

Cinquième partie :

Le contenu lexical.

5.1. Introduction.

“Peut-être qu’il faut tout abandonner. Laisser tout cela, ces mauvais oripeaux, ces masques, ces bagues, ces ceintures qu’on avait accumulés sur soi, qu’on avait laissé suspendre. On voulait bien croire que ce n’étaient que des mots, les plus inconsistants. Si on voulait ils s’effaceraient, eux qui disaient que, qui croyaient que... Puisqu’ils jugeaient, est-ce qu’on n’allait pas les juger, un jour ? Mais les mots ne sont pas seulement des mots. Ils ont de longues tenaces racines enfoncées dans la chair, enfoncées dans le sang, et les arracher fait mal. Mots appris, reconnus, habitudes, parasites, c’étaient eux qui distillaient le venin.”

Vers Les Icebergs, page 64.

Le Clézio est marqué très tôt par la confrontation incessante des mots et des signes dans notre société moderne : “Le langage a repris son ballet dément ; les mots se traversent, s’ajoutent, se divisent. Ils zèbrent la nuit sans cesse, pareils à des fusées d’artifice.” écrit-il en 1966 dans *Le déluge*⁴⁷⁴, et il n’est pas le seul à s’intéresser à ce phénomène :

“Les différents signes qui nous entourent, comment les catégoriser et former des classes ?, demande Roland Barthes. Un vêtement, une automobile, un plat cuisiné, un geste, un film, une musique, une image publicitaire, un ameublement, un titre de journal, voilà en apparence des objets bien hétéroclites.

Que peuvent-ils avoir en commun ? Au moins ceci : ce sont tous des signes. Lorsque je me déplace dans la rue – ou dans la vie – et que je rencontre ces objets, je leur applique à tous, au besoin sans m’en rendre compte, une même activité, qui est celle d’une certaine lecture : l’homme moderne, l’homme des villes passe son temps à lire.”⁴⁷⁵

⁴⁷⁴ *Le déluge*, p. 266.

⁴⁷⁵ R. Barthes, “La cuisine du sens”, in *L’aventure sémiologique*, Paris, Points essais, Seuil, 1985, p. 227.

Il est vrai que dans notre vie quotidienne nous sommes constamment confrontés à cette richesse de signes et mots que nous essayons – avec plus ou moins de succès – de comprendre et de doter d’un sens. Lorsque l’on s’intéresse à la signification des mots et que l’on éprouve le besoin de faire une analyse un peu plus affinée que celle première que propose Roland Barthes, on est confronté à un phénomène qui semble refléter le grand amour pour les mots qu’éprouvent les sémanticiens et les sémiologues : une grande prolifération de termes et de notions qui se complètent et parfois se contredisent...

Il convient donc, avant toute étude qui porte sur le sens des mots, de rappeler quelques notions fondamentales ainsi que les définitions principales de ce carrefour qu’est, malgré tout, l’analyse statistique du contenu lexical d’un corpus littéraire, entre la lexicologie, la statistique, la sémantique, la sémiotique ainsi que les analyses purement littéraires de différents genres.

Le terme de “sens” doit tout d’abord être manié avec une extrême prudence. Lorsque Benveniste écrit⁴⁷⁶ que “le mot n’apparaît pas nécessairement dans la phrase avec le sens qu’il a comme unité autonome”, il souligne l’importance d’étudier le contexte et pas seulement l’unité linguistique individuellement, ouvrant ainsi la voie à l’étude de thèmes. “Le thème, écrit Évelyne Bourion, est considéré comme une macrostructure sémantique, qu’on postule composée de différentes structures sémantiques stables et reliées entre elles (le réseau associatif).”⁴⁷⁷ Ce réseau associatif est pourtant assez fragile, notamment à cause de sa large part de subjectivité.

“La notion de thème est imprécise, écrit François Rastier⁴⁷⁸ ; en général un thème est un lexème utilisé comme dénomination générique. Par exemple, une thèse sur le thème de l’eau chez Bosco (il se dépose bon an mal an en France des dizaines de sujets de ce genre) exploitera les

⁴⁷⁶ É. Benveniste (1966) : p. 124.

⁴⁷⁷ É. Bourion, “Le réseau associatif de la peur”, in *L’analyse thématique des données textuelles*, É. Martin, F. Rastier (éds.), Paris, 1995, p. 108.

⁴⁷⁸ F. Rastier (1989) : p. 54.

occurrences fleuve, rivière, mer, étang, etc. Bien entendu les résultats seraient différents si l'on avait choisi le thème de la rivière ou, pourquoi pas, celui du liquide. Chaque thème ainsi conçu ne doit son existence qu'à un choix arbitraire. Chacun peut ouvrir une voie d'accès à un univers sémantique, mais peut aussi conduire à des impasses, en masquant le caractère systématique de cet univers."

Les critiques littéraires ainsi que les travaux universitaires qui s'occupent souvent, comme le souligne François Rastier, de l'analyse des thématiques dans les textes, sont – qu'on le veuille ou non – influencés par une large proportion de subjectivité. Jean-Marie Viprey fait l'objection suivante à propos des critères qui régissent les études thématiques⁴⁷⁹ :

“Plusieurs facteurs exercent leur contrainte sur un tel réglage : l'intuition plus ou moins affichée du lecteur-chercheur, sa compétence particulière vis-à-vis de l'œuvre et de son intertexte ; la tradition critique, prompte parfois à monosémiser autour de quelques *thèmes* solidement installés et profondément rapportés à *la psychologie de l'auteur* ou à d'autres entités de l'horizon extratextuel.”

Georges Molinié et Alain Viala soulignent également un autre aspect à prendre en compte⁴⁸⁰ : “Le lecteur fait le texte. Il y projette ses images, y trace ses chemins dans les entrelacs des significations possibles. Mais dans l'écriture même, l'image du destinataire telle que le texte la construit, telle que l'écrivain l'imagine, est un rouage essentiel à la machinerie de la création littéraire.”

Il est donc très intéressant de rechercher une méthode qui repose sur l'impartialité – ou tente de le faire – et il s'avère que les techniques de lexicométrie sont parfaitement adaptées comme plate-forme pour ces analyses. “Le calcul statistique,

⁴⁷⁹ J.-M. Viprey (1997) : p. 65.

⁴⁸⁰ G. Molinié, A. Viala (1993) : la quatrième de couverture.

signale ainsi Pierre Lafon⁴⁸¹, permet en effet de donner un contenu précis à la notion souvent subjective de vocabulaire caractéristique dans les études littéraires.”

“Or, écrit-il, le modèle statistique étant de nature totalement étrangère à la réalité linguistique, il pourrait ne constituer pour nous qu’un instrument de mesure permettant de détecter les formes qui s’éloignent le plus de lui – dans le sens positif aussi bien que dans le sens négatif - afin de donner une description précise de la réalité.”

Il convient ensuite de définir les méthodes pour l’exploitation des données qui nous sont fournies par les logiciels lexicométriques. Il ne suffit pas, en effet, de dégager les emplois les plus fréquents dans les textes.

Dans cette perspective, l’analyse littéraire purement traditionnelle semble en effet insuffisante. L’analyse du contenu thématique, et pas seulement l’extraction des emplois les plus significatifs, doit également reposer sur des bases solides et impartiales dans la mesure du possible.

Comment donc définir et identifier des thèmes, retracer leurs liens privilégiés, dessiner leur évolution dans un corpus ?

L’approche par la théorie de l’isotopie semble pouvoir répondre à ces questions et contribuer à assurer nos ambitions d’impartialité.

Le concept d’isotopie introduit par A.-J. Greimas est défini en ces termes par l’auteur : “Ensemble redondant de catégories sémantiques qui rend possible la lecture uniforme du récit telle qu’elle résulte des lectures partielles des énoncés et la réalisation de leurs ambiguïtés, qui est guidée par la recherche de la lecture unique.”

⁴⁸¹ P. Lafon (1984) : p. 102.

L'isotopie sémantique est définie comme effet de la récurrence d'un même sème⁴⁸². "Les relations d'identité entre les occurrences du sème isotopant⁴⁸³ induisent des relations d'équivalence entre les sémèmes qui les incluent."

Le concept de l'isotopie opère à plusieurs niveaux et rend notamment compte de la manière dont les textes entiers se trouvent situés à des niveaux sémantiques homogènes. L'isotopie permet la lecture d'un texte, en établissant son homogénéité. La cohérence du discours est fondée sur la répétition d'éléments semblables ou compatibles.

Un thème générique, écrit François Rastier⁴⁸⁴, est défini par un sème ou une structure de sèmes génériques⁴⁸⁵ récurrents. Cette récurrence définit une isotopie ou un faisceau d'isotopies génériques (c'est-à-dire un groupement de sèmes génériques co-récurrents), et particulièrement domaniales, déterminant le "sujet" (*topic*) du texte, et plus précisément induisant les impressions référentielles dominantes.

Il s'avère que les nouvelles techniques que nous offrent désormais les logiciels lexicométriques : la recherche thématique d'Hyperbase et l'analyse des univers lexicaux du logiciel Lexicométrie, ouvrent la voie à une analyse thématique dans cet esprit où la récurrence des différents éléments de l'entourage d'un mot et le caractère des liens qu'ils tissent entre eux sont identifiables par l'ordinateur.

Notre corpus Le Clézio, qui compte plus de 50.000 formes différentes, s'adapte parfaitement à cette technique et l'œil impartial de l'ordinateur permet bien de compléter la multitude d'études qui ont déjà été menées sur les thématiques dans l'œuvre de notre écrivain.

⁴⁸² F. Rastier, "La macrosémantique", in F. Rastier, M. Cavazza, A. Abeillé (1994) : p. 222-223.

⁴⁸³ Isotopant : se dit d'un sème dont la récurrence induit une isotopie.

⁴⁸⁴ F. Rastier "La thématique" in F. Rastier, M. Cavazza, A. Abeillé (1994) : p. 177.

⁴⁸⁵ La définition du sème générique est celle d'un élément d'un classème (l'ensemble des sèmes génériques d'un sémème), marquant l'appartenance du sémème (c'est-à-dire le signifié d'un morphème) à une classe sémantique (taxème, domaine, ou dimension).

Dans cette partie nous nous intéresserons d'abord aux spécificités lexicales, dans une perspective exogène - prenant le corpus de *Frantext* comme référence – aussi bien que dans une perspective endogène, en étudiant les excédents ainsi que les déficits de chaque œuvre par rapport à l'ensemble du corpus leclézien, avec le recours de deux logiciels différents : Hyperbase et Léxicométrie.

Cette étude mettra aussi en exergue des thématiques, ou isotopies, récurrentes dans l'œuvre. Nous en explorerons certaines plus en détail, notamment *la nature, la mer, la terre, le bestiaire* ainsi que *le regard*, omniprésent dans l'œuvre leclézienne.

Un autre palier, supérieur aux autres, de l'étude isotopique, est la notion d'intertextualité. L'intertextualité est un facteur très important chez Le Clézio et qui mérite également une approche technique ; nous nous intéresserons aussi dans ce chapitre à une particularité de l'écriture leclézienne : les récits intercalés.

Il convient toutefois, avant de procéder à une analyse sémantique d'un corpus, de bien connaître les différents textes de celui-ci et la relation qui les lie en fonction de leur contenu lexical. C'est donc par l'analyse de la distance lexicale – ou la connexion –, que nous nous proposons de commencer cette étude.

5.2. La distance lexicale, introduction.

“Je voudrais bien pouvoir écrire comme on parle. Je voudrais bien que cesse un jour le mince rempart de papier blanc qui me protège, qui me sépare. Qu’y a-t-il donc derrière ce carré éblouissant, quel paradis ou quel enfer se cache de l’autre côté de cette fenêtre opaque ? Oui, tout ça que je voudrais bien savoir. La grande hypocrisie de l’écriture - et aussi cette grande joie de la distance établie, des gants que je mets, pour toucher au monde, pour me toucher - c’est donc cette matière qui s’interpose, entre moi et moi, ce chemin détourné par lequel je m’adresse.”

Le livre des fuites, page 236.

Nous avons pu constater, dans les diverses études statistiques effectuées sur notre corpus, jusqu’ici s’appuyant uniquement sur les critères formels, externes, non seulement l’influence des différents genres littéraires dans les classifications mais aussi une opposition à l’intérieur même de la catégorie romanesque.

Cette opposition est-elle également observable lorsque l’on s’intéresse au contenu lexical et plus précisément à l’étude de la distance entre les vocabulaires des différents textes du corpus ? Partagent-ils les mêmes thèmes ? Y a-t-il une connexion thématique entre les œuvres d’un même genre littéraire chez Le Clézio ? La distinction en périodes, à l’intérieur d’un genre, est-elle confirmée par l’analyse de la distance lexicale ?

Le calcul de la distance entre les vocabulaires de plusieurs textes d’un corpus vise en effet à répondre à la question : quels sont les textes les plus proches et les plus éloignés du point de vue de leur contenu lexical et thématique ? Le problème a aussi été étudié sous le nom de connexion lexicale définie comme “l’intersection du vocabulaire de deux textes”⁴⁸⁶.

⁴⁸⁶ Ch. Muller (1977) : p. 145.

D'une manière générale, la distance absolue entre deux textes A et B sera l'union de ces deux textes moins leur intersection, soit : $(A \cup B) - (A \cap B)$.

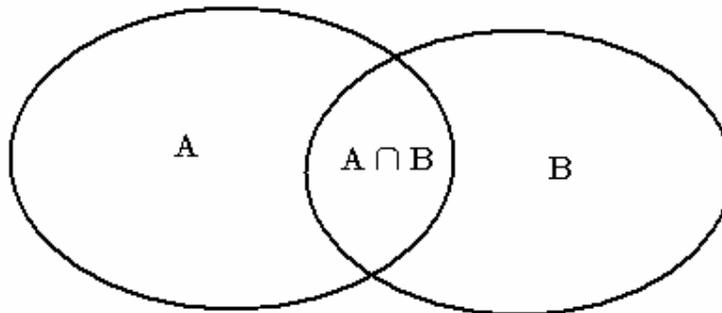


Figure n°143 : La distance – ou la connexion – lexicale.

Lorsque l'on mesure la distance lexicale, il s'agit de comparer le vocabulaire des textes, pris deux à deux, en neutralisant les différences de taille. Une faible distance indiquera que la plus grande partie de la surface du couple considérée est commune. A l'inverse, plus la distance s'accroît, plus cette partie commune est faible, laissant une large place au vocabulaire spécifique de chacun des textes.

Toutefois la mesure de la distance lexicale ne va pas de soi et crée de nombreuses difficultés :

“La distance entre deux textes, écrit Étienne Brunet⁴⁸⁷, c'est comme la distance entre deux êtres ou entre deux cultures. Cela n'a rien de mécanique ou métrique. Il ne semble pas qu'on puisse appliquer là la mesure”.

De nombreuses formules ont pourtant été proposées pour opérer des calculs de distance et des classements de textes parmi lesquelles deux démarches doivent être distinguées : d'une part, l'étude de la “connexion entre les vocabulaires”, lorsque

⁴⁸⁷ É. Brunet, “Peut-on mesurer la distance entre deux textes ?”, in *Corpus littéraires – Recueil et numérisation, analyses assistées, didactique*, Paris 20-21 octobre 2000, F. Rastier (éd.), Paris, 2000, p. 2.

l'on considère la présence ou l'absence de chaque vocable, dans chacun des textes, sans considération de sa fréquence (la démarche traditionnelle en analyse des données⁴⁸⁸), d'autre part la "connexion des textes", quand la comparaison porte sur les surfaces de textes (en tenant compte de la fréquence d'emploi de chaque vocable) et vise à répondre à la question : combien de mots sont communs ou encore : quelle proportion des deux textes leur est propre ?

"For some purposes, écrit Zdenek Hubalek⁴⁸⁹, however, other 'admissible' coefficients would be more optimal, and the choice of a measure should be related to the nature of the data. It is tentatively suggested that three or so alternative coefficients be used and the results compared to the same data basis; moreover, significance tests on association should be carried out whenever possible."

Il convient en effet de toujours exploiter plusieurs méthodes avant de tirer des conclusions quant à la distance lexicale.

Dans notre étude de la distance lexicale de l'œuvre de Le Clézio nous nous proposons d'étudier des résultats issus de différentes méthodes en soumettant notre corpus, dans ses différentes versions, à plusieurs traitements statistiques, et de mettre en parallèle la forme et le lemme, pour pouvoir comparer les résultats de l'étude de la distance, ou connexion lexicale, sous ces deux modes d'appréhension.

Nous nous appuierons dans cette analyse sur quatre calculs différents :

- sur V qui est le calcul "absence/présence" : on compte les vocables en commun sans se préoccuper de leur fréquence. Ce calcul effectué par le logiciel Hyperbase portera sur les formes graphiques et sur les lemmes.

⁴⁸⁸ Cf. J.-P. Benzecri (1976), R. Tomassone (éd.) (1988), É. Brunet, "Une mesure de la distance intertextuelle : la connexion lexicale", in *Le nombre et le texte. Revue informatique et statistique dans les sciences humaines*, Université de Liège, 1988 et D. Labbé, P. Hubert, "La connexion des vocabulaires", in *4e journées internationales d'analyse statistique des données textuelles*, S.Mellet (éd.), Université de Nice, 1998, p. 361-369.

⁴⁸⁹ Z. Hudalek (1982) : p. 687.

- sur N qui est le calcul sur la fréquence : on prend en considération la totalité du texte (mots x fréquences). Cette analyse se base sur deux méthodes différentes ; d'un côté exploitant la technique du logiciel Lexicométrie de Dominique Labbé, effectué sur les vocables, et de l'autre côté sur Hyperbase et s'intéresse de nouveaux aux formes graphiques et aux lemmes.

Pour ces deux premières méthodes, la représentation graphique sera celle de l'A.F.C.

- Enfin, nous recourrons aussi au calcul de distance de l'analyse arborée élaborée par Xuan Luong et Jean-Pierre Barthélemy.

Les différentes analyses s'appuyant sur ces méthodes permettent de tirer des conclusions précises sur la distance lexicale dans l'œuvre de Le Clézio.

Nous nous intéressons en premier lieu au plus traditionnel des calculs, celui qu'avait proposé Jaccard, suivi par les travaux de Charles Muller, qui tient compte de la présence et de l'absence de chaque mot dans les différents textes étudiés.

5.2.1. La distance lexicale calculée sur la présence V.

La distance lexicale s'intéresse, comme nous l'avons dit, au vocabulaire intégral de chacun des textes du corpus et au repérage de ceux qui partagent des thèmes semblables. Traditionnellement, pour un mot donné, seule compte sa présence - ou son absence - dans le texte considéré. Ou plus exactement, pour deux textes dont on cherche à apprécier la connexion, un mot contribue à rapprocher ces textes s'il est commun aux deux, et à augmenter la distance s'il est privatif et ne se rencontre que dans un seul. Ce calcul est appelé "calcul sur la présence V".

Pour chaque paire considérée, la distance obtenue tient compte de l'étendue de l'un et l'autre vocabulaires, selon la formule ; $d = ((a - ab)/a) + ((b - ab)/b)$, où ab désigne la partie commune aux vocabulaires a et b ($a - ab$ et $b - ab$ recouvrent les parties privatives).

Le tableau ci-dessous rend compte de la distance globale des textes deux à deux, s'appuyant sur les formes graphiques du corpus A.

Distance établie sur V (méthode Jaccard), formes graphiques.

