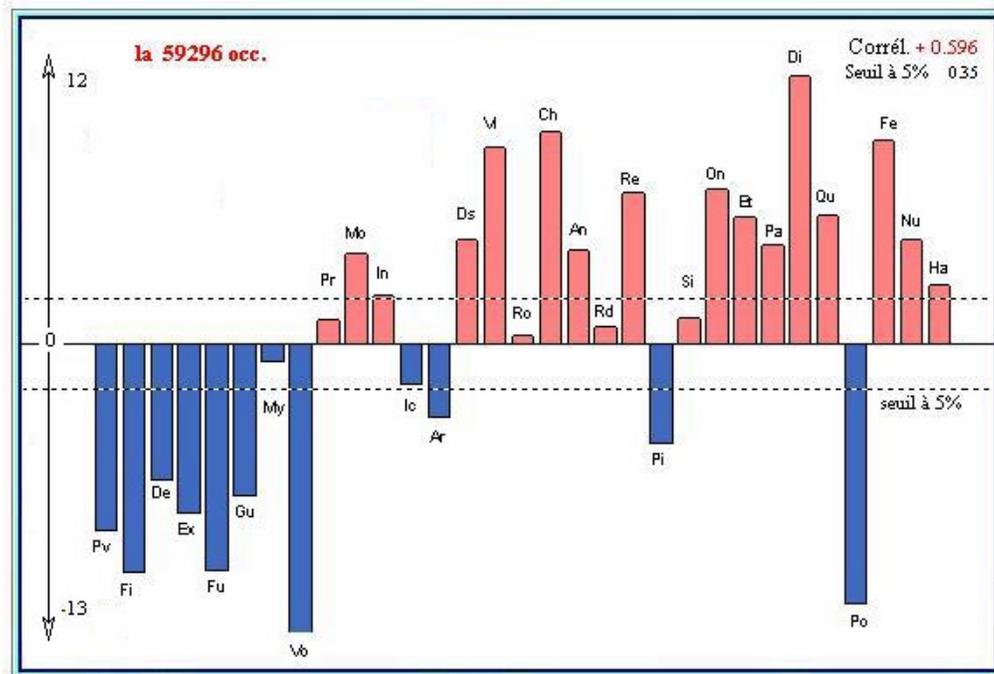


Figure n°91 : La distribution relative de l'article la dans le corpus B (écarts réduits).

Que la première période “nouveau roman” soit largement déficitaire n’est pas étonnant avec cette écriture particulière qui ne favorise pas l’abstraction. Bien au contraire, elle est favorable au concret, car c’est une réalité bien concrète, la société de consommation avec ses attributs et ses articles, que Le Clézio dénonce durant cette période. Par la suite, l’écriture semble tendre de plus en plus vers l’abstraction dans tous les genres littéraires (accentuée dans les ouvrages ethnologiques, les essais et la biographie), à l’exception des livres riches en dialogues et en oralités comme *Voyage au pays des arbres*, *Printemps et autres saisons* et *Poisson d’or*.

Un autre critère de reconnaissance du genre en français qui mérite d’être étudié de façon isolée est l’article indéfini.

L’article indéfini s’exprime par *un* ou *une* au singulier³⁷⁵. La forme masculine *un* occupe 60% des 37774 occurrences et la forme féminine *une* 40%. Cette

³⁷⁵ La forme du pluriel *des* est exclue de cette étude car elle ne distingue pas les genres.

prédominance n'a cependant rien d'original, elle apparaît partout dans les études statistiques sur des corpus littéraires³⁷⁶. Le diagramme ci-dessous rend compte de la distribution relative à l'intérieur du corpus :