Procès	99	100	100	114	109	109	121	115	135	108	115	121	118	110	126	109	116	113	122	132	112	130	115	112	121	127	114	115	129	128	112	
Fièvre	100	89	90	107	100	102	113	107	134	101	107	120	114	103	120	104	112	110	120	132	110	131	110	108	120	127	110	116	128	125	109	
Déluge	100	90	89	106	97	99	112	106	132	100	107	118	116	101	119	103	111	108	119	131	108	128	109	106	115	127	109	113	128	125	108	
Extase	114	107	106	104	111	112	113	116	131	120	105	117	123	115	123	116	116	125	119	125	122	125	123	121	129	120	116	126	124	126	122	
Fuite	109	100	97	111	95	96	110	103	133	101	104	117	115	101	114	103	110	111	118	131	112	128	110	108	117	127	110	117	128	124	113	
Guerre	109	102	99	112	96	95	109	102	135	103	105	118	114	102	116	104	111	114	121	133	112	130	113	110	117	128	112	117	129	126	116	
Mydras	121	113	112	113	110	109	105	106	147	115	106	131	133	111	126	115	114	131	129	133	125	148	122	120	137	130	116	129	129	133	124	
Voyage	115	107	106	116	103	102	106	98	140	99	99	118	111	99	115	101	106	116	123	137	113	136	114	109	120	133	108	119	134	127	117	
Prophét	135	134	132	131	133	135	147	140	112	139	128	141	148	135	134	142	134	148	132	113	142	143	136	138	149	129	133	144	116	136	142	
Mondo	108	101	100	120	101	103	115	99	139	90	106	123	114	91	120	97	106	104	122	134	102	137	101	97	114	132	103	110	128	125	106	
Inconnu	115	107	107	105	104	105	106	99	128	106	98	113	110	101	110	106	105	120	114	126	117	125	117	113	122	122	108	124	121	121	119	
Iceberg	121	120	118	117	117	118	131	118	141	123	113	112	139	118	135	121	117	134	125	125	125	142	124	121	141	125	117	127	125	132	125	
Arbres	118	114	116	123	115	114	133	111	148	114	110	139	109	112	129	112	115	126	132	131	122	149	121	113	135	129	115	123	124	137	123	
Désert	110	103	101	115	101	102	111	99	135	91	101	118	112	88	112	89	99	104	117	129	102	129	102	94	115	126	97	107	124	115	105	
Villes	126	120	119	123	114	116	126	115	134	120	110	135	129	112	109	118	116	130	130	126	126	144	120	118	136	130	116	132	120	131	128	
Ronde	109	104	103	116	103	104	115	101	142	97	106	121	112	89	118	88	102	110	120	134	101	135	106	97	119	128	100	106	128	126	106	
Cherche	116	112	111	116	110	111	114	106	134	106	105	117	115	99	116	102	94	109	101	128	108	125	106	103	111	125	95	115	123	118	109	
Angoli	113	110	108	125	111	114	131	116	148	104	120	134	126	104	130	110	109	103	126	130	108	144	105	104	118	131	107	114	125	128	105	
Rodrigu	122	120	119	119	118	121	129	123	132	122	114	125	132	117	130	120	101	126	100	124	125	129	120	121	125	122	111	129	122	120	121	
Rêve	132	132	131	125	131	133	133	137	113	134	126	125	131	129	126	134	128	130	124	99	135	121	131	131	129	116	127	138	100	125	133	
Printem	112	110	108	122	112	112	125	113	142	102	117	125	122	102	126	101	108	108	125	135	96	134	105	97	117	126	103	100	130	127	103	
Siranda	130	131	128	125	128	130	148	136	143	137	125	142	149	129	144	135	125	144	129	121	134	118	133	132	147	123	122	134	119	133	133	
Onitshal	115	110	109	123	110	113	122	114	136	101	117	124	121	102	120	106	106	105	120	131	105	133	100	101	115	127	103	111	126	124	105	
Étoile	112	108	106	121	108	110	120	109	138	97	113	121	113	94	118	97	103	104	121	131	97	132	101	93	111	128	99	105	127	123	103	
Pawana	121	120	115	129	117	117	137	120	149	114	122	141	135	115	136	119	111	118	125	129	117	147	115	111	109	133	110	123	129	132	119	
Diego	127	127	127	120	127	128	130	133	129	132	122	125	129	126	130	128	125	131	122	116	126	123	127	128	133	115	121	129	117	125	124	
Quarant	114	110	109	116	110	112	116	108	133	103	108	117	115	97	116	100	95	107	111	127	103	122	103	99	110	121	94	107	122	119	102	
Poisson	115	116	113	126	117	117	129	119	144	110	124	127	123	107	132	106	115	114	129	138	100	134	111	105	123	129	107	99	133	132	107	
Fête	129	128	128	124	128	129	129	134	116	128	121	125	124	124	120	128	123	125	122	100	130	119	126	127	129	117	122	133	99	122	129	
Nuages	128	125	125	126	124	126	133	127	136	125	121	132	137	115	131	126	118	128	120	125	127	133	124	123	132	125	119	132	122	114	126	
Hasard	112	109	108	122	113	116	124	117	142	106	119	125	123	105	128	106	109	105	121	133	103	133	105	103	119	124	102	107	129	126	101	
Prc Fiv Déu Exa Fut Gur Myr Voa Prp Mod Ino Icb Arr Dée Vil Rod Chr Ano ror Rée Prn Sia Ont Éti Paa Dig Qur Pos Fée Nug Haa																																

Dans ce tableau chaque élément correspond à la somme de deux quotients ($a - ab/a$ et $b - ab/b$), c'est-à-dire du rapport du vocabulaire exclusif au vocabulaire, somme qui évolue nécessairement entre 0 et 2^{490} , le rapport ayant

⁴⁹⁰ Pour faire des économies de place, ont été ôtées les virgules dans les tableaux : pour, par exemple, 114 il faut comprendre 1,14.

pour limites 0 et 1. C'est cette somme qui constitue la mesure de la distance entre les textes.

Les valeurs du tableau de distance se reflètent en miroir de chaque côté de la diagonale principale (car la distance de A vers B est la même que de B vers A). Quant à la diagonale principale, elle devrait ne comporter que la valeur 0 ; la distance d'un texte à lui-même est nécessairement nulle. Si pourtant une valeur non nulle s'y trouve, c'est par un artifice nécessité par le traitement des profils⁴⁹¹. C'est pourquoi le logiciel a porté sur la diagonale une valeur qui correspond au minimum de la série en question, en admettant que la distance de soi à soi est inférieure ou au plus égale à toute autre distance observée dans la série.

Il faut admettre que ce tableau n'est pas facile à lire et l'interprétation n'est pas très aisée pour une matrice comportant 31 sous-corpus comme le nôtre. La représentation graphique de l'analyse factorielle illustrant les multiples accords qui lient les membres du réseau rend l'interprétation plus confortable :

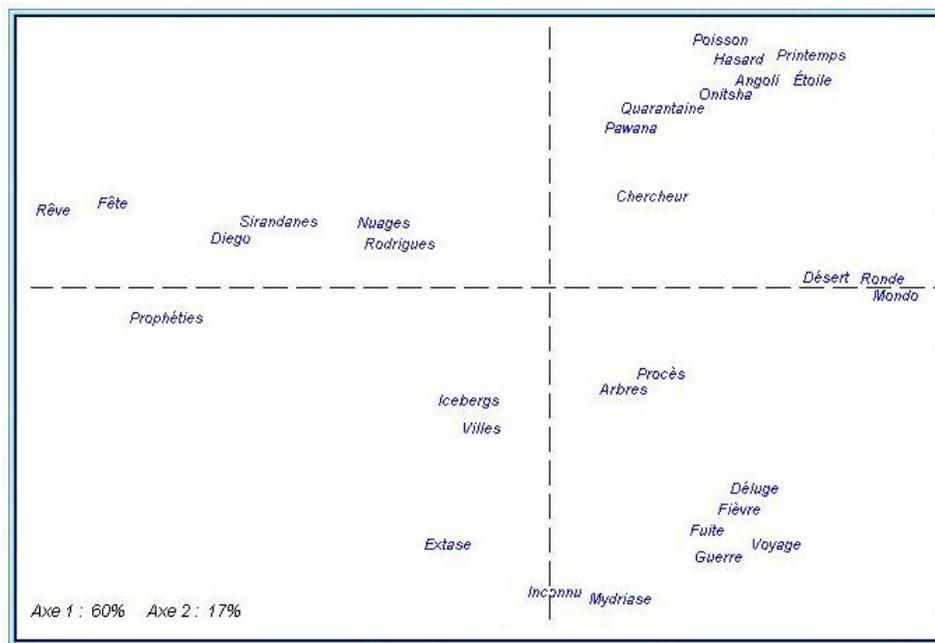


Figure n°144 : Analyse factorielle de la distance lexicale du corpus A. Le calcul est établi sur les formes sans considération des fréquences (V).

⁴⁹¹ É. Brunet signale que les valeurs nulles ont une influence dans les calculs multidimensionnels (analyse de correspondance, ou analyse arborée), et leur influence est d'autant plus nuisible que leur valeur s'écarte plus fortement des autres éléments de la même ligne (ou de la même colonne).

Pour qui connaît l'œuvre de Le Clézio, la lecture de ce graphique est aisée. Le premier facteur – horizontal – est sensible au genre. Il oppose le rêve à la réalité, les textes mythiques, à gauche, où l'action s'efface au profit de la contemplation, aux romans de la moitié droite, dont l'écriture est plus traditionnelle ; cet axe illustre bien la diversité de l'écrivain. A l'extrême gauche le graphique rend compte des thèmes partagés par les ouvrages qui traitent des rites du monde amérindien : *Le rêve mexicain* et *La fête chantée* ainsi que *Les prophéties de Chilam Balam*. Les thèmes de l'exotisme, du voyage et de la quête sont ensuite retrouvés avec *Diego et Frida*, *Sirandanes*, *Gens des nuages* et *Voyage à Rodrigues* qui se détache d'ailleurs des autres romans, non seulement – comme nous l'avons vu auparavant – par des critères formels, mais aussi par son vocabulaire, avoisinant celui des récits de voyages. Ce “journal”, inspiré de celui du grand-père, est effectivement, sous de nombreux aspects – notamment les nombreuses indications géographiques –, plus proche de ces récits :

“N'était-ce pas le même rêve qui s'était dissipé lorsque Coronado, croyant découvrir les cités de Cibola aux toits d'or et de pierreries, arriva enfin devant les villages de boue séchée des pueblos, n'était-ce pas le même rêve, lorsque René Caillié entra pour la première fois dans Tombouctou et vit que la cité mystérieuse n'était en fait qu'un rendez-vous de chameliers ? Comment mon grand-père a-t-il pu croire à la légende du trésor de Golconde, et surtout à celle de la dot de la fille d'Aurangzeb capturée par Avery ?”⁴⁹²

En bas à gauche du graphique nous trouvons les livres qui privilégient le mythe comme *Les icebergs*, *Trois villes saintes*, ainsi que l'essai *L'extase matérielle* qui se détache toutefois des autres. Les ouvrages de vocation poétique *L'inconnu sur la terre* et *Mydriase* semblent constituer la frontière entre la réalité et le mythe, l'action

⁴⁹² *Voyage à Rodrigues*, p. 43.

et la contemplation, en ouvrant la voie vers les livres de fiction qui se trouvent de l'autre côté de l'axe vertical.

Les romans inspirés par le “nouveau roman” se trouvent ensemble, à l'écart des autres en bas à droite. Les trois livres qui suivent la rupture dans l'œuvre : *Mondo*, *Désert* et *La ronde*, s'agglutinent les uns aux autres, partageant en grande partie le même vocabulaire. Les romans et les recueils de nouvelles les plus traditionnels se groupent également, en haut du graphique, signe qu'ils ont des thèmes communs.

Le second facteur de l'analyse factorielle parcourt en effet la chronologie du bas vers le haut du graphique et il rend ainsi compte de l'évolution de l'écrivain. A partir des premiers livres “nouveau roman” tout en bas du graphique, nous suivons au fur et à mesure que l'œuvre progresse “l'ascension” de l'évolution thématique qui vers la fin de l'œuvre est située tout en haut du graphique avec *Hasard* (1999), *Poisson d'or* (1997) et *Printemps et autres saisons* (1989), qui d'après notre analyse partagent le même univers intellectuel et sont écrits dans un style semblable.

Lorsque l'on parle d'un vocabulaire commun il ne s'agit évidemment pas d'un calque exact : les mots peuvent être placés dans un ordre différent et les thèmes intervenir à des moments différents du récit. Un autre facteur important de variation est celui des nombreuses flexions en français. Par exemple, les temps des verbes ou le genre et le nombre des adjectifs jouent-ils en rôle important dans cette analyse ?

Pour répondre à cette question nous effectuons la même étude avec le recours au logiciel Hyperbase, mais cette fois-ci l'analyse s'appuie sur les lemmes du corpus B. La représentation graphique de l'analyse factorielle ci-dessous permet la comparaison :

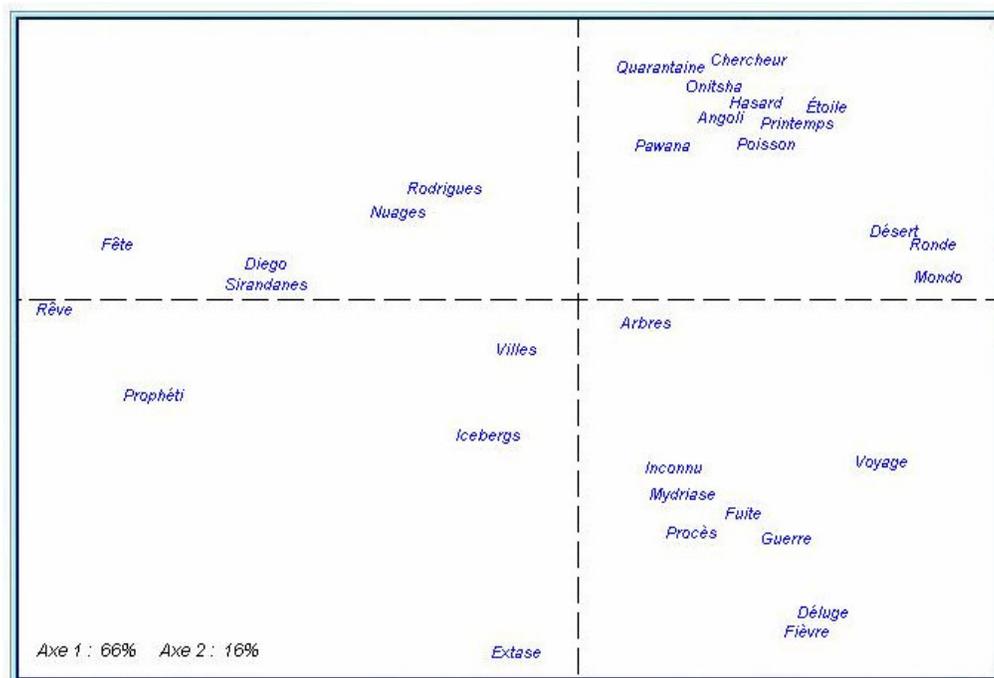


Figure n°145 : Analyse factorielle de la distance lexicale du corpus B. Le calcul est établi sur les lemmes sans considération des fréquences (V).

Les deux graphiques sont très semblables. Les grandes distances, qui s’expliquent par des différences de genre, sont aussi bien marquées et les constellations entre les différents ouvrages du corpus sont sensiblement les mêmes. Il y a un léger décalage des ouvrages “poétiques”, *L’inconnu sur la terre* et *Mydriase*, par rapport au graphique précédent, et il semblerait que dans ce cas la lemmatisation pourrait avoir une petite influence sur les résultats. Toutefois, cette différence est minime, ce qui nous amène à confirmer⁴⁹³ que la lemmatisation n’a pas une importance fondamentale sur les résultats de l’analyse de la distance lexicale, selon les deux méthodes employées, dans notre corpus.

Etant donné que l’opposition des genres littéraires est vraiment un facteur prépondérant dans l’analyse de la distance lexicale dans notre corpus, et que les

⁴⁹³ Remarque déjà faite à plusieurs reprises par É. Brunet, notamment dans “Qui lemmatise dilemme attise” in *Scolia, 11èmes rencontres linguistiques en Pays Rhénan*, L. José, A.Theissen (éds.), Publications de l’Université Marc Bloch n°13, Strasbourg, 2000, p. 7-32 et dans “Le lemme comme on l’aime”, in *JADT 2000, 6èmes Journées internationales d’Analyse statistique des Données Textuelles*, A. Morin, P. Sébillot (éds.), Irisa, Inria, Saint-Malo 13-15 mars, p.221-233.

variations à l'intérieur du genre romanesque sont relativement larges, la distance lexicale à l'intérieur de ce genre mérite d'être vue de près.

5.2.2. La distance lexicale à l'intérieur du corpus romanesque.

A l'intérieur du genre romanesque (le corpus E), l'analyse de la distance lexicale rend compte de la diversité de l'écrivain et de la formation des sous-groupes qui séparent les "nouveaux romans" des ouvrages dont l'écriture est plus traditionnelle :

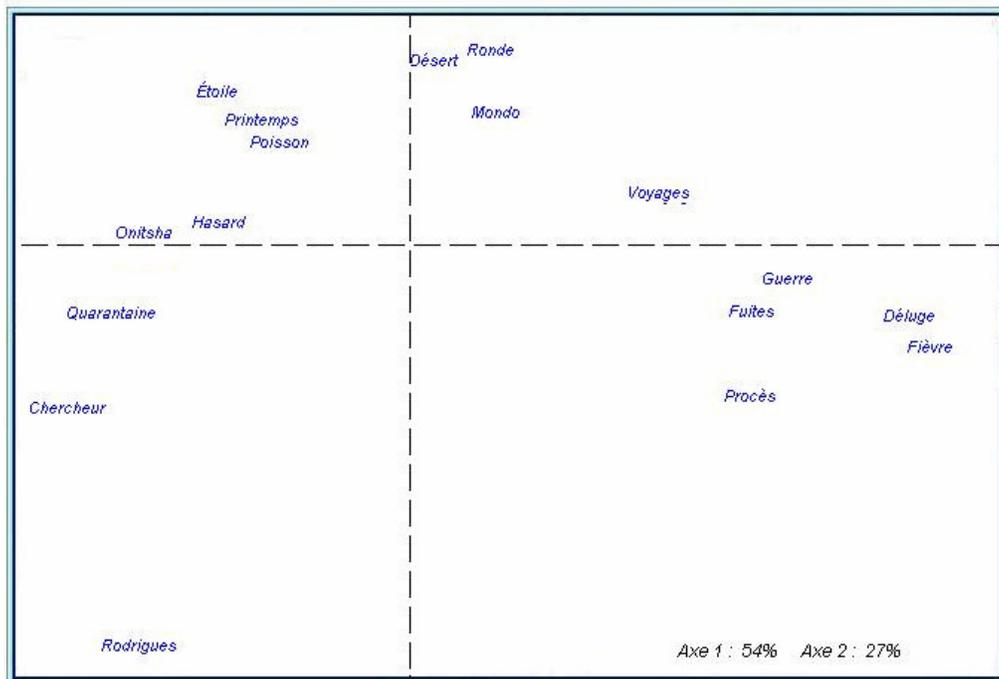


Figure n°146 : Analyse factorielle de la distance lexicale du corpus E. Le calcul est établi sur les lemmes sans considération des fréquences (V).

En effet, le graphique met en exergue la diversité et l'évolution de la production romanesque de Le Clézio. A l'extrême droite de la figure nous trouvons, groupés ensemble, les premiers livres de l'auteur teintés du style "nouveau roman". Le roman *Voyages de l'autre côté* annonce le changement dans l'œuvre en s'orientant vers la deuxième période de l'œuvre romanesque leclézienne représentée par

Mondo, *Désert* et *La ronde*, étroitement unis. A gauche, plus à l'écart, nous trouvons les romans qui suivent et qui ont pour thème l'Océan Indien : *Le chercheur d'or*, *La quarantaine* et *Voyage à Rodrigues*, le dernier comme toujours un peu isolé du reste.

Le graphe de répartition, avec sa forme caractéristique, permet également de constater l'importance de la chronologie que nous avons déjà observée auparavant mais qui, cette fois-ci, semble coïncider avec l'évolution du style et du genre et qui rend compte d'une progression thématique régulière dans l'œuvre leclézienne.

Nous avons déjà évoqué dans notre étude le fait que certains calculs semblent plus ou moins sensibles à la différence de taille des sous-corpus. C'est le cas du calcul de la distance lexicale, qui demanderait une certaine homogénéité en ce qui concerne la taille des différents textes du corpus.

5.2.3. La distance lexicale du corpus homogénéisé.

Dominique Labbé⁴⁹⁴ n'a cessé d'avertir que "la technique du calcul ne doit pas être employée sur des textes trop brefs et, quand les tailles sont fortement différentes, il faut résister à la tentation de mesurer les distances avec trop de décimales."

Le calcul risquerait en fait d'être influencé par la différence de taille entre les textes et cette influence sera d'autant plus forte que certains des textes seront très brefs. De manière générale, il est donc recommandé d'exclure les "géants" et les "nains" de l'analyse et de ne pas avoir un texte bref inférieur au dixième du plus grand texte, ce rapport de 1/10 étant empiriquement la limite de la comparaison possible entre des textes. De même, dans les corpus très étendus, la comparaison n'a guère de sens pour les trop petits textes.

⁴⁹⁴ D. Labbé (2000) : p. 85.