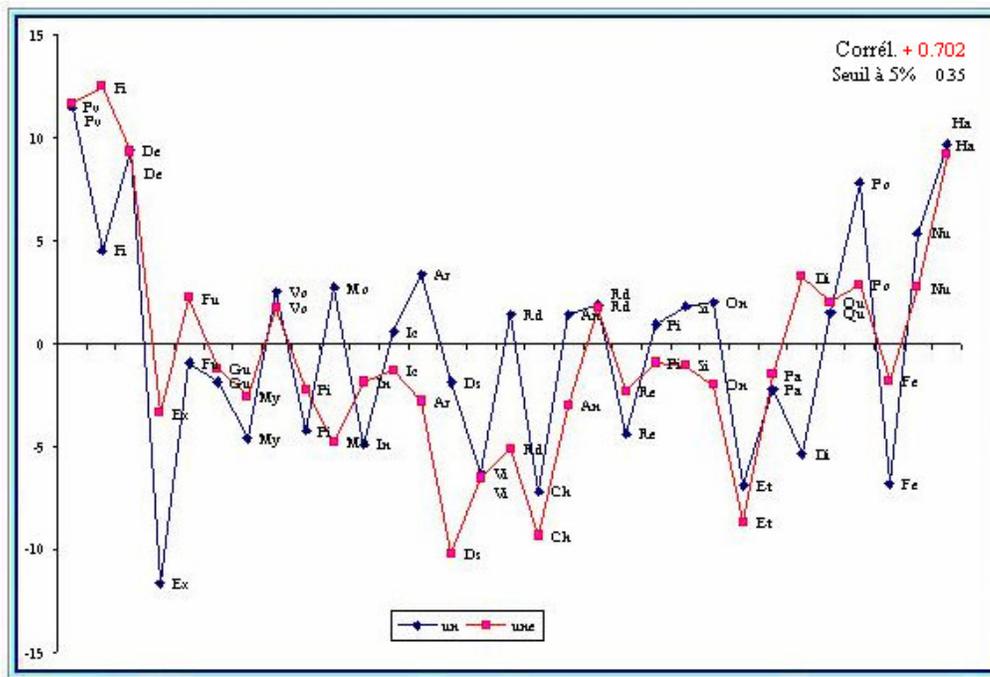


Figure n°92 : La distribution relative des articles un et une (écarts réduits).

Ces courbes témoignent, comme celles que nous venons de voir, d'un parallélisme notable, toutefois moins fidèle que pour celles que nous avons observé précédemment et le coefficient de corrélation de +0.70 en témoigne.

La comparaison de cette figure avec celle illustrant la distribution de *le* et de *la* met en lumière une autre caractéristique du corpus : celle de l'opposition entre le défini et l'indéfini :

³⁷⁶ É. Brunet l'observe chez Hugo "comme partout ailleurs" (cf. É. Brunet (1988) : p. 193.).