Nous avons donc eu recours à notre corpus homogénéisé, le corpus C, où la longueur des différents textes est sensiblement la même, afin d'éliminer d'éventuelles erreurs dues à une inégalité de taille trop importante des sous-corpus. L'illustration graphique de l'analyse factorielle de la distance lexicale est calculée, toujours, sur "V" :

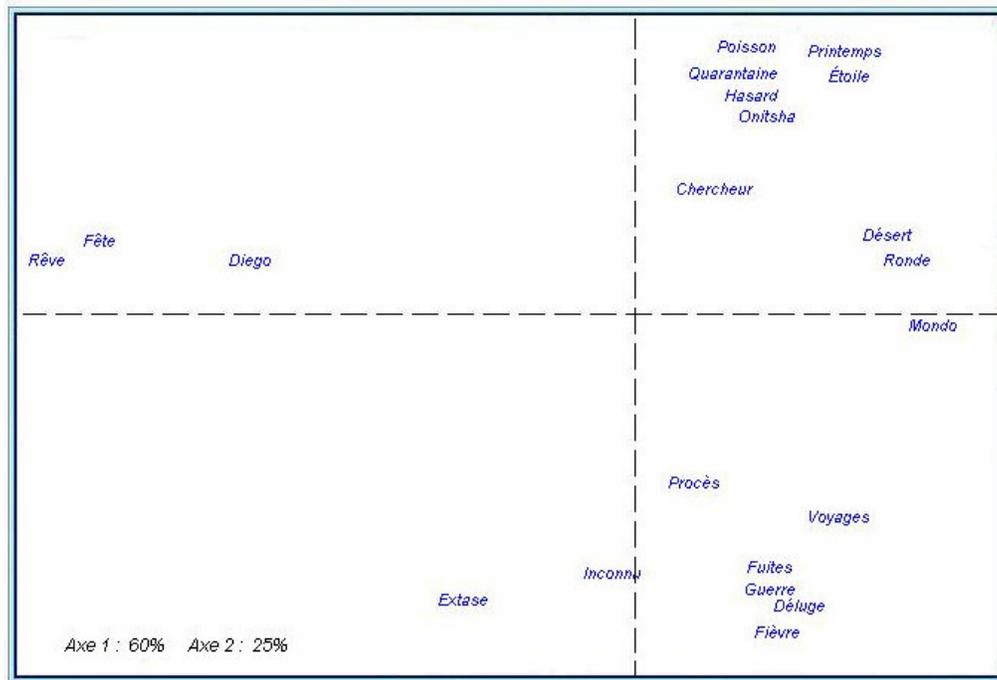


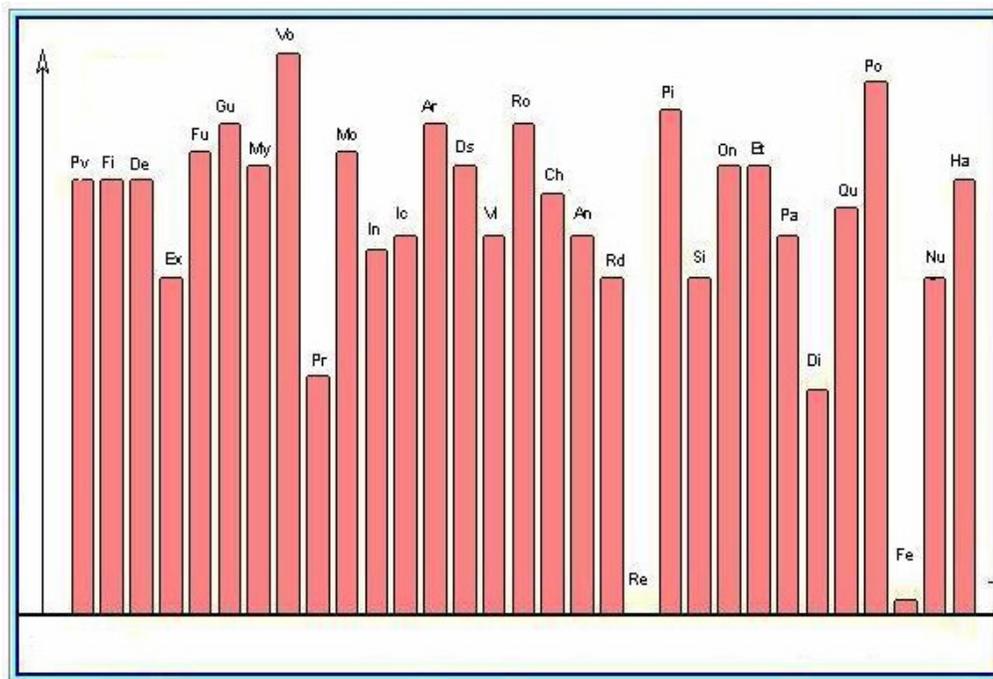
Figure n°147 : Analyse factorielle de la distance lexicale du corpus C. Le calcul est établi sur les formes sans considération des fréquences (V).

D'après ce graphique, les différences de taille des différents livres de notre corpus n'ont pas d'influence remarquable sur les résultats. Les tendances sont les mêmes, les oppositions et les attirances sont réparties de la même façon que dans les calculs précédents et soulignent une divergence générique importante.

En effet certains livres, notamment *Le rêve mexicain*, se trouvent dans toutes nos analyses à l'extrême gauche du graphique de l'analyse factorielle. Il est intéressant de regarder de près la distance entre cet ouvrage – ainsi que celle des autres textes très excentrés – par rapport au reste du corpus.

5.2.4. La distance lexicale texte par texte.

L'histogramme ci-dessous présente les résultats du calcul de la distance lexicale de l'essai *Le rêve mexicain* par rapport aux autres livres du corpus. Chaque bâton de la figure a une longueur proportionnelle à la distance qui sépare le texte correspondant du texte choisi comme pôle d'observation, *Le rêve mexicain* :



Figure°148 : *Le rêve mexicain*. Distances des autres textes du corpus (le calcul est établi sur les lemmes, sans considération des fréquences).

La distance reste faible tant qu'on a affaire à des livres qui traitent de la culture amérindienne, même s'il s'agit d'un genre différent, dont relèvent l'ouvrage ethnologique *La fête chantée*, l'introduction de l'ouvrage *Les prophéties de Chilam Balam* ainsi que *Diego et Frida* qui se déroule au Mexique. Elle est très longue pour tous les autres textes.

Est-ce que le portrait diffère beaucoup lorsqu'on établit la distance sur les seules formes ? On prendra la mesure de l'écart en considérant l'ouvrage qui marque la

rupture dans l'œuvre leclézienne : *Mondo et autres histoires*, et cette fois-ci nous nous limitons au corpus romanesque E.

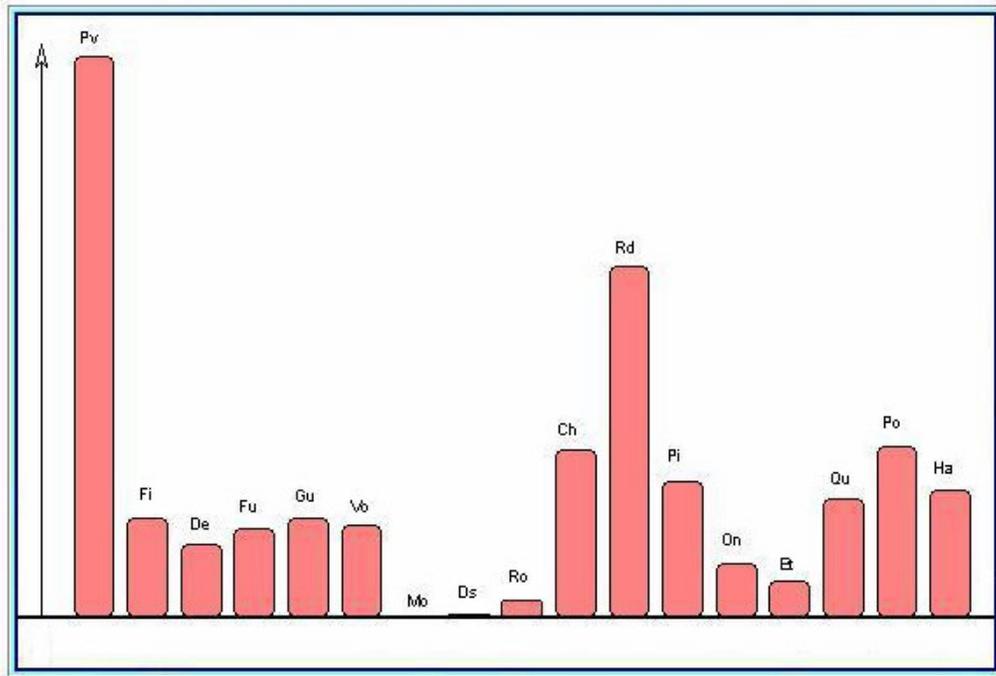


Figure n°149 : *Mondo et autres histoires*. Distances des autres textes du corpus E (le calcul est établi sur les lemmes, sans considération des fréquences (V)).

Le mouvement de ce graphique est tout à fait remarquable car il rend compte des trois mêmes périodes que nous avons pu constater si souvent dans notre étude formelle et qui se confirment maintenant du point de vue du contenu lexical. Non seulement notre écrivain change de style en cours de route, mais il change également de vocabulaire et de thématique. La distance est minime tant qu'on a affaire aux livres qui suivent *Mondo : Désert* et *La ronde*. Elle est plus grande quand on s'éloigne dans le temps, aussi bien vers la période précédente qui caractérise les débuts de l'écriture leclézienne que vers la période suivante qui instaure de nouvelles thématiques ; on notera bien sûr la très grande distance du premier livre, *Le procès-verbal*.

Les conclusions seront très proches d'un portrait à l'autre du corpus, chacun portant témoignage des traits caractéristiques communs. Le genre apparaît –

comme il le fait le plus souvent dans d'autres corpus aussi - comme le premier principe discriminant dans ce corpus. Néanmoins, lorsque les ouvrages de genres littéraires divers partagent le même univers lexical, c'est-à-dire quand ils traitent de mêmes thèmes, ils arrivent parfois à combler l'écart générique.

Cependant, ce calcul de la distance basé sur la simple présence ou absence mesure une distance essentiellement sensible aux variations thématiques, donne une grande importance aux hapax et aux mots rares et néglige les fluctuations de densité parfois importantes chez les mots plus fréquents (comme les mots outils et les verbes usuels). Les classes de fréquences élevées, reflétant dans une grande mesure les paramètres stylistiques, perdent ainsi tout poids dans le calcul, puisqu'elles se trouvent nécessairement dans la partie commune et inévitable du vocabulaire (*ab*).

“Il est évident, écrit Étienne Brunet⁴⁹⁵ à ce propos, que si le partage des fréquences est inégal (par exemple 1 occurrence dans le texte A et 19 dans le texte B), il contribue moins à rapprocher les deux textes que si la répartition était équilibrée, soit 10 occurrences dans chacun (en considérant que les deux textes sont de même étendue). Dans les deux cas le calcul précédent rangeait le mot à l'intersection des deux textes, ne tenant compte que de la présence/absence, en ignorant les disparités des fréquences.”

Dominique Labbé ajoute une autre objection à cette première : l'intersection des deux textes est comptée deux fois. On donne donc plus d'importance aux vocables communs qu'aux vocables propres à chacun.”

C'est pourquoi nous introduisons un autre calcul en complément de celui qui s'appuie sur la présence ou l'absence d'un mot : “le calcul sur la fréquence N”, qui considère pour un même mot le dosage des fréquences dans les deux textes

⁴⁹⁵ É. Brunet (2001) : p. 6.

comparés et vise ainsi à fournir une juste appréciation de la distance entre deux textes.

5.2.5. La distance calculée sur N.

Les outils informatiques permettent désormais de recenser exhaustivement les textes : l'association à chacun des vocables de sa fréquence d'emploi donne en effet le nombre total des mots d'un texte (son étendue ou sa dimension, notée N) et le calcul peut porter sur l'ensemble des textes "N" et non sur le seul vocabulaire "V".

Ce calcul sur "N" qui a été proposé depuis longtemps par Charles Muller, en complément de celui qui repose sur "V"⁴⁹⁶, opérant sur le modèle binomial, est ici ajusté par un algorithme proposé par Dominique Labbé, qui pour chaque mot apprécie la distribution réelle des fréquences dans les deux textes en la comparant non plus à la répartition théorique, mais à l'écart maximal possible dans cette distribution⁴⁹⁷. Il explique ainsi l'algorithme :

" V_a et V_b : nombre de vocables dans A et B (vocabulaire) ;

F_{ia} : fréquence du vocable i dans A ;

F_{ib} : fréquence du vocable i dans B ;

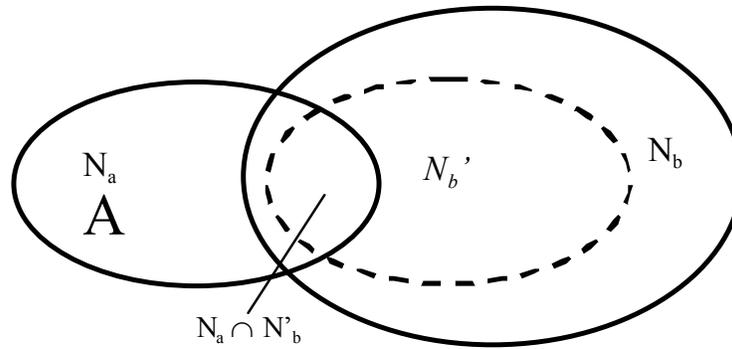
N_a et N_b : nombre de mots dans A et B (taille) ;

Nous proposons de simuler la réduction du plus grand des deux textes à la taille du plus petit.

Soit B' cette réduction de B en fonction de la taille de A :

⁴⁹⁶ Ch. Muller (1977) : p. 144-154.

⁴⁹⁷ D. Labbé (2000) : p. 85-94.



Soit $U(a,b)$, le coefficient de proportionnalité entre A et B :

$$U(a,b) = \frac{N_a}{N_b}$$

Tout vocable i de fréquence F_i dans B aura une fréquence attendue dans A égale à :

$$E_{ia(u)} = F_{ib} * U(a,b)$$

D'où l'on tire que :

$$N'_b = \sum_{v_b} E_{ia(u)}$$

Le minimum théorique (zéro) sera atteint quand le petit texte sera une sorte de modèle réduit du grand. Dans ce cas, tous les vocables de A se retrouvent dans B avec une fréquence telle que : $F_{ia} = E_{ia(u)}$. La distance absolue sera donc :

$$D_{V_a, b(u)} = \sum_{V_a, V'_b(E)} |F_{ia} - E_{ia(u)}|$$

A l'inverse, le maximum théorique (l'unité) devrait être atteint quand les deux textes n'ont aucun mot en commun. Au numérateur, comme au dénominateur, figureront $N_a + N'_b$. C'est donc le numérateur de la formule permettant d'obtenir la distance relative :

$$D(a,b) = \frac{\sum_{V_a, V'_b(E)} |F_{ia} - E_{ia(u)}|}{\sum_{V_a} F_{ia} + \sum_{V'_b} E_{ia(u)}} = \frac{\sum_{V_a, V'_b(E)} |F_{ia} - E_{ia(u)}|}{N_a + N'_b}$$

Enfin, pour tenir compte des deux objections ci-dessus : le résultat de cette méthode permet de ne considérer qu'une seule fois l'intersection des deux textes et on limite le calcul à l'ensemble des vocables de A mais seulement à V'_b ,

ensemble des vocables de B dont la fréquence est telle que l'on en attend au moins 1 dans A ($E_{ia(u)} \geq 1$). La somme de ces espérances donne N'_b .

Afin de mesurer cette distance, on utilise un indice qui varie entre 0 et 1 et permet d'établir une échelle des distances. Pour interpréter ces distances, l'indice obtenu correspond à celui de (N). Dominique Labbé suggère d'utiliser l'échelle suivante :

- en dessous de 0,15 : les vocabulaires employés sont très semblables, les textes appartiennent au même genre, ils développent un même thème avec des formulations extrêmement proches.
- De 0,15 à 0,25 : les textes appartiennent au même registre mais leur genre est différent ou ils développent des thèmes éloignés.
- De 0,25 à 0,35 : pour un même auteur, les textes appartiennent probablement à des genres ou à des registres différents.
- Au-dessus de 0,35 : les auteurs ou les registres sont différents ; les thèmes sont éloignés.

Lorsque l'indice est de 1 cela implique que les deux textes n'ont aucun mot en commun. Cette situation est aussi théorique que celle d'un indice de 0 – qui impliquerait deux textes en tous points identiques, car du moment où deux textes sont écrits dans la même langue, ils sont contraints d'utiliser un minimum de vocabulaire commun. Mais plus on se rapprochera de 1, plus les textes appartiendront à des univers intellectuels différents et plus ils développeront des thèmes éloignés.

Nous nous proposons d'étudier la distance lexicale, en prenant en compte la fréquence, selon deux méthodes : en appliquant d'abord la technique que propose Hyperbase puis celle fournie par le logiciel de Dominique Labbé, qui procède de façon un peu différente.

5.2.6. La distance lexicale calculée sur N, selon la technique d’Hyperbase.

Le logiciel Hyperbase permet désormais de calculer la distance lexicale s’appuyant sur la fréquence selon cette formule, introduite sous le nom “calcul Labbé”.

Étienne Brunet synthétise ainsi l’algorithme qu’il a introduit au programme :

$$D(A,B) = \sum d_i / \sum d_{max_i}$$

pour i variant du premier au dernier mot du vocabulaire des textes A et B.

Le tableau ci-dessous rend compte des distances lexicales entre les différentes œuvres dans la totalité du corpus, le corpus B :

Distance établie sur N, formes graphiques⁴⁹⁸.

Procès	.20	.20	.20	.29	.25	.26	.35	.30	.41	.30	.32	.35	.38	.31	.39	.27	.32	.31	.33	.40	.28	.37	.31	.30	.33	.37	.30	.28	.38	.34	.28
Fièvre	.20	.15	.16	.28	.21	.24	.32	.28	.40	.26	.28	.33	.39	.28	.35	.25	.30	.29	.34	.40	.28	.37	.29	.28	.31	.38	.29	.30	.38	.34	.28
Déluge	.20	.16	.15	.29	.20	.22	.31	.26	.39	.25	.28	.35	.37	.27	.34	.25	.30	.27	.33	.38	.28	.36	.27	.26	.30	.38	.28	.29	.36	.33	.27
Extase	.29	.28	.29	.26	.27	.29	.28	.32	.38	.36	.26	.30	.40	.34	.36	.31	.32	.37	.32	.37	.31	.33	.36	.33	.34	.36	.31	.33	.37	.35	.35
Fuite	.25	.21	.20	.27	.18	.18	.28	.23	.38	.26	.25	.33	.36	.25	.32	.24	.28	.30	.32	.38	.28	.36	.27	.27	.30	.38	.28	.30	.36	.32	.30
Guerre	.26	.24	.22	.29	.18	.18	.27	.21	.39	.26	.25	.33	.35	.25	.32	.24	.30	.31	.34	.37	.28	.36	.29	.27	.31	.37	.28	.29	.36	.34	.30
Mydrias	.35	.32	.31	.28	.28	.27	.24	.27	.40	.34	.24	.33	.38	.30	.32	.29	.33	.37	.37	.37	.36	.38	.34	.32	.37	.40	.34	.38	.37	.35	.36
Voyage	.30	.28	.26	.32	.23	.21	.27	.21	.42	.25	.22	.33	.32	.23	.31	.23	.29	.32	.38	.41	.30	.38	.30	.28	.33	.40	.28	.31	.39	.34	.32
Prophét.	.41	.40	.39	.38	.38	.39	.40	.42	.30	.42	.37	.41	.48	.38	.37	.41	.39	.43	.38	.30	.44	.37	.40	.39	.40	.39	.41	.46	.30	.35	.42
Mondo	.30	.26	.25	.36	.26	.26	.34	.25	.42	.21	.29	.38	.34	.21	.35	.22	.29	.24	.38	.41	.28	.40	.24	.23	.32	.40	.27	.29	.39	.36	.26
Inconnu	.32	.28	.28	.26	.25	.25	.24	.22	.37	.28	.22	.29	.36	.25	.26	.25	.28	.34	.32	.36	.32	.32	.32	.28	.31	.38	.30	.35	.35	.31	.34
Iceberg	.35	.33	.35	.30	.33	.33	.33	.33	.41	.38	.29	.29	.45	.35	.38	.34	.36	.40	.38	.34	.37	.37	.38	.36	.40	.40	.37	.38	.40	.37	.38
Arbres	.38	.39	.37	.40	.36	.35	.38	.32	.48	.34	.36	.45	.32	.36	.40	.35	.41	.39	.45	.43	.40	.47	.40	.38	.45	.45	.39	.40	.42	.43	.40
Désert	.31	.28	.27	.34	.25	.30	.23	.38	.21	.25	.35	.36	.18	.29	.18	.25	.27	.33	.37	.27	.36	.25	.22	.30	.36	.25	.31	.35	.30	.28	
Villes	.39	.35	.34	.36	.32	.32	.32	.31	.37	.35	.26	.38	.40	.29	.26	.32	.32	.38	.39	.36	.40	.37	.35	.33	.36	.40	.34	.42	.34	.34	.38
Ronde	.27	.25	.25	.31	.24	.24	.29	.23	.41	.22	.25	.34	.35	.18	.32	.18	.26	.27	.33	.39	.25	.36	.25	.22	.32	.36	.24	.28	.38	.33	.26
Cherche	.32	.30	.30	.32	.29	.30	.33	.29	.39	.29	.28	.36	.41	.25	.32	.26	.20	.29	.27	.38	.28	.36	.27	.24	.26	.36	.20	.30	.36	.30	.30
Angoli	.31	.29	.27	.37	.30	.31	.37	.32	.43	.24	.34	.40	.39	.27	.38	.27	.29	.23	.38	.39	.30	.42	.23	.25	.32	.38	.27	.31	.37	.37	.26
Rodrigu	.33	.34	.33	.32	.32	.34	.37	.38	.38	.38	.32	.38	.45	.33	.39	.33	.27	.38	.27	.34	.37	.35	.36	.33	.33	.34	.30	.38	.33	.30	.35
Rêve	.40	.40	.38	.37	.38	.37	.38	.40	.27	.41	.36	.40	.43	.37	.36	.39	.38	.39	.34	.20	.44	.33	.39	.39	.36	.32	.39	.44	.20	.31	.41
Printem.	.28	.28	.28	.31	.28	.28	.36	.30	.44	.28	.32	.37	.40	.27	.40	.25	.28	.30	.37	.44	.18	.40	.26	.22	.31	.39	.23	.18	.41	.37	.25
Siranda	.37	.37	.36	.33	.36	.36	.38	.38	.37	.40	.32	.37	.47	.36	.37	.36	.36	.42	.35	.33	.40	.32	.39	.36	.38	.36	.37	.40	.32	.34	.40
Onitsh	.31	.29	.27	.36	.27	.29	.34	.30	.40	.24	.32	.38	.40	.25	.35	.25	.27	.23	.36	.39	.26	.39	.22	.22	.29	.38	.24	.29	.37	.33	.24
Étoile	.30	.28	.26	.33	.27	.27	.32	.28	.39	.23	.28	.36	.38	.22	.33	.22	.24	.25	.33	.39	.22	.36	.22	.21	.26	.38	.21	.25	.37	.33	.25
Pawana	.33	.31	.30	.34	.30	.31	.37	.33	.40	.32	.31	.40	.45	.30	.36	.32	.26	.32	.33	.36	.31	.38	.29	.26	.26	.38	.27	.34	.35	.32	.32
Diego	.37	.38	.38	.36	.38	.37	.40	.40	.39	.40	.38	.41	.45	.36	.40	.36	.36	.38	.34	.32	.38	.36	.38	.38	.38	.31	.34	.40	.31	.33	.36
Quarant.	.30	.29	.28	.31	.27	.28	.34	.28	.41	.27	.30	.37	.39	.25	.34	.24	.20	.27	.30	.39	.23	.37	.24	.21	.27	.34	.20	.26	.36	.31	.24
Poisson	.28	.30	.29	.33	.30	.29	.38	.31	.46	.29	.35	.38	.40	.31	.42	.28	.30	.31	.38	.44	.18	.40	.29	.25	.34	.40	.26	.18	.42	.39	.26
Fête	.38	.38	.36	.37	.36	.36	.37	.39	.30	.39	.35	.40	.42	.35	.34	.38	.36	.37	.33	.20	.41	.32	.37	.37	.35	.31	.36	.42	.20	.29	.39
Nuages	.34	.34	.33	.35	.32	.34	.35	.34	.35	.36	.31	.37	.43	.30	.34	.33	.30	.37	.30	.31	.37	.34	.33	.33	.32	.33	.31	.39	.29	.29	.34
Hasard	.28	.28	.27	.35	.30	.30	.36	.32	.42	.26	.34	.38	.40	.28	.38	.26	.30	.26	.35	.41	.25	.40	.24	.25	.32	.36	.24	.26	.39	.34	.24

⁴⁹⁸ Le 0 avant la virgule ainsi que les troisièmes décimales fournies par le logiciel ont été ôtés dans la matrice pour des raisons d’économie de place, en arrondissant selon les règles établies.