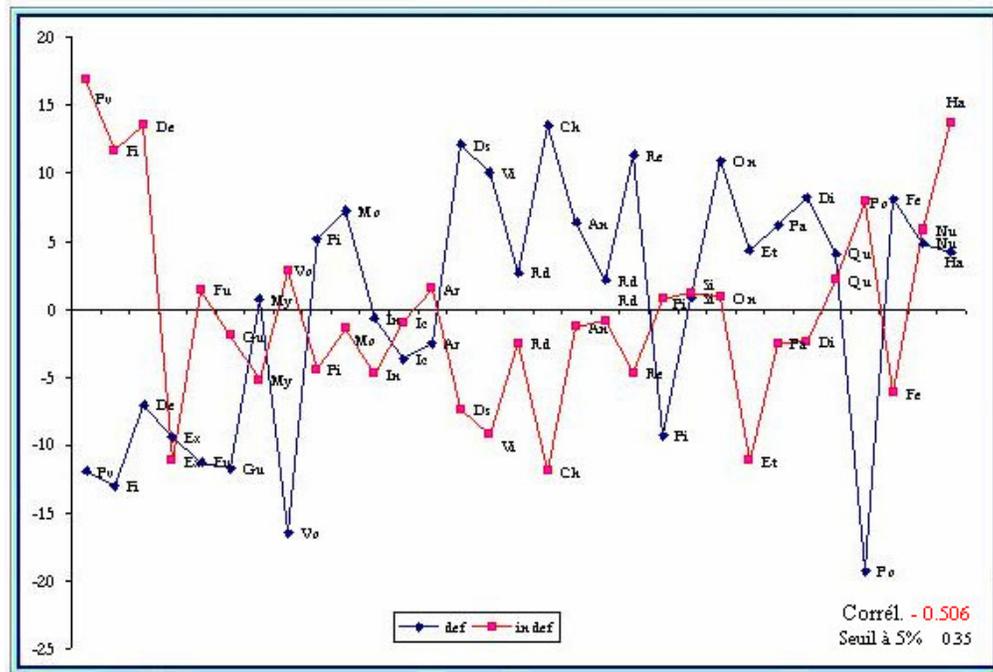


Figure n°93 : La distribution relative des articles définis et indéfinis dans le corpus B (écarts réduits).

Observons d’abord l’importance de l’article défini (effectifs absolus : 104.009 contre 37.774 pour l’article indéfini) de notre corpus. L’article défini est en effet excédentaire dans la plupart des livres du corpus tandis que l’article indéfini est déficitaire. C’est dans les premiers ouvrages ainsi que dans les derniers, surtout dans *Poisson d’or* et dans *Hasard*, que nous observons des valeurs déficitaires de l’article défini. Les ouvrages inspirés de l’école du “nouveau roman” sont, comme le sont également *Poisson d’or* et *Hasard*, riches en dialogues et en oralités qui semblent favorables selon notre analyse à l’article défini.

Les courbes permettent ensuite de constater l’opposition importante des deux paires d’articles dans quasiment tout le corpus. L’histogramme de la courbe de quotient permet de mieux cerner cette opposition :

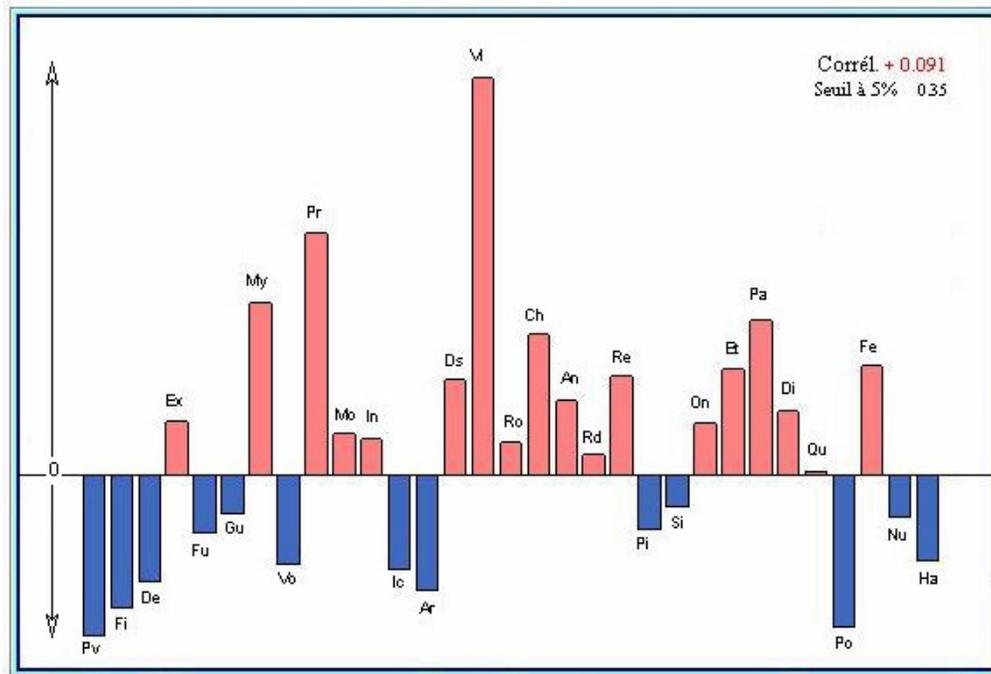


Figure n°94 : Quotient article défini/article indéfini dans le corpus B.