L'échelle proposée par Dominique Labbé permet d'interroger le tableau et de déterminer la distance entre chaque couple de livres du corpus. Nous pouvons donc constater que la distance entre les "nouveaux romans" est inférieure à 0,15 et que *Le rêve mexicain* et *La fête chantée* (même univers, mais genre différent) ont une distance comprise entre 0,15 et 0,25.

Toutefois, cette lecture n'étant pas très commode, nous avons encore une fois recours à la commodité de la représentation graphique de l'analyse factorielle :

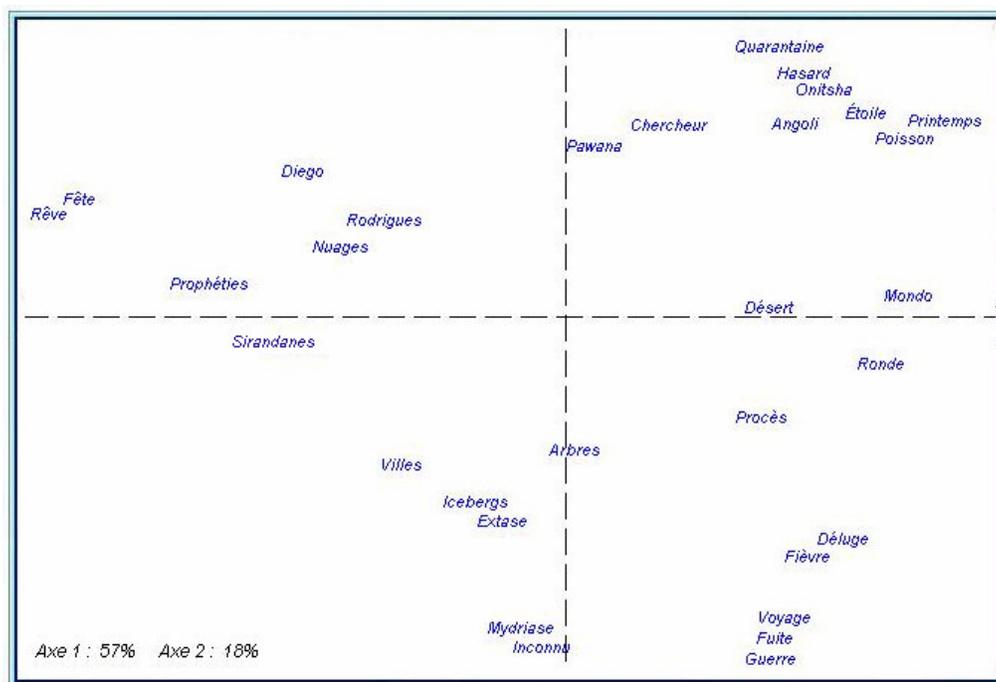


Figure n°150 : Analyse factorielle de la distance lexicale du corpus B. Le calcul est établi sur les fréquences des formes (N).

Il est aisé de constater que, malgré un calcul différent, les résultats sont quasi identiques à ceux calculés sur "V" auparavant. Nous retrouvons les mêmes oppositions et les mêmes unions entre les différents textes. Le premier facteur oppose les genres littéraires et le deuxième, sur l'axe vertical, rend compte de la chronologie de l'œuvre.

Le graphique ci-dessous rend compte des mêmes résultats mais cette fois-ci en s'appuyant sur les lemmes :

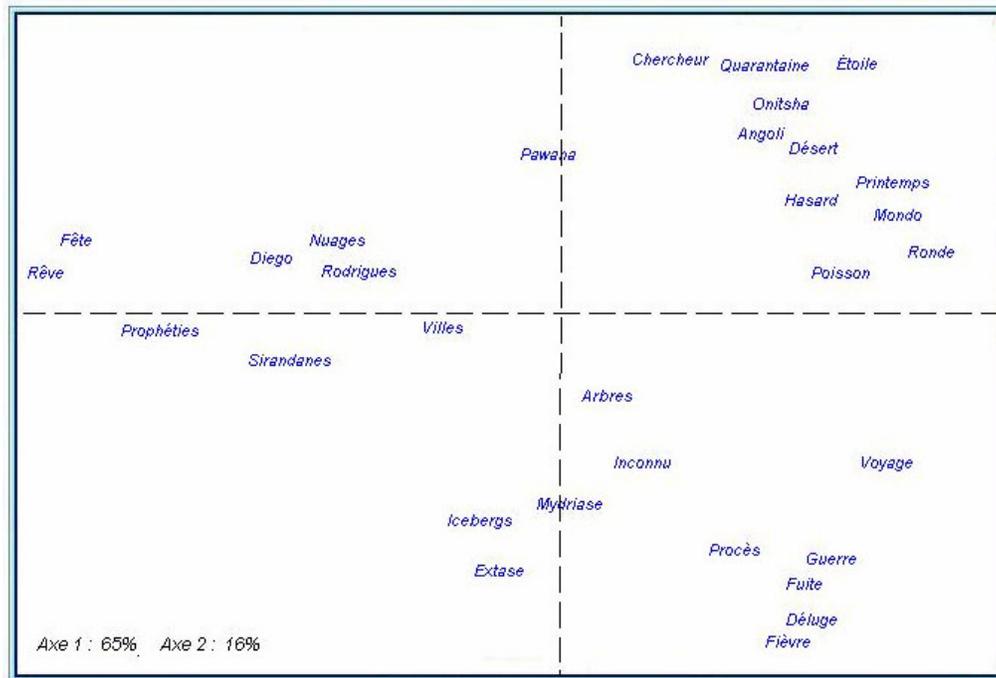


Figure n°151 : Analyse factorielle de la distance lexicale du corpus B. Le calcul est établi sur les fréquences des lemmes (N).

Les graphiques sont encore une fois presque superposables. En effet, il semble que ni la lemmatisation, ni l'introduction des fréquences ait une incidence notable dans les résultats du calcul de la distance lexicale de notre corpus.

Une autre approche à la distance lexicale est celle que propose Dominique Labbé avec son logiciel qui procède à une méthode de classification interne.

5.2.7. La méthode Labbé.

Le logiciel de Dominique Labbé calcule, pour chaque texte étudié, la distance qui le sépare de tous les autres textes pris un à un, suivie par une classification automatique traditionnelle, comme dans les analyses précédentes, et l'on obtient

ainsi une matrice carrée avec en fin de ligne et de colonnes les valeurs moyennes pour chaque texte. Plus la moyenne est faible, plus le texte considéré est central dans le corpus, à l'inverse plus la distance augmente plus le texte est décalé par rapport à l'ensemble. La matrice fournit également la valeur moyenne sur l'ensemble, permettant ainsi de juger de l'homogénéité ou de l'hétérogénéité du corpus.

Dans cette étude, qui repose sur le corpus de 12 œuvres (le corpus H), nous nous appuyons sur les vocables issus de la lemmatisation selon la méthode Labbé.

Après une division visant à homogénéiser la taille des différents sous-corpus, l'analyse englobe au total, 15 œuvres ou groupes de nouvelles⁴⁹⁹ :

Le procès-verbal, *La fièvre*, *Mydriase*, *Voyages de l'autre côté*, *Les bergers*⁵⁰⁰, *Mondo* (la nouvelle seule), *Mondo*, les autres nouvelles (*Lullaby*, *La montagne du dieu vivant*, *La roue d'eau*, *Celui qui n'avait jamais vu la mer*, *Hazaran*, *Peuple du ciel*), *Désert*, *La ronde et autres faits divers*⁵⁰¹ (*La ronde*, *Moloch*, *L'échappé*, *Ariane*, *Villa Aurore*, *Le jeu d'Anne*, *La grande vie*, *Le passeur*, *Ó voleur, voleur, quelle vie est la tienne ?*, *Orlamonde*, *David*), *Voyage à Rodrigues*, *Printemps* (la nouvelle seule), *Printemps et autres saisons*, les autres nouvelles (*Fascination*, *Le temps ne passe pas*, *Zinna*, *La saison des pluies*), *Pawana*, *La quarantaine* et *Hasard*.

La distance moyenne de ce corpus est de 0,309, ce qui correspond à une diversité assez forte, avec un écart-type de 0,043 et un coefficient de dispersion relative de 14,0 %, c'est-à-dire relativement élevé. Le texte le plus central de la série est *Désert* avec un indice de 0,262, le plus décalé de la série est *Voyage à Rodrigues* avec un indice de 0,370.

Les 20 couples les plus proches du corpus sont :

⁴⁹⁹ Les recueils de nouvelles ont été divisés selon la taille des nouvelles.

⁵⁰⁰ Il s'agit de la 8^{ème} et dernière nouvelle du recueil *Mondo et autres histoires*.

⁵⁰¹ Nous avons traité les nouvelles de ce recueil ensemble, leur tailles étant à peu près égales.

<i>Désert</i>	<i>La ronde</i>	0,173
<i>MondoAutres</i>	<i>La ronde</i>	0,181
<i>MondoAutres</i>	<i>Désert</i>	0,184
<i>Le procès-verbal</i>	<i>La fièvre</i>	0,196
<i>Voyages</i>	<i>MondoAutres</i>	0,203
<i>Mondo</i>	<i>La ronde</i>	0,204
<i>Mondo</i>	<i>MondoAutres</i>	0,204
<i>Voyages</i>	<i>La ronde</i>	0,208
<i>Voyages</i>	<i>Désert</i>	0,213
<i>La quarantaine</i>	<i>Hasard</i>	0,221
<i>Bergers</i>	<i>Désert</i>	0,224
<i>Mondo</i>	<i>Désert</i>	0,226
<i>PrintempsAutres</i>	<i>Hasard</i>	0,228
<i>La ronde</i>	<i>Hasard</i>	0,228
<i>Désert</i>	<i>La quarantaine</i>	0,229
<i>Printemps</i>	<i>PrintempsAutres</i>	0,229
<i>Voyages</i>	<i>Mondo</i>	0,230
<i>La ronde</i>	<i>PrintempsAutres</i>	0,230
<i>La ronde</i>	<i>La quarantaine</i>	0,232
<i>PrintempsAutres</i>	<i>La quarantaine</i>	0,232

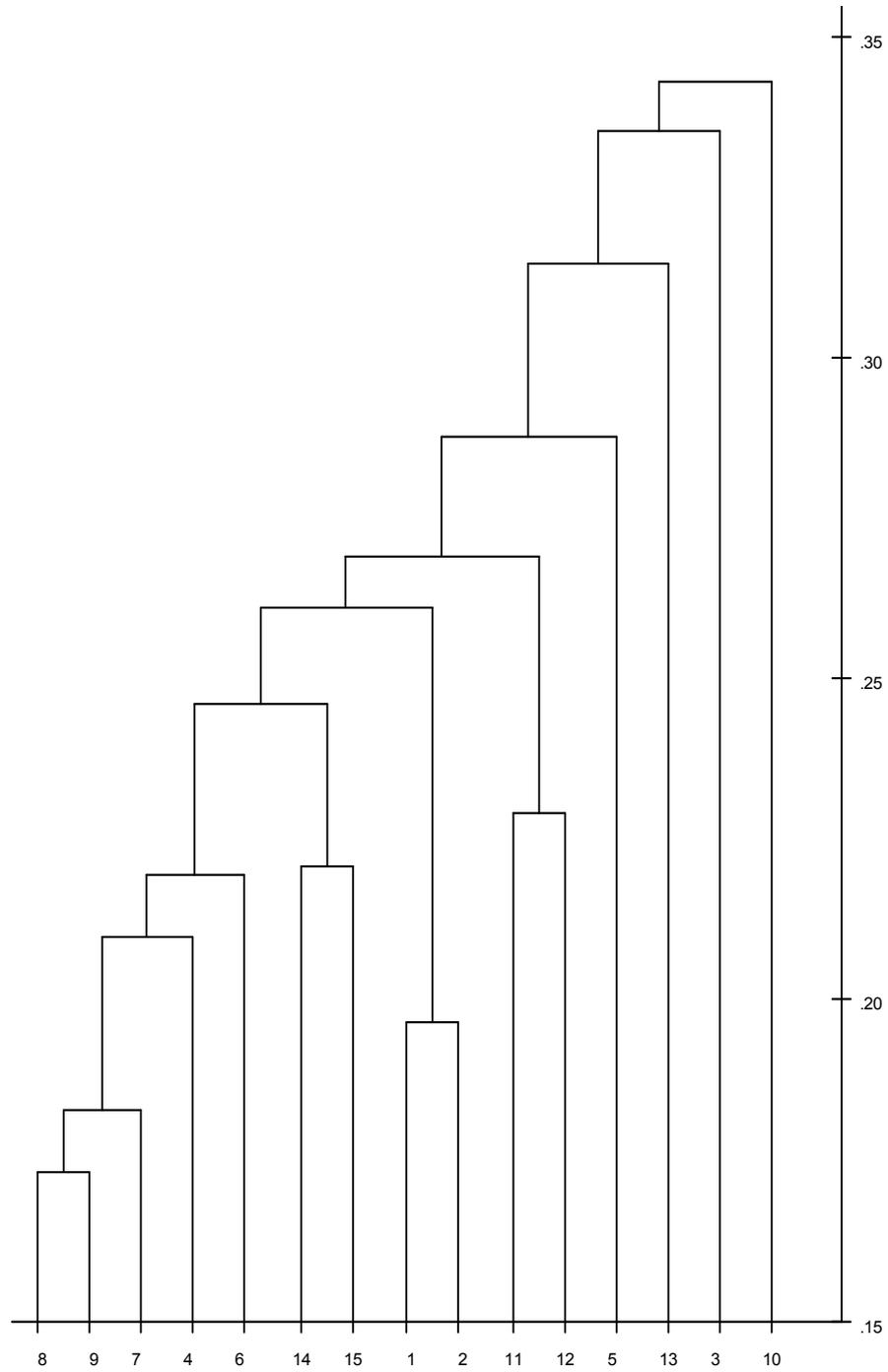
et les 20 couples les plus distants se regroupent ainsi :

<i>Rodrigues</i>	<i>Hasard</i>	0,352
<i>La ronde</i>	<i>Rodrigues</i>	0,353
<i>Mydriase</i>	<i>Pavana</i>	0,356
<i>Mydriase</i>	<i>Bergers</i>	0,358
<i>MondoAutres</i>	<i>Rodrigues</i>	0,358
<i>Le procès-verbal</i>	<i>Mydriase</i>	0,360
<i>Mydriase</i>	<i>Hasard</i>	0,361
<i>Bergers</i>	<i>Pavana</i>	0,362
<i>Printemps</i>	<i>Pavana</i>	0,362
<i>Voyages</i>	<i>Rodrigues</i>	0,365
<i>Rodrigues</i>	<i>PrintempsAutres</i>	0,366
<i>Mondo</i>	<i>Pavana</i>	0,367
<i>Mydriase</i>	<i>PrintempsAutres</i>	0,369
<i>Mydriase</i>	<i>Rodrigues</i>	0,374
<i>Bergers</i>	<i>Printemps</i>	0,379
<i>Mydriase</i>	<i>Mondo</i>	0,379
<i>Bergers</i>	<i>Rodrigues</i>	0,388
<i>Mondo</i>	<i>Rodrigues</i>	0,392
<i>Mydriase</i>	<i>Printemps</i>	0,393
<i>Rodrigues</i>	<i>Printemps</i>	0,403

Le classement des textes en fonction de leur distance moyenne par rapport à tous les autres est le suivant :

N°	Titre	Distance
1	<i>La ronde</i>	0,247
2	<i>Désert</i>	0,253
3	<i>Mondo.Autres</i>	0,256
4	<i>Voyages</i>	0,265
5	<i>La quarantaine</i>	0,266
6	<i>La fièvre</i>	0,272
7	<i>Hasard</i>	0,274
8	<i>Mondo</i>	0,280
9	<i>Printemps.Autres</i>	0,283
10	<i>Le procès-verbal</i>	0,288
	Moyenne	0,289
11	<i>Printemps</i>	0,306
12	<i>Bergers</i>	0,314
13	<i>Pavana</i>	0,330
14	<i>Mydriase</i>	0,349
15	<i>Rodrigues</i>	0,356

L’algorithme opère en réalité sur la matrice des distances, en procédant à la construction d’une classe regroupant les deux textes séparés par la distance la plus faible (dans le corpus le Clézio, il s’agit des deux recueils de nouvelles *Mondo et autres histoires* et *La ronde et autres faits divers*) : ils sont donc les premiers à être fondus en une “classe”, puis le calcul recalcule les distances des autres textes par rapport à ce nouvel ensemble, fusionne les deux textes ou groupes de textes les plus proches, et ceci jusqu’à la constitution d’un ensemble unique. Ces regroupements successifs par la technique du saut minimal sont représentés dans le dendrogramme suivant, où sont portées en ordonnées les distances correspondant aux différents niveaux d’agrégation et en abscisses l’ordre des regroupements :



De gauche à droite :

- | | | |
|-----------------------------------|----------------------------|----------------------|
| 8. <i>Désert</i> | 14. <i>La quarantaine</i> | 13. <i>Pavana</i> |
| 9. <i>La ronde</i> | 15. <i>Hasard</i> | 3. <i>Mydriase</i> |
| 7. <i>MondoAutres</i> | 1. <i>Le procès-verbal</i> | 10. <i>Rodrigues</i> |
| 4. <i>Voyages de l'autre côté</i> | 2. <i>La fièvre</i> | |
| 6. <i>Mondo</i> | 11. <i>Printemps</i> | |
| | 12. <i>PrintempsAutres</i> | |
| | 5. <i>Les bergers</i> | |

En coupant le graphique horizontalement, nous pouvons isoler les groupes de textes très proches, relativement proches, etc. Il convient de noter que, plus l'on s'élève dans le graphe, plus les classes constituées sont hétérogènes et plus l'interprétation des différences devient complexe.

Dans ce type de graphe, les textes les plus centraux se trouvent généralement dans la moitié gauche du graphe (ce sont les premiers à être agrégés ensemble). Plus le texte est situé à droite dans le dendrogramme, plus il a de chances d'être "décalé" par rapport au reste du corpus.