Les écarts sont en effet de grande ampleur, avec notamment des “pics positifs” très importants dans *Trois villes saintes*, *Les prophéties de Chilam Balam* et *Mydriase*. L’opposition des deux articles semble être plus notable dans les livres qui n’appartiennent pas au genre romanesque.

A l’intérieur du genre romanesque cette opposition est encore plus forte, comme en témoigne l’histogramme de la figure n°95 :

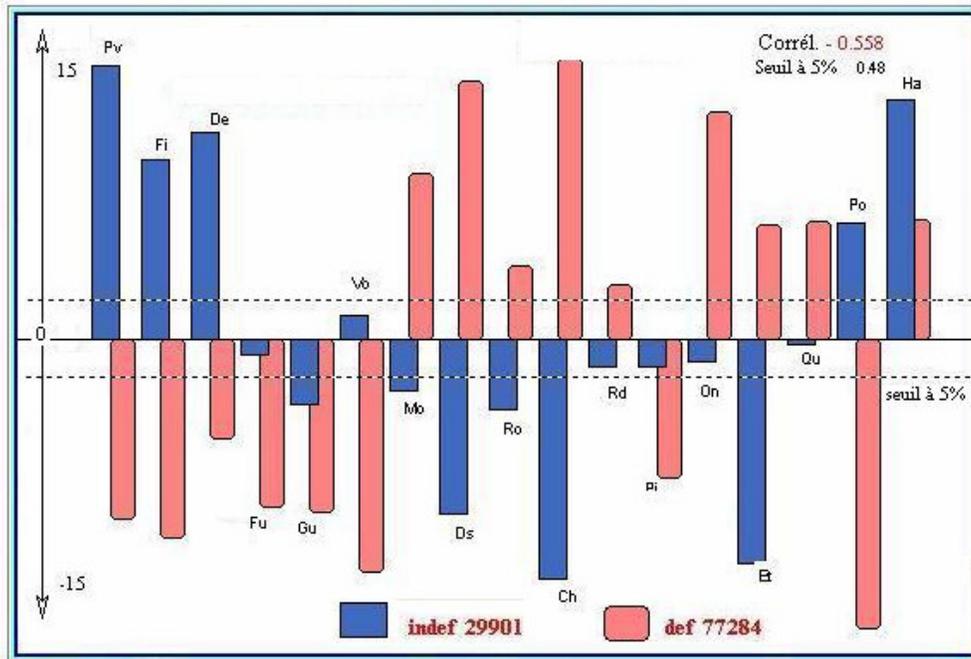


Figure n°95 : La distribution des articles définis et indéfinis dans le corpus E (écarts réduits).

Il est également intéressant de comparer l'utilisation du singulier et du pluriel dans les articles. Le diagramme ci-dessous regroupe dans une courbe les articles *les*, *des* et *aux* (84.496 occurrences) et dans l'autre les articles en singulier *la*, *le*, *une*, *un* et *au* (141.783 occurrences) :