Ceci semble également être vrai pour notre corpus. Nous trouvons à gauche un groupe d'œuvres très proches (distance < 0.25), dans lequel il est intéressant de noter l'extraordinaire proximité des nouvelles de *La ronde* avec *Désert*, puis avec les autres nouvelles de *Mondo* (hors *Les bergers* et *Mondo*) et *Voyages de l'autre côté*. Il s'agit en effet d'une proximité chronologique aussi bien que thématique. Ces ouvrages semblent vraiment constituer le cœur de ce corpus réduit. Ensuite se situent *La quarantaine* et *Hasard*, très proches du noyau. Il y a certes une évolution mais elle n'est pas très grande, ces deux œuvres semblant bien appartenir au même "monde" lexical que les précédentes. Toutefois, le classement gauche-droite peut parfois s'avérer trompeur : le couple constitué par *La fièvre* et *Le procès verbal* est plus proche du groupe *Voyages de l'autre côté* – *La ronde* que du couple *La quarantaine* – *Hasard*. Les deux premiers romans sont en effet très proches mais plus décalés par rapport au noyau. La rupture que nous avons constatée auparavant entre ces deux livres et la période 1975-1982 semble donc un peu plus nette. Notons au passage que dans l'étude formelle effectuée auparavant *Voyages de l'autre côté* était le plus souvent près des ouvrages "nouveau roman", alors qu'ici il se trouve proche de *Désert*, *La ronde* et *Mondo et autres nouvelles* qui suivent la rupture, ce qui pourrait souligner son caractère transitoire vers une nouvelle écriture où les changements thématiques précèdent ceux de caractères plus formels et stylistiques. La singularité relative des nouvelles recueillies dans *Printemps et autres saisons* apparaît clairement, *Le poisson d'or*, très proche lexicalement, étant absent de ce corpus. Enfin, dans une irréductible singularité se trouvent *Les bergers*, *Pawana*, *Mydriase* et

Voyage à Rodrigues qui s'éloignent thématiquement du reste du corpus en reflétant également l'ordre chronologique de l'écriture.

Il convient toutefois de remarquer l'inconvénient de la classification automatique traditionnelle dans cette analyse. En effet, quand il n'existe pas de sous-groupes bien individualisés, l'algorithme, après avoir réuni quelques paires bien individualisées, agrège les textes un par un, générant ainsi une structure en "escalier" qui interdit de découvrir des structures fines dans les données. La technique employée (le "saut minimal") semble également produire parfois des "effets de chaîne". Certaines proximités entre textes ne sont pas discernables car les sommets qui les relient sont effacés par des agrégations effectuées à un niveau inférieur.

Il ne faut donc pas se fier aveuglément à cette technique. L'appartenance de chacun des textes à un regroupement donné doit également être contrôlée par d'autres analyses et représentations graphiques. La technique de la représentation par l'analyse arborée semble limiter cet inconvénient et convenir pour combler l'analyse de la distance lexicale de notre corpus.

5.2.8. L'analyse arborée.

La technique de l'analyse arborée élaborée par Xuan Luong⁵⁰² permet de représenter les résultats du calcul de la distance lexicale d'une façon différente. L'algorithme produit des graphes qui rendent compte de la proximité – ou de l'éloignement – des textes étudiés en une seule représentation graphique, sous forme de dendrogramme ou bien sous forme radiale.

⁵⁰² Pour le détails des calculs de topologie cf., N. X. Luong, (éd.), 1989. et J.-P. Barthelemy, X. Luong, "Représenter les données textuelles par les arbres...", in *4e journées internationales d'analyse statistique des données textuelles*, S. Mellet (éd.), Université de Nice, 1998, p. 49-71.
N.X. Luong, "L'analyse arborée des données textuelles : mode d'emploi.", in *Travaux du cercle linguistique de Nice* n° 16, Université de Nice, 1994, p.27-42..

Le modèle de l'arbre est un "graphe connexe et sans circuit"⁵⁰³ et il est caractérisé par l'ensemble des distances entre ses éléments, la longueur des "branches" représentant fidèlement la distance entre les textes. Par ailleurs, sa structure est aussi importante, faisant apparaître l'ordre et la force des regroupements ou des oppositions entre les différents éléments.

"L'avantage de cette technique par rapport à l'analyse factorielle, écrit Étienne Brunet⁵⁰⁴, est qu'on n'a plus à distinguer et à croiser des facteurs, dont chacun n'explique qu'une partie de la variance."

L'analyse arborée permet également, grâce à la représentation par "branches", d'éviter les inconvénients du "saut minimal", la technique employée dans les dendrogrammes⁵⁰⁵.

Dans notre étude nous nous intéresserons à la totalité de l'œuvre, en ayant recours au corpus B et en nous appuyant sur les formes graphiques aussi bien que sur les lemmes afin de représenter – sous la forme radiale de l'analyse arborée - les résultats du calcul de la distance calculés sur V ainsi que sur N.

Afin de pallier les inconvénients dus à une importance prépondérante de la grande variabilité des éléments, nous avons préalablement calculé les distances de χ^2 (chi2) qui possèdent notamment la propriété d'équivalence distributionnelle⁵⁰⁶.

L'arbre ci-dessous représente la distance lexicale des formes graphiques du corpus B, en s'appuyant sur le calcul absence/présence (le calcul sur V) :

⁵⁰³ N.X. Luong (1994) : p. 29.

⁵⁰⁴ É. Brunet (2000) : p.11.

⁵⁰⁵ Cf. p. 458.

⁵⁰⁶ Pour plus de détails sur ce calcul cf. S. Körner, L. Wahlgren (1998) : p. 147-156.

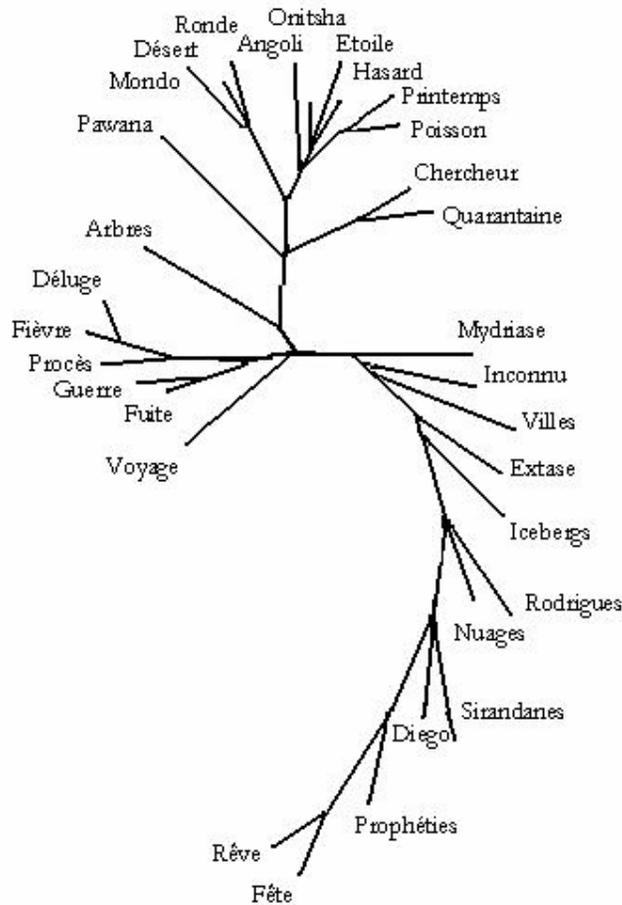


Figure n°152 : Analyse arborée de la distance lexicale s'appuyant sur les formes graphiques sans considération de la fréquence (V).

L'arbre permet de constater plusieurs regroupements de textes et rend compte de la chronologie de l'œuvre : les premiers livres du genre "nouveau roman" se trouvent sur les rameaux de gauche attachés à une même branche ; en haut de l'arbre, sur une autre branche fondamentale, nous trouvons les ouvrages romanesques plus traditionnels et ceux de la deuxième période – *Mondo*, *Désert* et *La ronde* – attachés à une même branche. Du côté droit de l'arbre est à remarquer la structure "en escalier" qui semble être surtout régie par le genre ; dans le sens descendant nous trouvons d'abord les ouvrages à vocation poétique, puis les essais, et tout à fait en bas de l'arbre les ouvrages ethnologiques : *Les prophéties de Chilam Balam*, *Le rêve mexicain* et *La fête chantée*. Les regroupements sont rares et peu

pertinents dans cette troisième partie de l'arbre, à l'exception de celui qui rapproche *Le rêve mexicain* et *La fête chantée*.

Les divers regroupements du corpus sont toutefois plus faciles à distinguer dans l'analyse suivante, qui repose sur le calcul prenant en compte la fréquence (le calcul sur N) :

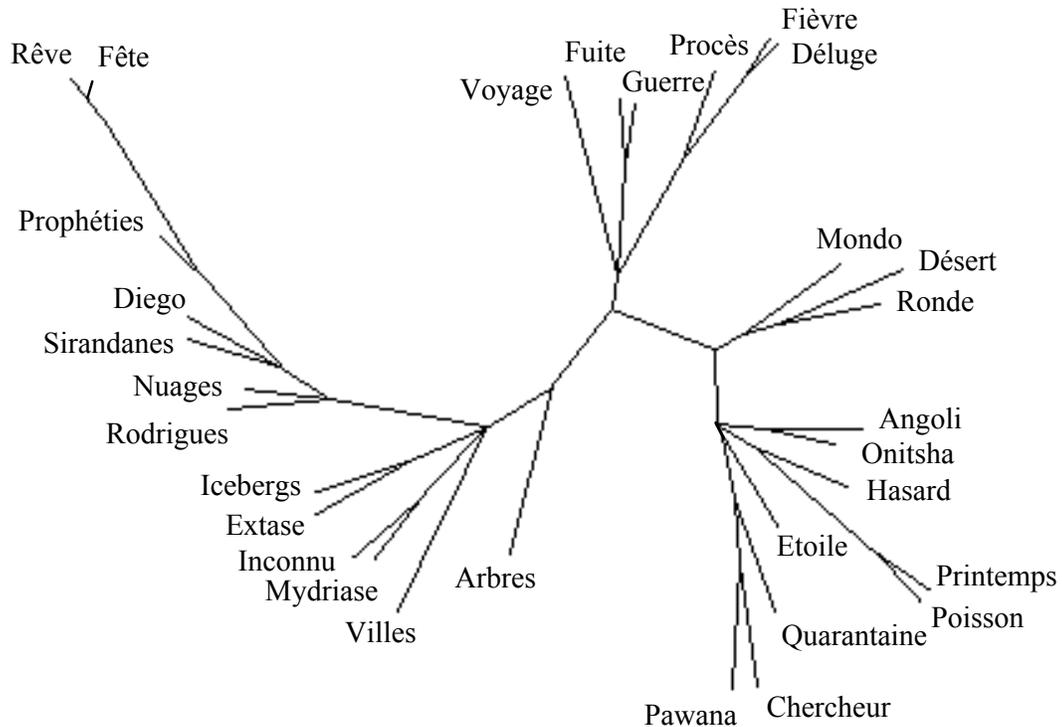


Figure n°153 : Analyse arborée de la distance lexicale s'appuyant sur les formes graphiques prenant en considération la fréquence (N).

A la cime de l'arbre sur la même branche maîtresse se trouve, à l'écart des autres, la ramification correspondant au début de l'œuvre, avec les ouvrages inspirés de l'école "nouveau roman". Ces livres se distinguent nettement des autres et l'analyse rend compte de leur proximité thématique. Avec *Voyages de l'autre côté* de 1975 qui est le dernier livre de cette période et qui se trouve à l'extrémité de la

ramification, Le Clézio ouvre la voie vers un changement thématique de l'écriture et l'arbre le montre bien. Germaine Brée écrit ceci à ce propos⁵⁰⁷ :

“*Voyages de l'autre côté* semble avoir libéré Le Clézio de l'angoisse née de la confrontation avec la technologie moderne et le vaste arsenal des stratégies subliminales pour contrôler et conditionner la société humaine.”

Le Clézio écrit en effet à cette époque *Mondo et autres histoires* et *L'inconnu sur la terre*, qui constituent une sorte de plaque tournante, étape de la pensée et de l'écriture lecléziennes. “Il faut quitter le monde établi, le monde compliqué et sérieux des adultes” écrit-il dans la préface de *L'inconnu sur la terre*, et l'analyse arborée confirme à la fois son propos et le changement important de cette période. Il est à noter que cet ouvrage n'est pas du tout sur la même branche que *Mondo et autres histoires* et que les deux livres semblent ouvrir chacun une voie différente d'évolution.

En effet, c'est à ce branchement de l'arbre que vont se ramifier deux autres branches maîtresses : d'un côté l'œuvre romanesque et de l'autre les autres genres littéraires.

Les ramifications qui portent l'œuvre romanesque distinguent sur la première branche, à l'écart des autres romans, *Mondo et autres histoires*, *Désert* et *La ronde*, qui constituent la deuxième période dans l'œuvre romanesque leclézienne. Ces ouvrages, que nous avons pu distinguer des autres dans les analyses formelles du vocabulaire comme dans celle de l'accroissement lexical, s'écartent en effet thématiquement des précédentes et marquent ainsi la rupture de l'œuvre également par ce changement. Le Clézio nous introduit à cette époque dans un nouveau monde. Il quitte ici le monde urbain avec ses autos, ses grands magasins

⁵⁰⁷ G. Brée (1990) : p. 97.

et ses rues pour s'intéresser à un univers plus magique, plus lumineux, plus beau, celui des enfants.

Angoli Mala et *Onitsba* ont en commun l'aventure d'un jeune garçon, l'un à l'intérieur de l'Afrique noire et l'autre dans les forêts du Panama.

Printemps et autres saisons et *Poisson d'or* sont des livres où l'action se déroule autour des adolescents du monde occidental et qui privilégient le dialogue et la langue des jeunes ; ils se trouvent sur la même branche de notre arbre.

La quarantaine, *Le chercheur d'or* et *Pawana* sont également des livres où le personnage principal est un jeune garçon qui vit une aventure ; cependant, à la différence de ceux cités auparavant, il ne s'agit plus de découvrir un pays exotique mais de connaître le monde marin, celui de la navigation, des explorateurs et des îles.

Le roman *Etoile errante*, tout en se trouvant au milieu de ce regroupement, est porté par une branche solitaire. C'est en fait le seul livre du corpus à traiter – à travers l'aventure de la jeune Esther – des aspects de la deuxième guerre mondiale, de l'histoire juive et de l'exode vers le nouvel état d'Israël.

Si nous remontons le côté gauche de l'arbre, le côté non romanesque, nous trouvons, écarté des autres, *Le voyage au pays des arbres* qui est le seul livre spécifiquement pour enfants de notre corpus.

Les différentes ramifications semblent en effet surtout sensibles au genre, qui les divise assez nettement. Les essais littéraires se regroupent avec les ouvrages difficilement catégorisables que sont *Mydriase* et *Vers les Icebergs*. Par la suite nous trouvons sur la même branche le récit de voyage *Gens de voyage*, avec le seul roman qui se trouve de ce côté de l'arbre, *Voyage à Rodrigues*, qui se singularise non seulement dans les analyses purement statistiques mais également par son contenu lexical, où il semble s'approcher davantage d'un récit de voyage que d'un roman.

L'explication de la proximité de la biographie *Diego et Frida* et de *Sirandanes* est plus délicate ; il convient de prendre en compte la taille de *Sirandanes* qui ne compte que 20 pages et ne permet pas de tirer des conclusions solides.

Au bout de cette branche maîtresse nous trouvons le monde amérindien avec *Les prophéties de Chilam Balam*, *Le rêve mexicain* et *La fête chantée*, tout comme nous l'avons remarqué dans toutes les analyses sur la distance lexicale.

Afin de poursuivre notre exploration méthodologique nous avons procédé à ces deux mêmes analyses (les calculs sur V et sur N), mais cette fois-ci l'étude se basera non plus sur les formes graphiques mais sur les lemmes. L'arbre ci-dessous présente l'analyse selon la méthode Jaccard (calcul sur "V") s'appuyant sur les lemmes du corpus B :

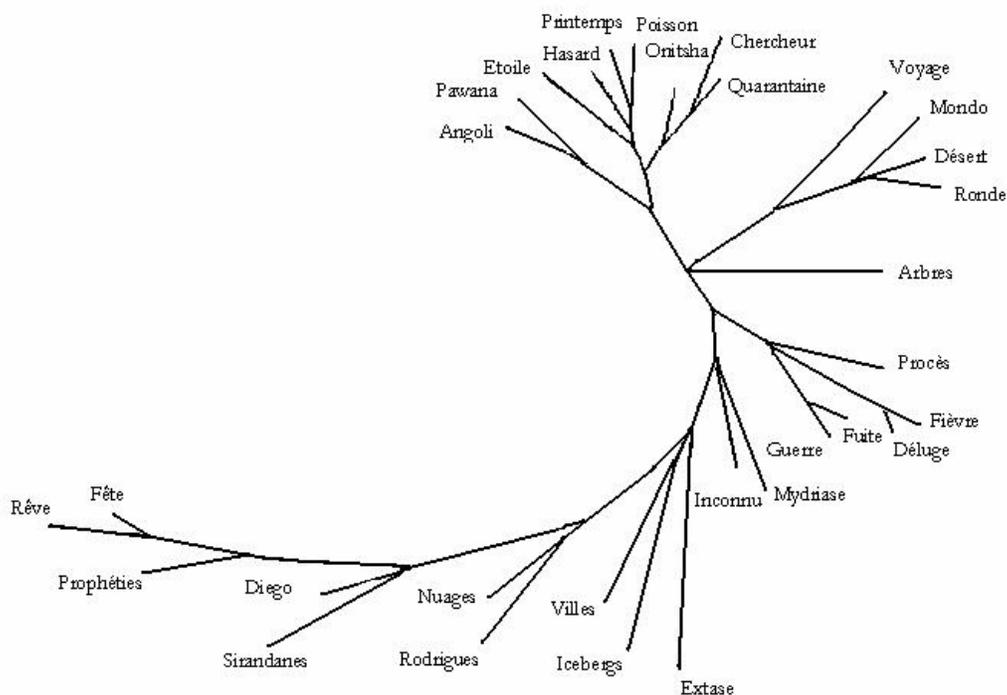


Figure n°154 : Analyse arborée de la distance lexicale s'appuyant sur les lemmes sans considération de la fréquence (V).

La lemmatisation, qui modifie surtout les nombreuses formes de flexion, ne change pas la sémantique d'un texte et par conséquent cet arbre ressemble fort à celui qui reposait sur l'analyse des formes graphiques⁵⁰⁸ ; il n'y a qu'à deux endroits que nous pouvons observer des différences. La position d'*Angoli Mala* est légèrement décalée par rapport à l'autre arbre bien qu'il appartienne toujours au même regroupement. Le décalage le plus notable est celui de *Voyages de l'autre côté*, qui sur cet arbre semble déjà avoir franchi le pas vers la rupture thématique de l'œuvre et se trouve près de *Mondo*, *Désert* et *La ronde*, au lieu de se compter parmi les ouvrages inspirés du "nouveau roman" ; il contredit ainsi quelque peu la précédente analyse. Son décalage semble en effet témoigner d'un effet stylistique précurseur à la rupture de l'œuvre.

Il convient, avant de tirer des conclusions, de procéder à la même comparaison des deux méthodes en s'appuyant sur le calcul qui prend en compte la fréquence (le calcul sur N) :

⁵⁰⁸ Cf. figure n°152, p. 460.

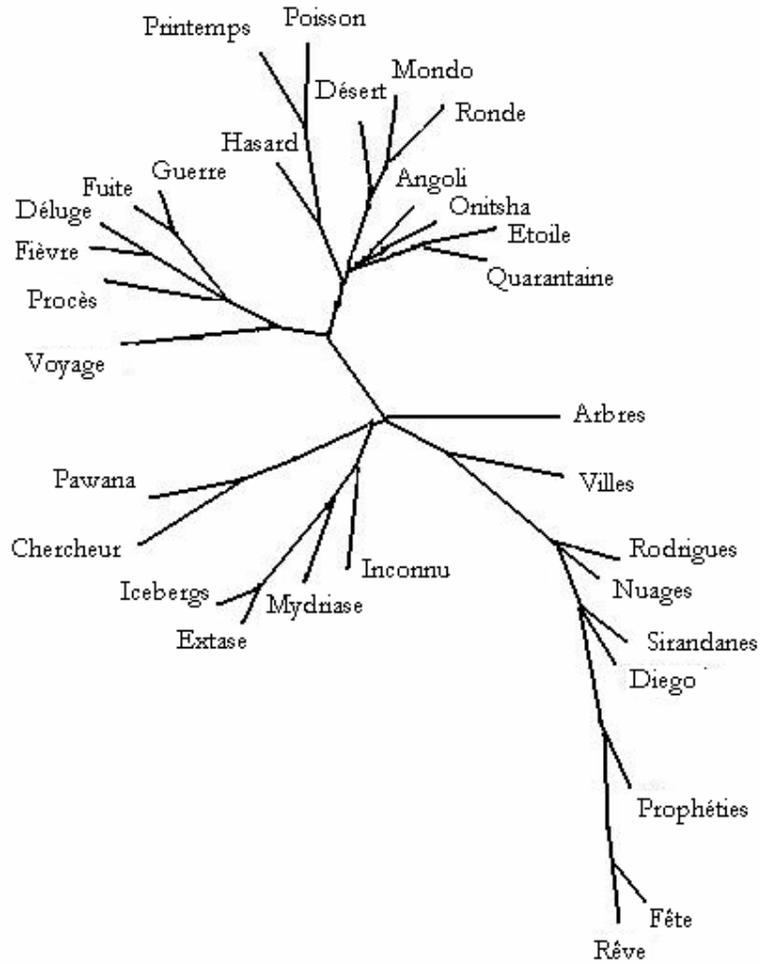


Figure n°155 : Analyse arborée de la distance lexicale s'appuyant sur les lemmes prenant en considération la fréquence (N).

Le petit décalage constaté dans l'analyse précédente est cette fois-ci quasi nul, si l'on prend garde à ne pas mesurer les distances à "vol d'oiseau", et les regroupements sont quasi identiques. Il semble donc que la lemmatisation ait moins de conséquences lorsque l'on se base sur une étude prenant en considération la fréquence que sur une étude reposant sur l'absence ou la présence d'un mot donné.

Toutefois, les quatre arbres sont très semblables et les résultats pratiquement les mêmes. L'importance donnée, d'un côté aux hapax et mots rares, de l'autre aux

mots plus fréquents, ne semble pas modifier les résultats de façon significative. Les résultats restent sensiblement égaux et permettent une conclusion assurée de l'analyse de la distance lexicale dans l'œuvre de Le Clézio.

5.2.9. Conclusion de l'étude de la distance lexicale.

En effet, des résultats qui se dégagent des différentes méthodes et techniques que nous avons utilisées dans cette étude – le recours aux formes graphiques ou aux lemmes véritables, avec ou sans considération de fréquences, et l'application des différents logiciels pour le calcul des distances et pour leur représentation graphique – se dessinent les mêmes lignes de forces et les mêmes influences s'exerçant dans les différents textes.

Les différentes analyses ont permis de répondre aux questions que nous nous sommes posées avant d'effectuer cette étude sur le contenu lexical : nous avons pu faire à peu près les mêmes observations précédemment lorsque nous nous intéressions uniquement aux aspects formels et purement quantitatifs du vocabulaire de notre auteur.

Nous avons trouvé, comme auparavant – dans les analyses structurelles et stylistiques du corpus –, des oppositions fortes entre les différents genres littéraires et un regroupement des livres appartenant à une même variété générique. La division tripartite à l'intérieur du genre romanesque écarte les ouvrages initiaux inspirés de l'école "nouveau roman" des autres – en indiquant que le changement thématique intervient déjà avec *Voyages à l'autre côté* – pour ensuite distinguer les autres ouvrages romanesques en un deuxième et un troisième regroupements du corpus qui tiennent bien compte de la chronologie de l'œuvre et de son évolution.

Nous allons maintenant, après cette étude de distance lexicale, nous intéresser plus précisément au sens des mots ou des thématiques qui constituent les intérêts majeurs dans l'analyse d'un corpus littéraire et dans l'étude du vocabulaire d'un écrivain.

5.3. Les spécificités de Le Clézio.

5.3.1. Introduction.

“Ecrire seulement sur les choses qu'on aime. Ecrire pour lier ensemble, pour rassembler les morceaux de la beauté. Alors les arbres qui sont dans les mots, les rochers, l'eau, les étincelles de lumière qui sont dans les mots, ils s'allument, ils brillent à nouveau, ils sont purs, ils s'élancent, ils dansent !”

L'inconnu sur la terre, page 10.

“Ailleurs. C'est avec le voyage le maître-mot de l'œuvre et de la vie de Jean-Marie Gustave Le Clézio” écrit Marc Dupuis dans *Le monde de l'éducation*⁵⁰⁹. L'écrivain passe en effet pour l'écrivain-voyageur par excellence ; aimantée par le désert, la mer et les pays sauvages, sa littérature serait celle des errances et des mythologies. Peut-on le vérifier par les chiffres? Peut-on, avec la statistique, dévoiler les mots et les thèmes préférés de Le Clézio ?

Les mots que nous sentons comme typiques de l'écriture leclézienne, les reconnaissons-nous à leur fréquence ou à la signification particulière que Le Clézio leur prête ? Selon Étienne Brunet⁵¹⁰ :

“...il faut croire que les deux faits sont liés. [...] et cette liaison entre la fréquence, la préférence et l'importance constitue le postulat fondamental de la statistique lexicale.”

Nous nous intéressons dans ce chapitre à la spécificité de l'écriture leclézienne à partir de deux perspectives : d'abord la perspective exogène, c'est-à-dire Le Clézio comparé aux autres, et ensuite la perspective endogène, les variations à l'intérieur du corpus.

⁵⁰⁹ *Le Monde de l'éducation de la culture et de la formation*, mensuel n°257, mars 1998, p. 27

⁵¹⁰ É. Brunet (1988) : p. 426-427.

5.3.2. Les spécificités externes.

Il va de soi qu'une fréquence ne saurait devenir "caractéristique" que comparée à une fréquence théorique, donc par référence à un texte ou un corpus plus étendu que celui qui est en question et cet ensemble plus grand est alors pris comme source du modèle théorique.

Pour connaître, de façon impartiale, les mots spécifiques et caractéristiques de Le Clézio, nous nous utiliserons comme point de comparaison dans cette étude *Frantext* (qui s'appuie sur les fréquences du Trésor de la langue française (T.L.F.) avec ses 86 millions d'occurrences), et plus précisément le corpus littéraire du XX^{ème} siècle. Dans Hyperbase, en effet, ces données sont insérées pour servir de norme et servent de base de calcul en indiquant la différence entre deux grandeurs, celle des fréquences dans le corpus Le Clézio et celle de *Frantext*. Les valeurs obtenues mettent en relief les excédents et les déficits du vocabulaire leclézien par rapport à celui de *Frantext*.

“Selon Charles Muller⁵¹¹, la méthode consiste à, pour un fragment d'un texte, calculer l'écart réduit de chacun des vocables du texte par rapport à la sous-fréquence théorique, et à les classer en fonction de cet écart réduit. On obtiendra ainsi, en tête de liste, le vocabulaire caractéristique positif du fragment, l'ensemble des vocables dont la sous-fréquence est plus élevée que la fréquence dans le texte ne le fait prévoir ; en fin de liste, le vocabulaire caractéristique négatif.”

Il convient de rappeler à ce sujet que la comparaison avec l'usage observé dans le *Trésor de la langue française* doit être interprétée prudemment.

⁵¹¹ Ch. Muller (1968) : p. 204.

“D’une part, écrit Étienne Brunet⁵¹², le T.L.F. reflète l’usage littéraire de la langue, dans un registre élevé, et si le corpus qu’on traite se trouve éloigné de ce niveau de langue, la valeur de la comparaison en est amoindrie. D’autre part toutes les formes n’ont pas été soumises à la comparaison, parce que le calcul de l’écart réduit perd de sa légitimité quand la fréquence théorique est trop faible, ce qui dépend certes de la taille du corpus traité, mais aussi de la fréquence du mot en question. Cela permet, en fixant le point de départ et le point d’arrivée, de rapprocher les deux corpus comparés et de justifier ainsi leur confrontation. Mais il peut se faire que des mots soient très significatifs d’un corpus donné, même s’ils ne figurent pas dans la liste des 100.000 [mots retenus], soit parce qu’ils sont trop récents, soit parce qu’ils sont trop techniques.”

Prenant ces considérations en compte, le traitement informatique nous permet d’extraire le vocabulaire spécifique positif et négatif de notre corpus, c’est-à-dire les mots particulièrement suremployés et sous-employés par rapport à la base de référence afin de nous donner une idée précise des thèmes traités et non traités.

Nous nous intéressons en premier lieu aux mots significativement suremployés par rapport à notre référence externe, *Frantext* du XX^{ème} siècle.

Quels sont les mots caractéristiques de notre écrivain ? Afin de répondre à cette question nous avons établi une liste des mots les plus excédentaires du corpus A. Le regroupement en catégories grammaticales permet une meilleure lisibilité des résultats. Rappelons que le seuil significatif généralement admis est aux alentours de la valeur 2 (en laissant 5 chances sur 100 au hasard). Nous nous sommes arrêtée à la valeur 3 en adoptant un seuil plus sévère.

⁵¹² É. Brunet (1995) : p. 199.

		22.00	8980	1015	vite	6.61	3047	264	arriver		
Vocabulaire spécifique du corpus (excédents)		16.57	3456	431	autrefois	6.47	3351	284	courant		
		14.59	4027	453	doucement	6.46	6872	528	savait		
		14.51	2472	314	dela	6.30	21592	1481	va		
		14.21	4153	458	tellement	6.12	5260	412	arrive		
		13.23	19555	1579	ici	6.08	2753	236	boire		
Ecart TLF Corpus Mot		13.12	5170	524	dehors	5.94	8923	654	vient		
		13.00	10157	902	longtemps	5.87	3566	291	tomber		
Grammaticaux		12.64	15067	1246	près	5.69	11928	844	aller		
		11.51	7812	696	vraiment	5.57	15907	1096	faisait		
Prépositions		9.07	31798	2241	alors	5.56	4252	334	restait		
		8.04	42317	2865	aussi	5.47	12590	881	parler		
134.78 5755 2737 jusqu'		6.35	7828	590	parfois	5.37	13176	916	pouvait		
60.20 15630914736 sur		5.62	4210	332	simplement	5.37	6951	512	venir		
49.60 31105 3874 vers						5.37	3320	267	attendait		
46.65 25808620669 dans						5.02	3667	286	faisaient		
28.18 11059 1343 autour	Verbes					4.58	3258	252	voulais		
21.99 26077 2360 devant	55.41	417	290	brille	4.23	3148	240	commence			
21.62 75370 5805 sans	51.64	44055	5123	sont	4.13	3949	292	entrer			
8.77 21832 1582 contre	45.34	1035	403	bouger	4.06	3601	268	pensait			
8.70 128127 8230 avec	43.21	16668213897	a		3.93	3993	292	rester			
8.02 22065 1571 avant	42.51	433	233	brillait	3.84	7264	502	pense			
7.40 11631 868 derrière	40.12	853	325	brûle	3.73	3824	278	passait			
4.87 6933 501 malgré	34.63	25599	2799	étaient	3.71	9357	632	sorte			
	34.34	7073	1092	regarde	3.50	7024	480	pouvoir			
grammaticaux	30.03	2371	482	marcher	3.05	7740	516	peux			
	28.89	4849	756	marche							
123.22 49276949146 les	28.52	897	253	glisse							
94.64 32374531590 des	27.78	14870711219	était								
65.88 428 345 el	26.27	971	249	traverse	Part. Adjectifs						
59.83 80153559489 la	26.08	5004	726	regarder							
57.65 26818 3786 leurs	25.74	124605	9426	avait	42.12	2515	643	noirs			
54.79 23583420049 elle	22.65	1463	289	marchait	41.43	3903	836	bleu			
50.84 89300 8793 ils	22.11	3807	543	vont	40.45	6972	1201	noir			
48.56 21196017654 du	21.35	4756	624	regardait	40.11	3552	769	noire			
42.88 15172 2128 chaque	20.23	28930	2501	ont	38.55	4769	904	vide			
41.30 12626310835 c'	18.33	1332	235	portant	38.40	1472	432	noires			
37.87 29552322130 qui	18.53	7016	775	voit	37.52	515	230	lisse			
37.38 51596 5013 quand	17.74	2544	359	entend	35.59	742	271	plate			
35.96 562 233 j	15.93	44793	3413	peut	34.30	3041	622	sombre			
27.99 23016 2344 elles	15.60	4699	526	parlait	31.09	5490	862	blanc			
27.91 35759 3332 cela	15.53	2536	332	monte	30.86	2877	557	gris			
27.23 67340344660 le	15.42	1567	235	retourner	29.28	1399	339	nus			
26.05 109016 8400 y	15.00	17863	1516	avaient	28.99	10230	1287	haut			
23.40 17624012621 au	14.88	2235	296	regardant	28.36	6548	922	rouge			
21.51 90585 6822 ses	14.43	3394	396	faisant	27.98	1409	329	disparu			
21.00 1231 245 b	14.41	1883	257	souviens	27.03	2213	428	rouges			
19.67 70102644898 et	13.68	2217	281	viennent	25.67	4340	651	assis			
19.63 72689 5497 où	13.45	2213	278	allaient	25.58	1410	308	froide			
17.06 1519 245 h	13.22	2317	285	arrêter	24.80	2210	403	commencé			
13.70 6199 616 tandis	12.75	7155	672	parle	23.96	2246	398	blancs			
12.45 18553 1484 eux	12.68	3168	353	sent	23.90	1139	256	grise			
11.88 56953 3999 aux	12.38	2793	317	cherche	21.94	2005	348	marché			
11.66 64607 4477 ou	12.29	2278	271	oublier	21.45	1413	272	venus			
11.34127976877917 de	12.04	2039	247	courir	20.23	2297	362	jaune			
11.20 45594 3230 autre	11.91	2778	310	dormir	20.03	4587	587	blanches			
10.48 2544 273 p	11.31	3949	398	manger	18.89	1697	282	blanches			
10.59 51679 3590 rien	11.07	2414	269	attend	18.70	1749	286	assise			
8.71 27920 1976 autres	11.02	6808	612	passé	18.46	1391	243	bleue			
5.41 23703 1583 toutes	10.99	10642	889	allait	18.16	1396	241	hautes			
6.75 143578 9004 son	10.94	6723	604	voulait	18.10	8363	878	grands			
5.96 21446 1460 ni	10.70	34284121537	est		17.74	2524	357	drôle			
5.75 3134 259 uns	10.66	2858	301	entendait	16.86	3500	439	immense			
3.22 18294811031 s'	9.89	3408	335	regarda	16.79	2890	381	doux			
	9.89	2654	275	appelait	16.79	2858	378	douce			
	9.86	4464	416	appelle	16.64	1499	239	regardé			
Adverbes	9.78	6054	533	comprendre	16.58	3217	409	étrange			
	9.51	5073	456	vit	16.21	2432	330	longues			
57.78 3821 1062 lentement	9.23	2367	244	devient	15.70	2062	288	dure			
38.63 14119 1904 loin	8.93	6159	525	font	14.92	1973	271	longs			
34.74 13668011015 comme	8.20	2635	253	arrête	14.71	2276	298	pâle			
26.69 5179 754 partout	7.90	5324	447	partir	14.41	2582	323	chaud			
25.66 18018 1863 maintenant	7.50	5992	487	voyait	14.06	1714	237	inconnu			
25.18 12437 1387 dessus	7.41	3956	341	avance	13.50	26647	2077	grand			
23.58 2938 472 quelquefois	7.23	9017	689	sait	13.04	4737	488	longue			
22.53 34354 2991 très	7.21	3335	293	entends	12.89	4730	485	écrit			
22.38 36888 3168 puis	7.00	3333	290	écoute	12.65	2673	310	devenu			

12.38	2573	298	pareil	40.17	4041	836	sol	18.31	4054	511	jambes
12.19	2820	317	arrivée	39.92	6202	1101	cheveux	17.92	13711	1295	personne
11.31	2370	268	vivant	39.86	2040	542	rivière	17.69	4025	499	rêve
11.16	14988	1198	grande	39.72	1042	362	navire	17.44	20843	1811	voix
9.86	3096	310	anglais	39.59	13877	1907	côté	17.11	3702	461	beauté
9.06	16142	1215	jeune	39.51	943	340	nuage	16.84	9015	903	instant
8.92	13256	1017	bas	37.61	3364	709	pluie	16.74	2471	340	forêt
8.79	2444	245	vert	36.51	5053	905	froid	16.40	21880	1850	air
8.40	4930	427	grandes	35.38	3245	663	souffle	16.38	1941	283	couleurs
7.04	3190	280	terrible	35.13	20861	2412	nuit	16.16	15277	1363	maison
6.44	11069	807	fort	34.78	1992	481	noms	15.81	3296	406	ventre
5.18	3243	259	arrivé	34.66	3028	625	chaleur	15.73	1511	232	toit
5.12	11563	806	première	34.43	3635	700	maisons	15.38	2758	351	langue
5.09	5733	426	venu	34.14	1294	364	villes	15.22	1633	240	ailes
4.84	3264	256	large	33.97	30000	3137	yeux	15.19	1804	257	payage
4.59	2947	231	dur	33.76	2995	609	feuilles	15.10	2452	319	sac
4.56	6290	453	droite	33.65	2191	498	montagne	14.85	7363	730	feu
4.52	4251	318	calme	33.36	11228	1487	regard	14.20	6637	660	train
4.46	17184	1143	mort	32.20	1827	430	fumée	14.18	1804	247	aube
4.22	7053	496	vieille	32.10	1800	425	sud	14.13	5009	528	dos
4.13	23936	1551	quelque	31.89	8251	1163	travers	14.09	37706	2849	temps
3.35	3188	231	venue	31.65	4318	741	odeur	13.92	3278	379	entrée
3.30	3844	273	morts	31.40	2411	503	immobile	13.82	5544	566	coups
				31.37	2672	537	cris	13.43	2679	320	herbe
				31.33	10297	1349	enfants	13.08	4842	497	mur
				31.21	881	269	trous	12.65	3578	387	fer
				30.85	1824	416	visages	12.36	2002	247	histoires
				30.77	25166	2619	monde	12.25	5497	535	mouvement
				30.11	913	267	gouttes	11.90	2270	266	côte
				30.09	908	266	silhouette	11.62	2881	315	croix
				29.94	10310	1317	femmes	11.62	2067	245	pointe
				29.71	4910	776	bord	11.43	2033	240	univers
				29.30	1499	354	cigarette	11.42	4041	407	papier
				29.19	2086	435	étoiles	11.42	2519	282	oreilles
				29.08	1095	290	colline	11.15	2557	282	portes
				29.07	2808	526	endroit	11.12	3361	348	colère
				28.91	7513	1028	ombre	11.02	11279	935	silence
				28.69	4739	741	musique	11.00	3754	378	voyage
				28.57	6759	947	pierre	10.44	5557	508	garçon
				28.40	992	268	milliers	10.24	4310	410	peuple
				28.33	1782	385	langage	10.15	3187	321	robe
				28.30	1217	303	ouest	10.13	29635	2144	jour
				28.16	930	256	puits	10.10	3754	365	passage
				26.82	1467	327	signes	9.71	2571	266	solitude
				26.31	2253	425	pont	9.36	3372	325	lune
				25.37	1196	276	surface	9.36	2313	241	soldats
				25.19	1063	255	toits	8.61	2597	255	vague
				25.18	971	241	sommet	8.53	3722	340	verre
				24.94	7935	986	route	8.42	6573	545	chemin
				24.42	9869	1147	peur	8.42	2421	239	poids
				24.28	1274	278	boue	8.38	3418	315	douleur
				24.16	1721	336	chiens	8.26	3506	320	foule
				23.64	3069	487	arbre	7.97	7313	588	sang
				23.29	1629	316	trottoir	7.95	7969	633	cause
				23.25	2640	435	champs	7.91	2602	247	fatigue
				23.16	3381	514	cesse	7.76	3009	276	poitrine
				22.71	3116	480	nord	7.09	3912	333	village
				22.52	1227	257	lunettes	6.93	2625	237	fête
				22.34	10311	1136	bout	6.85	9774	731	fin
				22.21	14245	1456	gens	6.77	3654	310	secret
				22.06	1317	265	abri	6.77	3253	281	réalité
				21.86	2365	388	gorge	6.72	3721	314	genre
				21.82	1419	276	sueur	6.65	2590	231	gestes
				21.50	13158	1349	fond	6.42	7776	588	bouche
				21.41	1236	249	animaux	6.34	4827	386	paroles
				21.38	8428	954	milieu	6.21	3639	301	ligne
				21.08	7139	836	pieds	6.14	7577	569	histoire
				20.86	1248	246	étendue	6.14	3740	307	chien
				20.71	1281	249	routes	6.09	4699	373	jardin
				20.46	1851	315	creux	5.96	3647	298	hasard
				20.13	1611	284	port	5.77	5479	421	lèvres
				19.83	1679	289	matière	5.68	4458	350	signe
				18.77	5537	652	forme	5.53	4435	346	forces
				18.71	1931	306	branches	4.98	10169	713	force
				18.55	2190	332	faim	4.97	8715	619	pays
				18.48	2507	364	fenêtres	4.60	5623	410	années
				18.47	10767	1080	nom	4.34	3054	235	cou
Substantifs											
153.11	6819	3367	mer	30.11	913	267	gouttes	11.90	2270	266	côte
107.20	8196	2758	lumière	30.09	908	266	silhouette	11.62	2881	315	croix
95.85	7270	2344	vent	29.94	10310	1317	femmes	11.62	2067	245	pointe
94.10	1062	782	plage	29.71	4910	776	bord	11.43	2033	240	univers
93.19	10986	2936	ciel	29.30	1499	354	cigarette	11.42	4041	407	papier
89.19	10706	2793	soleil	29.19	2086	435	étoiles	11.42	2519	282	oreilles
85.50	15601	3420	terre	29.08	1095	290	colline	11.15	2557	282	portes
84.91	1124	734	montagnes	29.07	2808	526	endroit	11.12	3361	348	colère
80.82	12487	2851	eau	28.91	7513	1028	ombre	11.02	11279	935	silence
78.55	1740	871	nuages	28.69	4739	741	musique	11.00	3754	378	voyage
78.52	1164	697	vallée	28.57	6759	947	pierre	10.44	5557	508	garçon
76.95	7241	1961	bruit	28.40	992	268	milliers	10.24	4310	410	peuple
76.72	567	462	rochers	28.33	1782	385	langage	10.15	3187	321	robe
71.27	2266	929	centre	28.30	1217	303	ouest	10.13	29635	2144	jour
69.55	1647	759	fleuve	28.16	930	256	puits	10.10	3754	365	passage
69.34	2186	889	sable	26.82	1467	327	signes	9.71	2571	266	solitude
66.27	759	473	collines	26.31	2253	425	pont	9.36	3372	325	lune
65.24	1579	701	vagues	25.37	1196	276	surface	9.36	2313	241	soldats
63.92	326	290	ciment	25.19	1063	255	toits	8.61	2597	255	vague
61.75	266	252	immeubles	25.18	971	241	sommet	8.53	3722	340	verre
58.36	2208	773	pierres	24.94	7935	986	route	8.42	6573	545	chemin
58.14	272	241	ravin	24.42	9869	1147	peur	8.42	2421	239	poids
54.35	2158	719	poussière	24.28	1274	278	boue	8.38	3418	315	douleur
53.18	1563	585	bruits	24.16	1721	336	chiens	8.26	3506	320	foule
51.58	559	319	rocher	23.64	3069	487	arbre	7.97	7313	588	sang
51.51	2243	704	oiseaux	23.29	1629	316	trottoir	7.95	7969	633	cause
51.24	1767	609	dieux	23.25	2640	435	champs	7.91	2602	247	fatigue
48.80	3856	937	peau	23.16	3381	514	cesse	7.76	3009	276	poitrine
48.49	1898	607	désert	22.71	3116	480	nord	7.09	3912	333	village
47.75	2776	753	rues	22.52	1227	257	lunettes	6.93	2625	237	fête
47.70	8243	1499	ville	22.34	10311	1136	bout	6.85	9774	731	fin
47.40	9203	1606	long	22.21	14245	1456	gens	6.77	3654	310	secret
47.03	3800	903	murs	22.06	1317	265	abri	6.77	3253	281	réalité
46.62	14236	2139	corps	21.86	2365	388	gorge	6.72	3721	314	genre
46.48	1241	457	voitures	21.82	1419	276	sueur	6.65	2590	231	gestes
45.95	691	324	cailloux	21.50	13158	1349	fond	6.42	7776	588	bouche
45.70	13161	2001	visage	21.41	1236	249	animaux	6.34	4827	386	paroles
44.41	2027	588	île	21.38	8428	954	milieu	6.21	3639	301	ligne
44.11	450	246	insectes	21.08	7139	836	pieds	6.14	7577	569	histoire
43.56	3501	810	couleur	20.86	1248	246	étendue	6.14	3740	307	chien
43.25	1634	506	bateau	20.71	1281	249	routes	6.09	4699	373	jardin
43.01	1047	388	herbes	20.46	1851	315	creux	5.96	3647	298	hasard
42.71	2194	598	horizon	20.13	1611	284	port	5.77	5479	421	lèvres
42.62	2273	610	espace	19.83	1679	289	matière	5.68	4458	350	signe
41.85	4703	949	arbres	18.77	5537	652	forme	5.53	4435	346	forces
41.25	472	238	habits	18.71	1931	306	branches	4.98	10169	713	force
40.57	3621	785	intérieur	18.55	2190	332	faim	4.97	8715	619	pays
40.49	762	307	métal	18.48	2507	364	fenêtres	4.60	5623	410	années
40.43	18888	2410	hommes	18.47	10767	1080	nom				