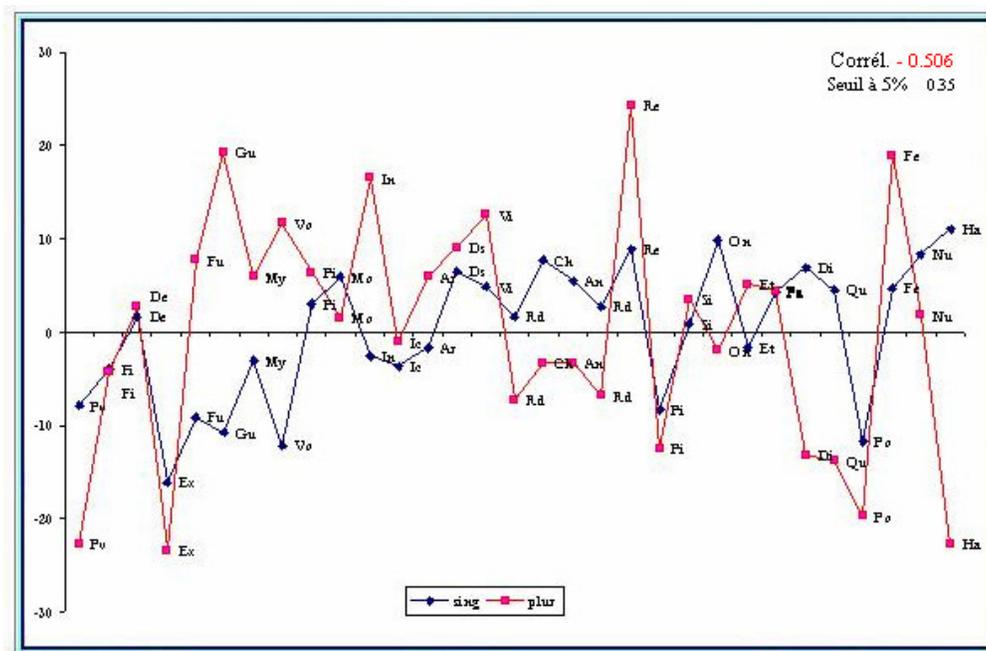


Figure n°96 : La distribution relative des articles au singulier et au pluriel du corpus B (écarts réduits).

Les courbes de ce diagramme permettent de constater un certain parallélisme bien qu'elles convergent à plusieurs reprises : elles avancent souvent des côtés opposés de l'axe des abscisses, la courbe du singulier déficitaire et la courbe du pluriel excédentaire, comme chez la plupart des écrivains. La courbe du quotient rend encore mieux compte de cette situation :

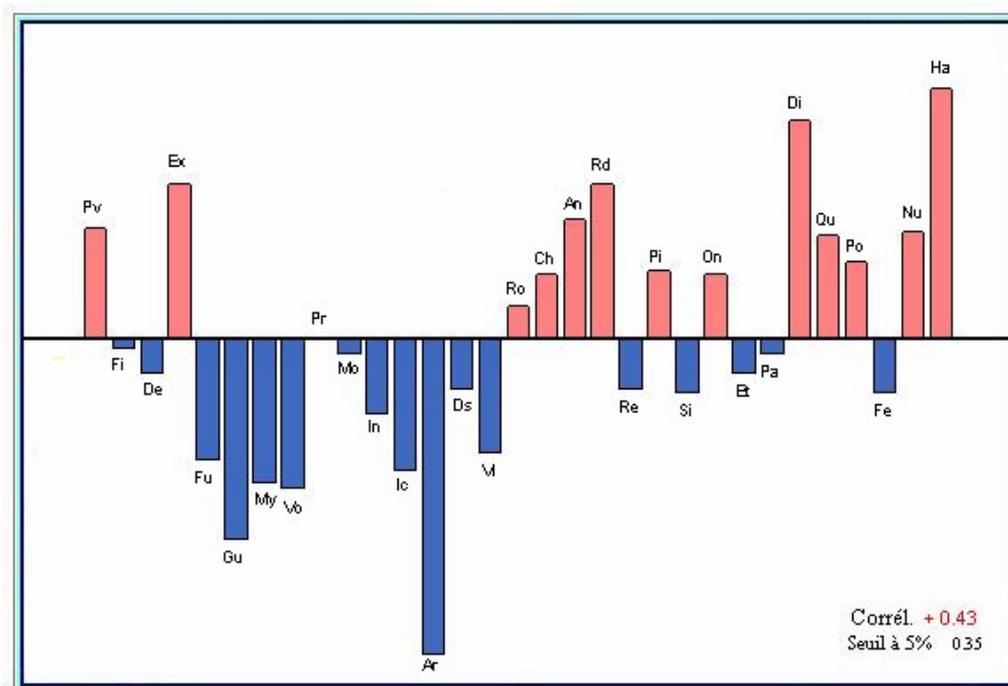


Figure n°97 : Courbe sur quotient singulier/pluriel (Corpus B).