4.31	11254	766	guerre	Noms propres			
4.26	14243	953	mains				ponctuation
3.73	3308	244	début	165.56	286	674	Diego
3.71	11795	785	rue	165.38	245	622	Esther
3.56	4912	346	filles	133.71	55	236	Geoffroy
3.34	5561	384	salle	125.25	180	405	Indiens
3.18	6823	461	midi	98.22	625	613	Besson
				86.28	450	456	Adam
				77.72	165	244	Mexique
				38.83	752	294	Martin
				28.27	955	261	Laure
				23.75	1255	271	Suzanne
				7.27	7022	554	Jacques

Les mots qui se trouvent en tête de liste n'étonneront aucun lecteur de *Le Clézio*. Nous sommes en présence des éléments de la nature, auxquels l'œil et le cœur de l'écrivain sont si sensibles. Les noms propres écartés, rien d'humain dans la liste des substantifs. C'est le règne du minéral : une terre (plus souvent une mer) que l'homme n'habite pas, et que le végétal et le vivant ne couvrent pas encore. Le mot le plus significativement suremployé du corpus est le substantif *mer*, suivi par *lumière, vent, plage, ciel, soleil, terre, montagne, eau, nuages, vallée* etc., liste qui constitue un hymne à la nature de la part de *Le Clézio*. En effet la contemplation de la nature est au cœur de l'œuvre dès le premier roman. L'extrait suivant de *Le procès-verbal* illustre bien l'importance de cet élément dans l'œuvre⁵¹³ :

“Elle était, c'est cela – elle était le seul point mobile dans tout le pays. Autour d'elle, la nature était pareille, immobile – excepté qu'elle lui formait, comment dire ? un halo autour de la tête, comme si la terre et le ciel étaient sa chevelure.”

Il est intéressant de constater dans le début de la liste la présence de plusieurs versants de la nature, notamment de la nature aquatique. En effet, le thème de la mer et le décor marin sont omniprésents dans l'œuvre leclézienne et notre liste le met en exergue avec des substantifs comme *plage, eau, fleuve, sable, vagues, bateaux, rivière, navire*, etc.⁵¹⁴

⁵¹³ *Le procès-verbal*, p. 249.

⁵¹⁴ Nous reviendrons plus en détail sur les thèmes significatifs de l'œuvre dans les chapitre suivants.

Les substantifs faisant référence au monde terrestre, l'antipode du monde marin, se trouvent également en haut de liste avec des mots comme *terre, montagnes, rochers, sable, collines, pierres* etc. projetant l'image d'une terre méditerranéenne, désertique et chaude, comme les adjectifs *doux, douce, blanche(s) et chaud* nous le rappellent.

Au-dessus de la terre et de la mer nous avons le *ciel* avec ses *nuages* et son *soleil* que reflète la *lumière* et qui *brille* (la forme verbale la plus significative du corpus), qui *brillait* et qui *brûle*.

Il est intéressant de noter que, parmi ces substantifs appartenant au registre de la nature, sont intercalés les substantifs *ciment, immeubles* et *bruits*. Il s'agit ici de l'antipode de la nature, l'environnement urbain, dénoncé et critiqué par Le Clézio dans ses débuts romanesques et dans ses essais tout au long de son œuvre.

Le thème de la ville est en effet un thème très répandu pendant les années 60 et 70. Selon Roland Barthes⁵¹⁵ :

“La cité est un discours et ce discours est véritablement un langage : la ville parle à ses habitants, nous parlons notre ville, la ville où nous nous trouvons, simplement en l'habitant, en la parcourant, en la regardant. Cependant, le problème est de faire surgir du stade purement métaphorique une expression comme ‘langage de la ville’.”

Dans l'écriture leclézienne, il s'agit plutôt de dénoncer cet environnement. En effet, les protagonistes des différents livres fuient l'agression du milieu urbain. Il s'agit, selon Jean-Xavier Ridon⁵¹⁶, d'une “situation étrange et angoissante des êtres qui créèrent ces villes et qui se retrouvent dans une position d'antagonisme par rapport à elles. Comme si les villes ne donnaient plus aux humains la place qu'ils

⁵¹⁵ R. Barthes “Sémiologie et urbanisme”, in *L'aventure sémiologique*, Paris, Points essais, Seuil, p. 265.

⁵¹⁶ J.-X. Ridon (1995) : p. 25.

s'étaient octroyés dans l'espace." Dans *Le livre des fuites* de 1969 Le Clézio s'exprime ainsi⁵¹⁷ :

“On fuyait la ville trop blanche, les murs trop droits, les bruits de pas, le va-et - vient des voitures, les douleurs de la connaissance. [...] Écrasée, rejetée, piétinée, la ville maudite. Couverte de cendres, de vieux papiers. Oublié le dépotoir plein de pourritures. On avait creusé sa tombe, puis on l'avait recouverte de fumier.”

Côté verbe, c'est la contemplation qui est mise en exergue par la liste des spécificités avec les verbes *regarder* et *voir*⁵¹⁸ dans leurs différentes formes flexionnelles. Le regard est en effet toujours très présent dans l'œuvre de Le Clézio ainsi que les verbes du mouvement comme *marcher*, *aller*, *courir*, *venir*, *partir*, *arriver* et *avancer*, confirmés par la liste des prépositions, toutes spatiales. Avec le regard et le mouvement on a deux façons complémentaires d'embrasser l'espace dans l'œuvre leclézienne.

La liste de spécificités positives témoigne également de l'importance qu'attache Le Clézio aux couleurs. Rien de plus naturel chez un écrivain qui privilégie le regard et l'observation que de “peindre une image en couleur”. L'écrivain est en fait très sensible aux couleurs et y attache volontiers une valeur symbolique souvent émanant de son grand savoir ethnologique :

“Pour les Aztèques, écrit Le Clézio dans l'essai *Le rêve mexicain*⁵¹⁹, comme pour les Purepecha et les Mayas, le monde est carré, et chacun de ses quatre côtés est symbolisé par une couleur, un arbre, un animal, et un esprit gardien. Pour les Mayas, les esprits gardiens des orientes sont des dieux importants : ce sont les Bacab, porteurs des pierres Ahcantun : au nord, le Bacab blanc (Zacal Bacab), au sud le Bacab jaune (Kanal Bacab), à l'est le Bacab rouge (Chacal Bacab), à l'ouest le Bacab noir

⁵¹⁷ *Le livre des fuites*, p. 52.

⁵¹⁸ Nous ferons une étude détaillée de ce thème dans le chapitre 5.4.5. “Le regard”.

⁵¹⁹ *Le rêve mexicain*, p. 126.

(Ekal Bacab). Au centre de ce carré, se trouve la couleur Yax, le vert du jade. L'origine de cette division du monde en orient et en couleurs est sans doute cosmogonique. Elle fait songer à l'ancienne représentation du monde des Chinois, où à chaque orient correspondait une couleur, et un organe vital (tsang) : poumons, cœur, rate, foie, reins. Chez les Mayas, cette division se rapporte aux premières observations des astres, sans aucun doute le plus ancien héritage des cultures de l'Amérique moyenne : monuments des Mound Builders (sud-est des États-Unis), pyramides toltèques ou mayas.”

La couleur préférée de Le Clézio, dit-il dans une interview à Jean-Louis Ezine⁵²⁰, est indéniablement le bleu (couleur de la mer et du ciel). Toutefois l'analyse statistique nous montre que la couleur la plus significativement excédentaire dans l'œuvre est le noir. Les autres couleurs privilégiées sont *le bleu*, *le blanc*, *le gris* ainsi que *le rouge*. Nous venons d'évoquer l'image du règne du minéral : “la mer et la terre que le végétal ne couvre pas encore”. Les couleurs sont par conséquent celles du minéral. Les paysages que décrit Le Clézio sont rarement verdoyants (le *vert*, quoique excédentaire, n'arrive qu'en bas de la liste), il s'agit plus souvent d'un désert aride ou torride, une montagne ou des rochers pierreux et inaccessibles, ou bien du paysage urbain avec son béton et son asphalte, les terrains vagues en marge de l'urbanisation.

Toutefois, le vocabulaire d'un auteur n'est pas seulement caractérisé par les mots qu'il emploie très fréquemment mais également par la rareté ou l'absence de certains mots.

Le vocabulaire spécifique négatif, ce sont les mots qui – de façon significative – dans le corpus faisant l'objet de l'étude, sont moins utilisés que dans le corpus de référence. La liste ci-dessous regroupe les formes déficitaires du corpus Le Clézio,

⁵²⁰ *Ailleurs*, Entretiens avec Jean-Louis Ezine, Paris, Arléa (Entretiens France-Culture), 1995, p. 79.

-18.67	52265	2057	deux	mots très ambiguës
-17.50	44679923405		un	
-16.22	20459	653	trois	-33.34 30581613571 pas
-12.97	9633	265	dix	
-12.26	8568	235	cinq	
-5.10	11107	524	quatre	Signes de ponctuation
				-22.16 40921 1343)
				-22.35 40245 1303 (

La liste permet de constater le déficit des pronoms personnels à la première et à la deuxième personne au singulier aussi bien qu'au pluriel. Nous avons déjà pu observer l'importance de la narration à la troisième personne dans l'œuvre leclézienne. Les fortes valeurs négatives de l'écart réduit de la première et la deuxième personnes (-64.56 pour *vous*, -50.50 pour *me*, -50.01 pour *je* etc.) soulignent encore le fait que l'œuvre contient peu de dialogue et nous notons également le déficit d'éléments d'articulation du discours. La littérature leclézienne serait-elle celle du silence et de la solitude comme on le dit souvent ? La statistique semble confirmer cette assertion.

Que les pronoms possessifs se trouvent aussi dans cette liste n'étonne aucun lecteur de Le Clézio, car non seulement ils appartiennent aux deux premières personnes et au dialogue, mais leur déficit témoigne aussi de ce que la possession⁵²¹ intéresse très peu notre écrivain, surtout la possession matérielle qu'il va même jusqu'à refuser. Le Clézio rejette en effet les valeurs qu'attribue la culture occidentale à la possession et à la matière⁵²² :

“La matière sans infini, sans éternité, mais aussi sans mesures et sans histoire, la matière est étendue de toutes parts, elle remplit tout. Elle n'a pas de limites, parce qu'elle n'a pas d'illimité. Être, ce n'est donc pas être vivant. La vie, la mort sont des modalités sans importance, comme végétal ou minéral. La vie et la mort sont des formes qu'adopte la matière, parmi tant d'autres. Quel est donc le principe fondamental de

⁵²¹ Si l'on peut se permettre de donner aux “possessifs” un sens un peu fort.

⁵²² *L'inconnu sur la terre*, p. 238.

cette matière ? Est-ce le mouvement ? Mais comment découvrir cette loi qui ne serait même plus une loi, mais une réalité ? Comment inventer la seule chose qu'on ne puisse récuser, le seul principe qui soit dans le monde et qui n'en découle pas ?”

L'immobilité est une autre caractéristique de la narration leclézienne. Le verbe *faire* est le plus déficitaire du corpus. Cette immobilité n'est pourtant pas passive dans l'écriture leclézienne, il s'agit plutôt d'une forme de sagesse où l'immobilité et l'observation du monde ne correspondent pas à l'inactivité mais à une prise de distance du quotidien⁵²³ :

“Les êtres naissaient, puis disparaissaient ; se divisaient sans cesse, comblaient le vide, comblaient le temps, goûtaient, et étaient goûtés. Les millions d'yeux, les millions de bouches, les millions de nerfs, d'antennes, de mandibules, de tentacules, de pseudopodes, de cils, de suçoirs, d'orifices tactiles étaient ouverts dans le monde entier et laissaient entrer les doux effluves de la matière.

Partout ce n'étaient que lumières, cris, parfums, froid et chaleur, duretés, nourritures. Partout ce n'étaient que frémissements, ondes et vibrations. Et pourtant, pour moi, c'était le silence, l'immobilité et la nuit. C'était l'anesthésie. Car ce n'était pas dans ces communications éphémères que résidait ma vérité. Ce n'était pas dans cette lumière, dans cette nuit, ni dans rien de ce qui était manifesté pour la vie.”

L'immobilité est en effet bien plus que le “non-mouvement” chez Le Clézio. Jean-Xavier Ridon parle⁵²⁴ même d'une dérive du langage où “l'immobilité des mots inscrits sur le papier vient donner naissance au mouvement de la pensée. Le sens de la dérive serait alors celui d'une pensée qui tend à se créer un espace nouveau,

⁵²³ *L'extase matérielle*, p. 13.

⁵²⁴ J.-X. Ridon (1995) : p. 22-23.

qui cherche à trouver une brèche par laquelle pourrait s'élaborer une idée autre du *différent* et de l'*ailleurs*'", et il cite ce qu'écrit Deleuze à propos de Nietzsche⁵²⁵ :

“Mais aussi, le nomade, ce n'est pas forcément quelqu'un qui bouge ; il y a des voyages sur place, des voyages en intensité, et même historiquement les nomades ne sont pas ceux qui bougent à la manière des migrants, au contraire ce sont ceux qui ne bougent pas, et qui se mettent à nomadiser pour rester à la même place en échappant aux codes.”

Le vocabulaire spécifique de Le Clézio, côté positif aussi bien que côté négatif, reflète parfaitement son écriture si personnelle et ses caractéristiques. On est au premier jour de la Création : les éléments viennent d'apparaître. Rien ne bouge : pas de verbe, peu d'adjectifs ou d'adverbes. Aucune articulation du discours : on dirait que la phrase est faite de substantifs juxtaposés. Seul le regard est actif, témoignant cette immobilité “active” que déplore Le Clézio dans notre culture occidentale et citadine.

Toutefois ces thèmes ne sont pas constants à l'intérieur de l'œuvre, l'étude interne permet de constater des variations importantes.

5.3.3. Les spécificités endogènes.

Un mot appartient au vocabulaire spécifique d'une partie d'un corpus lorsque sa fréquence relative dans ce texte, ou dans cette partie, s'écarte *significativement* de celle observée dans l'ensemble du corpus. Si la fréquence dans la partie étudiée est supérieure à la fréquence attendue, on dit que le mot est une *spécificité positive* ; dans le sens contraire, la spécificité est dite *négative*. En revanche, si l'écart n'est significatif dans aucune des parties du corpus, on dit que le mot est “non-

⁵²⁵ G. Deleuze et F. Gattari (1980) : p. 174.

spécifique” ou “commun”. Si les tests permettent de conclure avec plus ou moins de certitude à la spécificité d’un mot, la “non-spécificité” n’a pas le même statut : la normalité de la distribution est l’hypothèse la plus probable, mais elle n’est pas prouvée.

Le logiciel Hyperbase effectue ces calculs et permet l’observation du vocabulaire spécifique de chacune des 31 œuvres du corpus Le Clézio en s’appuyant sur les mêmes techniques déjà utilisées dans les analyses de structures et du rythme, la loi normale et les écarts réduits. Les spécificités propres au corpus peuvent aussi être repérées grâce à un système de segmentation interne du texte en plusieurs parties comparables. L’application fonctionne comme pour les spécificités externes, c’est-à-dire en accueillant des excédents de chaque partie du corpus par rapport aux autres, et la spécificité est déterminée par le calcul de l’écart réduit pour chaque forme dans chaque partie du corpus. Rappelons que les différences de taille entre les unités de comparaison dans le corpus sont nivelées grâce à une pondération statistique.

Les résultats de l’analyse sont très nets, les listes de mots obtenues reflètent parfaitement le thème de l’ouvrage et nous donnent le profil caractéristique de chaque livre. Le tableau ci-dessous rend compte des 10 mots les plus fréquents de chaque œuvre, les noms propres et les mots grammaticaux écartés⁵²⁶ :

⁵²⁶ En annexe n°5 les listes complètes de spécificités.