Le coefficient de corrélation, nettement significative (+0,43), permet de constater une tendance dans l'écriture leclézienne à aller vers le singulier au fur et à mesure que l'œuvre progresse. Toutefois, dans les ouvrages ethnologiques comme *Le rêve mexicain* et dans *La fête chantée* nous constatons la situation inverse, le pluriel primant nettement sur le singulier. Cette opposition générique a déjà été observée par Étienne Brunet, notamment dans son étude du vocabulaire de Victor Hugo³⁷⁷ où il a constaté que le singulier primait dans le théâtre et dans la poésie ainsi que

³⁷⁷ É. Brunet (1988) : p. 199.

dans la correspondance qui témoigne des relations individuelles et souvent intimes, tandis que le roman avait “un cadre et un cœur plus larges où les individualités peuvent s’y accumuler et y produire du pluriel.” Chez Le Clézio il semble que la situation soit différente. Dans les ouvrages d’ethnologie il est question d’une voix plurielle, celle des peuples amérindiens, tandis que dans le roman cette petite analyse formelle souligne un phénomène caractéristique de l’écriture romanesque de Le Clézio : la solitude et l’expérience individuelle, celle de l’homme seul devant la nature.

Si nous isolons pour un instant l’œuvre romanesque, la courbe du même quotient (singulier/pluriel) rend bien compte, non seulement de la tendance de notre écrivain à favoriser le singulier, mais surtout de la tendance chronologique positive très significative avec son coefficient de corrélation de + 0,567, favorisant de plus en plus le singulier dans l’écriture³⁷⁸ :

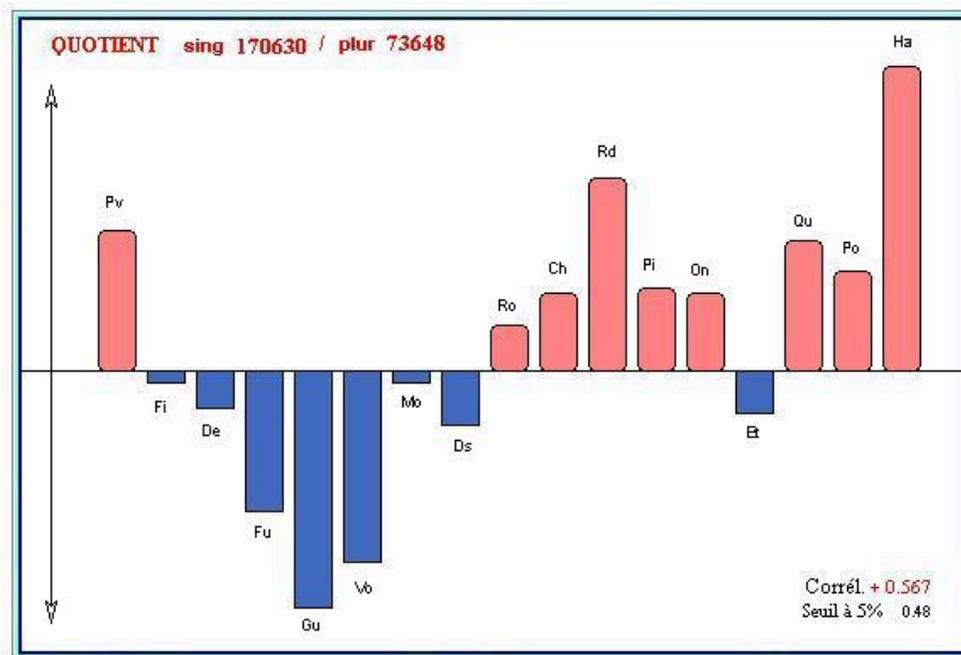


Figure n°98 : Courbe sur quotient singulier/pluriel (Corpus E).

³⁷⁸ L’accès au code grammatical du logiciel Hyperbase permet d’affiner l’analyse car grâce à l’analyse syntaxique qui désambiguïse des formes douteuses, il permet la division de la totalité des substantifs en singulier ou en masculin. Le singulier occupe 71,5% des substantifs et le pluriel 28,5%. L’analyse que nous venons de faire place 62,7% du côté du pluriel et 37,3% dans le camp du singulier. La différence qu’il y a dépend sans doute de la désambiguïsement et de la prise en compte de l’article éliidé l’.

C'est l'analyse factorielle qui nous donnera une synthèse de ces différentes relations et accords qui lient les articles et les différents ouvrages du corpus. L'analyse s'appuyant sur le corpus B prend en considération les articles *la, le, les, un, une*, l'article élide *l'* ainsi que les formes contractées *au, aux, du* et *des* :

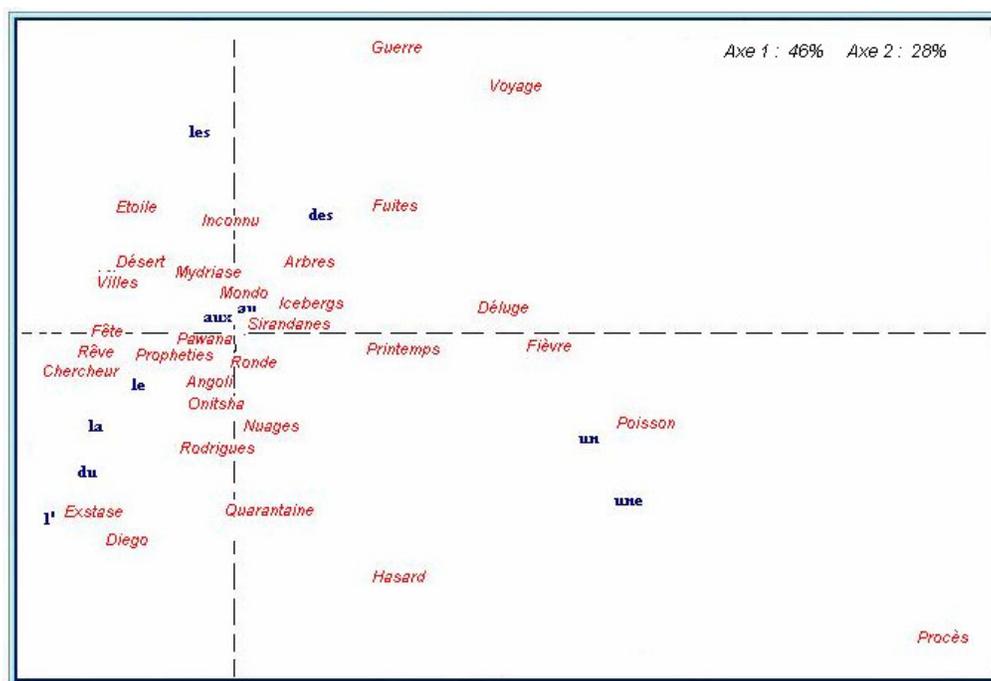


Figure n°99 : Analyse factorielle de la distribution des articles dans le corpus B.

L'analyse permet de poser plusieurs distinctions. Le premier facteur sépare l'article défini, autour duquel se trouvent les ouvrages ethnologiques, les essais, la biographie et les romans descriptifs, de l'article indéfini qui attire quant à lui les trois premiers romans du corpus ainsi que les deux derniers. Le deuxième facteur distingue le singulier du pluriel, les articles du pluriel se trouvant en haut du graphique en compagnie des romans de l'école "nouveau roman", à l'exception des trois premiers, et les autres livres proches des articles du singulier en bas du tableau. Il est à noter que l'analyse factorielle ne montre pas d'opposition entre le masculin et le féminin. Les facteurs 1-3 et 2-3 de l'analyse factorielle confirment ce constat en ne montrant aucune opposition entre les genres masculin et féminin. Il semble donc que l'opposition entre le défini et l'indéfini tout comme l'opposition

entre le singulier et le pluriel dominant nettement sur l'opposition entre le masculin et le féminin lorsque tous ces paramètres sont pris en compte.

Nous avons vu dans les différentes analyses factorielles des catégories grammaticales du chapitre précédent que le déterminant restait toujours, comme la logique le veut, très près de son maître, le substantif, auquel il témoigne d'une fidélité absolue. Une autre catégorie qui était du côté du substantif dans ces analyses était l'adjectif, étant son adjectif directement ou indirectement.