Liste des mots spécifiques des différents sous-corpus

Vocabulaire spécifique du				texte n°9 PROPHEITIES			texte n°17 CHERCHEUR				
texte n°1 PROCES				Evart	Corpus	Texte	Mot	33.05	74	73	timonier
Ecart	Corpus	Texte	Mot	48.11	418	64	peuple	23.76	188	90	capitaine
27.78	55	44	rat	32.19	232	32	parole	23.75	3033	502	mer
16.04	27	18	billard	31.83	604	52	dieux	23.25	91	59	corsaire
14.45	289	62	chien	25.46	250	27	livres	21.94	208	89	pirogue
14.38	126	38	savez	23.38	53	11	calendrier	21.48	295	107	navire
14.25	48	22	trucs	23.10	29	8	prophéties	21.40	517	149	rivière
14.16	123	37	type	18.45	115	13	prêtres	20.26	243	91	ravin
13.95	38	19	noyé	18.16	245	19	univers	20.16	118	60	anse
12.35	117	32	demanda	17.68	804	35	jours	19.60	274	95	entends
12.33	111	31	cas	16.50	83	22	mois				
11.24	100	27	paquet	texte n°10 MONDO				texte n°18 ANGOLI MALA			
				26.45	798	192	petite	20.16	54	43	
				25.09	36	33	fronde	17.91	340	66	forêt
				24.35	38	33	bouc	14.64	759	71	fleuve
				22.46	30	27	directrice	14.31	144	38	torrent
				22.15	308	95	croix	12.01	57	23	gardes
				12.84	624	108	regardait	11.00	30	17	fugitif
				21.48	26	24	lanière	9.65	63	18	jaguar
				19.82	182	64	plaine	9.58	1363	62	maison
				18.80	89	41	troupeau	8.80	90	18	plancher
				18.57	45	28	gitan	8.43	161	21	sauvage
				texte n°11 INCONNU				texte n°19 RODRIGUES			
				33.35	459	192	beauté	44.31	1114	214	père
				29.27	2601	495	lumière	40.19	166	71	trésor
				22.07	577	156	espace	35.00	243	76	ravin
				21.42	86	51	vertu	34.96	118	52	anse
				20.63	2759	409	ciel	32.60	1957	224	grand
				20.55	1755	300	vie	29.36	64	32	quête
				17.34	379	100	langage	25.96	39	22	documents
				14.84	774	140	nuages	25.61	40	22	commandeur
				14.72	420	94	visages	24.63	135	40	plan
				14.20	135	46	fleur	22.65	91	30	corsaire
				texte n°12 ICEBERGS				texte n°20 REVE			
				26.81	70	8	poèmes	47.52	362	192	indiens
				19.33	75	6	poésie	44.73	604	240	dieux
				17.88	337	12	langue	38.71	100	80	barbares
				17.81	1080	22	mots	36.95	114	82	aztèques
				11.00	57	3	poème	36.17	95	73	rites
				10.96	379	8	langage	31.75	126	75	conquérants
				10.34	29	2	tremblement	30.56	128	73	mayas
				9.10	37	2	pois	29.63	74	53	culte
				8.86	39	2	poète	28.58	738	182	Dieu
				8.85	549	8	savoir	28.49	175	81	conquête
				texte n°13 ARBRES				texte n°21 PRINTEMPS			
				59.37	25	11	chêne	29.56	59	42	colonel
				51.34	866	57	arbres	25.50	115	52	amie
				26.73	554	24	garçon	20.84	27	20	opéra
				26.26	260	16	forêt	19.06	138	44	appartement
				25.78	39	6	clairière	18.75	896	128	mère
				25.76	53	7	sifflant	17.94	264	60	souviens
				25.15	1016	31	peut	16.94	43	21	lycée
				20.18	286	13	branches	16.90	256	56	voulais
				16.47	42	4	sifflements	14.01	28	14	alcôve
				15.93	174	8	racines	13.19	90	25	madame
				texte n°14 DESERT				texte n°22 SIRANDANES			
				31.57	575	232	désert	33.95	60	9	langues
				24.60	118	76	manteau	23.44	337	15	langue
				23.06	216	101	dunes	20.00	34	4	langages
				21.50	126	70	cité	17.49	379	12	langage
				19.60	101	57	guerrier	15.93	53	4	créole
				18.78	818	193	sable	13.25	43	3	mépris
				17.92	26	25	figuier	12.38	49	3	amérindiennes
				17.80	2386	386	hommes	11.25	159	5	français
				12.67	773	152	pierres	10.57	30	2	commun
				12.28	2344	313	vent	10.03	478	8	pouvoir
				texte n°15 VILLES				texte n°23 ONITSHA			
				36.67	308	56	croix	46.26	659	241	fleuve
				24.74	866	67	arbres	26.20	71	43	reine
				23.71	135	24	fleur	24.83	208	73	pirogue
				22.37	260	32	forêt	21.74	47	29	épave
				17.31	202	22	seigneur	16.55	31	18	club
				15.60	3288	98	terre	16.37	35	19	embarcadère
				14.78	49	9	vengeance	15.27	86	29	varangue
				13.75	299	22	fleurs	14.97	112	33	coque
				13.63	232	19	parole	14.70	362	64	pont
				13.42	238	19	puits	13.21	25	13	membrures
				texte n°16 RONDE				texte n°24 ETOILE			
				39.22	203	15	pouce	31.58	150	96	maman
				22.27	79	41	mobile	26.33	68	53	italiens
				20.08	63	33	aurore	23.90	104	61	berger
				17.76	82	34	villa	23.77	191	85	camp
				16.28	319	68	sent	23.66	53	42	juifs
				15.70	66	27	dame	21.36	107	56	torrent
				15.18	33	18	vélocoteur	20.89	1351	248	enfants
				15.07	189	47	frère	20.82	25	25	carabiniers
				15.01	53	23	loup	19.60	73	42	piano
				14.32	104	32	guide	19.25	51	34	valises
								19.24	38	29	fugitifs
								18.51	239	78	soldats

texte n°25 PAWANA			24.00	213	109	baie	texte n°30 NUAGES			
88.11	63	49	23.56	71	58	ilot	29.25	100	30	tombeau
80.8	55	42	23.40	511	179	île	24.09	575	63	désert
51.52	37	22	22.46	261	116	plate	22.09	301	41	rocher
29.91	57	16					21.85	130	26	cheikh
19.76	25	7	texte n°28 POISSON				19.96	693	59	vallée
15.54	79	10	28.74	34	34	fondouk	17.90	83	17	nomades
15.09	31	6	26.56	35	32	princesses	17.14	39	11	tribus
13.69	65	8	15.10	55	24	disais	14.66	52	11	arabes
13.30	254	16	14.66	34	18	garage	13.80	58	11	troupeaux
12.76	213	14	14.58	139	39	bébé	13.33	26	7	saints
			14.06	87	29	aimais				
			13.81	59	23	comprendais	texte n°31 HASARD			
texte n°26 DIEGO			13.64	634	92	disait	17.01	36	36	rouf
56.50	171	139	13.57	183	43	cour	15.77	506	98	bateau
50.71	250	153	13.56	148	38	sentais	11.12	362	60	navire
45.95	104	88					10.48	108	33	barre
43.53	76	71	texte n°29 FETE				10.38	425	61	pont
42.07	153	99	41.31	604	200	dieux	10.28	90	30	cabine
36.72	49	48	35.67	135	78	seigneurs	9.99	150	36	coque
28.95	37	33	34.63	160	83	relation	9.36	106	29	quai
28.26	136	64	30.09	49	39	amérindien	9.03	46	20	rade
27.88	52	38	29.07	175	74	conquête	8.92	31	17	commissaire
26.80	77	45	28.99	45	36	Amérindiens				
26.76	74	44	28.49	195	77	lac				
			25.30	128	55	empire				
texte n°27 QUARANTAINE			23.94	202	6	seigneur				
38.07	136	128	23.41	738	137	Dieu				
35.78	124	115								
30.25	156	112								
29.08	78	74								
27.70	160	105								
25.80	59	57								

Cette technique résume les thèmes principaux de chaque livre. Nous trouvons le *chien errant* et le *rat*, thèmes itératifs dans *Le procès-verbal*, les trottoirs de la ville dans les trois premiers romans et dans *L'extase matérielle*, le premier essai littéraire du corpus ; nous reconnaissons des mots comme *matière*, *esprit*, *conscience*, *néant*, *réalité*, *vie* etc. qui reflètent les thèmes de la pensée et de la réflexion qui occupe l'auteur dans cet ouvrage.

Nous avons regroupé les dix mots les plus significatifs de *L'extase matérielle* et constitué une liste de ces dix mots dans l'ensemble de l'œuvre, qui a été ensuite représenté par le graphique suivant :

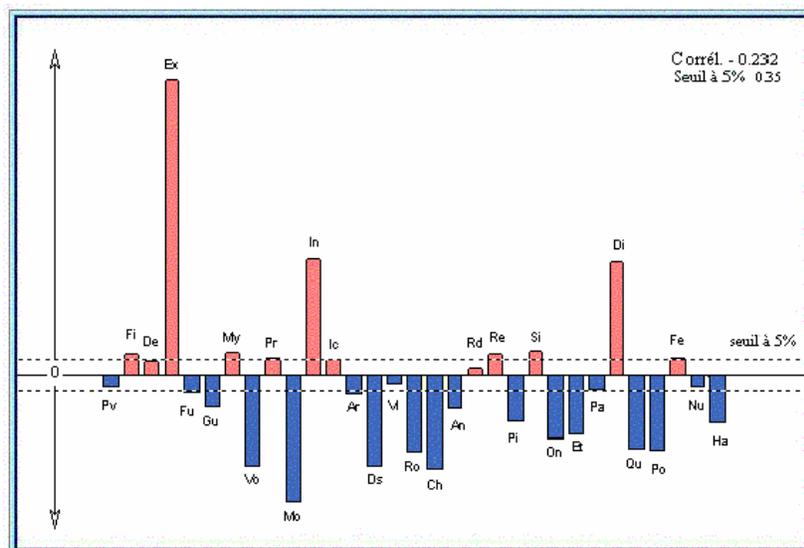


Figure n°156 : La distribution relative du regroupement des dix mots les plus fréquents dans *L'extase matérielle*.

L'histogramme montre que les questions métaphysiques qui préoccupent Le Clézio dans *L'extase matérielle* sont également présentes de façon significative dans l'essai *L'inconnu sur la terre* et dans la biographie *Diego et Frida*, tandis qu'elles occupent beaucoup moins de place dans l'œuvre romanesque. Quel est le lien entre les deux essais et *Diego et Frida* ? La biographie des deux peintres ne semble pas au premier abord s'occuper des mêmes questions que les essais littéraires. Afin de répondre à cette question, nous avons vérifié la distribution de ces dix vocables dans *Diego et Frida*. Le graphique ci-dessous, qui illustre la distribution, montre que les dix thèmes ne sont pas du tout significatifs dans cette biographie mais qu'il est surtout question de la "vie", chose naturelle pour une biographie, et plus précisément celle de Frida Kahlo, torturée par sa maladie :

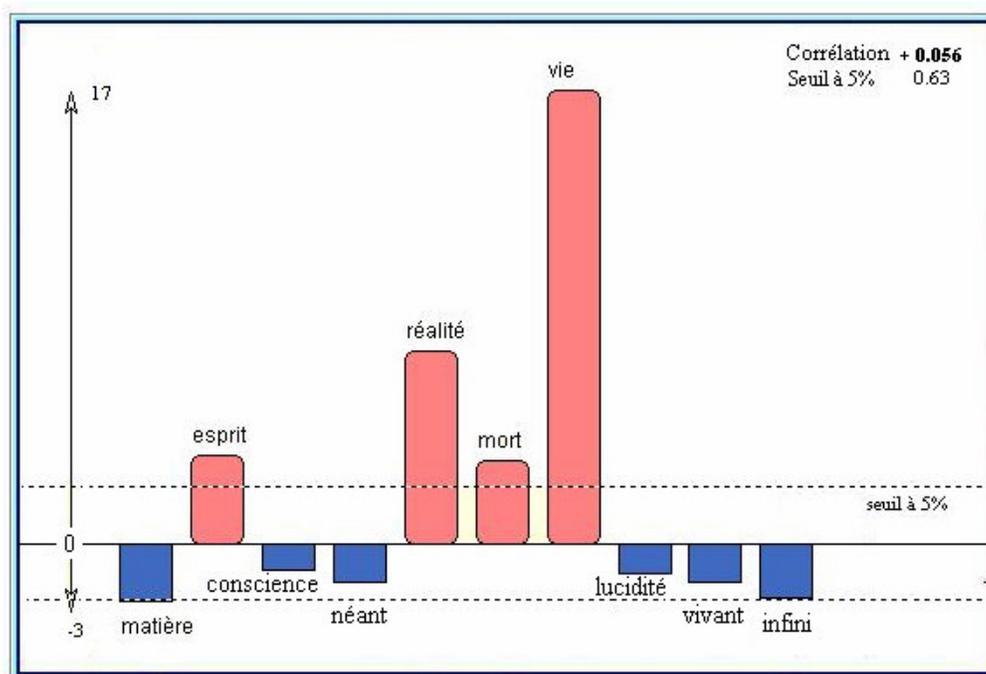


Figure n°157 : La distribution des 10 mots dans Diego et Frida.

C'est donc ce mot *vie* qui, à lui seul - ou presque -, déclenche le rapprochement entre *Diego et Frida* et *L'extase matérielle*. Le mot *matière* apparaît 4 fois dans *Diego et Frida*, l'*esprit* 21 fois, la *conscience* 7, le *néant* 2, la *réalité* 36, la *mort* 55, la *vie* 201, la *lucidité* 1, le *vivant* 6 et enfin l'*infini* 2.

L'étude interne montre bien que le milieu urbain dont nous avons déjà souligné l'importance est concentré dans les premiers ouvrages, dans *Le livre des fuites* et dans *La guerre*. En 1969, la même année que la parution de *Le livre des fuites*, Le Clézio répond à Pierre Lhoste⁵²⁷ :

“Je suis fasciné par la ville, par les agressions de la ville. Une ville c'est un amoncellement d'êtres humains qui se battent les uns les autres, qui cherchent à se dominer les uns les autres et mon regard s'accroche à tous les détails de cette domination. Domination par le mépris, domination par la violence, domination par les rites : le rite de l'automobile, le rite de la promenade, de l'achat dans les magasins. Ces choses-là me troublent et m'exaspèrent et cela crée cet état d'anxiété dont j'ai besoin aussi.”

On trouve, en effet, dans les ouvrages de cette époque non seulement le reflet d'un sentiment de captivité dans ce milieu urbain et une impossibilité de le fuir, mais également une atmosphère de menace. Dans le roman *La guerre* l'écrivain note dès la première page : “La guerre est en route pour durer dix mille ans, pour durer plus longtemps que l'histoire des hommes. Il n'y a pas de fuite possible, pas de désaveu.” Selon Le Clézio c'est justement le sens intime du livre⁵²⁸ :

“Pour moi, c'est peut-être cette impossibilité d'échapper au monde de l'urbanisme, d'une part parce qu'en tant qu'Occidental je suis obligé d'y vivre, d'autre part parce que ce monde s'étend et qu'il est en train de détruire les portions d'espaces où il n'y a pas encore de villes. Cette avancée de ce qu'on appelle civilisation ne s'effectue pas calmement, doucement mais d'une façon très agressive et cette avancée des villes amène une sorte de combat permanent entre les hommes et la nature.”

⁵²⁷ P. Lhoste (1969) : p. 54-55.

⁵²⁸ P. Lhoste (1969) : p. 63.

Les thèmes significatifs des ouvrages suivants peuvent presque être dotés d'une valeur symbolique : le texte n°7, *Mydriase*, ouvre l'œil et aiguise le regard avec des mots comme *yeux, pupilles, voit, astre, voient* etc. ; le voyage et la quête sont mis en exergue dans le texte n° 8 *Voyages de l'autre côté* avec des mots comme *pays, air, marin, voyager, voit, vite, regardes*, etc. Et soudain arrive dans la production, avec des mots significatifs tels *peuple, parole, dieux, livres, prophéties, univers* etc., *Les prophéties de Chilam Balam*, le neuvième texte du corpus, avec la découverte du monde et de la culture amérindienne, apportant non seulement une échappatoire possible à notre culture occidentale, mais également des réponses si importantes pour Le Clézio qu'elles ouvrent la voie vers le tournant décisif dans son œuvre.

C'est donc seulement à partir de *Mondo et autres histoires* que nous trouvons les éléments de la nature parmi les thèmes significatifs des livres. Dans *L'inconnu sur la terre* ces phrases, qui sont un véritable éloge à la beauté, témoignent du changement important⁵²⁹ :

“Le langage ne guide pas vers l'espace illimité ; il vous conduit pas à pas sur les sentiers réels de la terre. La beauté n'est pas secrète. Elle est libre, exposée de toutes parts. Le ciel est si grand, la mer, et la lumière respandit. Tout est si calme, si vaste, le silence est si profond, à travers lui passent les vols d'oiseaux blancs, lentement, voyageant le long du ciel.”

Avec *Désert* nous constatons l'introduction des thèmes exotiques, jusqu'alors peu exploités dans l'œuvre, avec les dunes, le sable et les pierres du désert. L'univers marin avec ses éléments est significatif dans *Le chercheur d'or*, dans *Voyage à Rodrigues, Pawana, La quarantaine* et *Hasard*. Nous trouvons les thèmes du monde amérindien et de leurs rites dans les ouvrages qui leur sont consacrés, *Le rêve mexicain* et *La fête chantée*, et celui du désert à nouveau dans *Gens des nuages*. La vie des jeunes et des adolescents se reflète dans la liste de *La ronde* et *Le printemps et*

⁵²⁹ *L'inconnu sur la terre*, p. 10.

autres saisons, tandis que la biographie *Diego et Frida* est le seul livre du corpus qui traite le thème de la révolution et du communisme.

Une autre technique pour aborder les spécificités d'un corpus est celle de Dominique Labbé qui, en employant les mêmes formules statistiques, complète bien la technique d'Hyperbase.

5.3.4. Le vocabulaire spécifique selon la méthode de Labbé (logiciel Lexicométrie).

La spécificité du vocabulaire de chaque partie est ici également jugée par rapport à la totalité du corpus. Le calcul des spécificités ne fait donc pas appel à une norme extérieure au corpus, ce qui résout l'épineux problème du "français standard" et du niveau de langue dans les corpus de référence (la langue très soutenue du T.L.F. a parfois été critiquée). Pour ces calculs sont utilisées les formules proposées notamment par Pierre Lafon⁵³⁰.

Ce programme trie d'abord les vocables en catégories grammaticales sur un tableau récapitulatif, fournit ensuite les occurrences dans les différentes catégories, et finalement recense les phrases les plus spécifiques en valeur absolue, sous-corpus par sous-corpus. Pour cette étude nous avons recours à notre corpus H de 12 œuvres en étudiant en particulier le vocabulaire du premier livre de l'œuvre, *Le procès-verbal*⁵³¹, qui couvre 9,1% du corpus total :

⁵³⁰ P. Lafon (1984). Pour une discussion de ces conventions et une présentation détaillée des calculs : C. Labbé, D. Labbé, *Que mesure la spécificité du vocabulaire ?*, Grenoble, CERAT, 1994.

⁵³¹ Une étude détaillée du vocabulaire spécifique des autres livres se trouve en annexe n°6.

Etude des spécificités du vocabulaire de *Le procès-verbal* :

Corpus de référence : II LeClezioTotal, Taille : 870 468

Sous-Corpus comparé : II 63PVTtotal, Taille : 78 783

Soit : 9.1 % du corpus total

Nombre de vocables spécifiques : 1 102

Nombre de spécificités positives : 604 soit 10 292 mots en excédent

Nombre de spécificités négatives : 498 soit 10 667 mots en déficit

Densités des catégories grammaticales

Catégories	A-B (Corpus-Sous corpus)	B Sous corpus	A/B	B-A
Noms propres	2.4	1.8	74.0	-26.0
Verbes	15.9	15.7	98.9	-1.1
<i>Formes fléchies</i>	11.0	10.5	95.2	-4.8
<i>Participes passés</i>	2.2	1.9	87.0	-13.0
<i>Participes présents</i>	0.4	0.5	122.0	+22.0
<i>Infinitifs</i>	2.2	2.8	124.0	+24.0
Substantif	19.5	19.3	99.2	-0.8
Adjectif	5.4	5.6	104.3	+4.3
<i>Adj participe passé</i>	1.1	1.2	106.6	+6.6
Pronom	12.4	13.1	105.6	+5.6
Pronoms personnels	12.4	13.1	105.6	+5.6
Déterminant	17.2	16.1	93.3	-6.7
Nombres	0.5	1.2	251.8	+151.8
Adverbes	6.9	7.2	103.5	+3.5
Prépositions	14.8	15.3	103.3	+3.3
Conjonction	5.2	5.6	106.5	+6.5
Mots étrangers	0.2	0.1	82.8	-17.2

- Colonne 1 : la catégorie grammaticale

- Colonne 2 : A. Densité de la catégorie grammaticale dans le reste du corpus
(corpus total - univers)

- Colonne 3 : B. Densité de la catégorie grammaticale dans l'univers

- colonne 4 : Rapport entre B et A (%)

- colonne 5 : 100 - (B/A)

Vocables significativement suremployés au seuil de 1% :

(Classement par catégories grammaticales et spécificité décroissante)

Noms propres : Adam, Michèle, Julienne, Pollo, Sonia, R, Américain, W, Sim, N, Hozniacks, Amadouy, Tweedsmuir, J, Prisunic, Hornatozi, Amadouay, Bosio, Guillaume, Frère, Sénégal, Corse, Carros, Beaujolais, Cécile, E. S.

Verbes : faire, dire, pouvoir, savoir, vouloir, penser, prendre, attendre, comprendre, falloir, laisser, demander, trouver, écrire, finir, répondre, fumer, importer, traîner, rappeler, acheter, ajouter, expliquer, réfléchir, fixer, foutre, écraser, fouiller, gagner, noyer, représenter, douter, intéresser, tromper, offrir, supposer, métamorphoser, noter, crever, aboutir, produire, composer, accomplir, bouffer, graver, renifler, questionner, devoir, vomir, considérer, servir, baisser, créer, prolonger, tasser, deviner, agir, réussir, passer, lire, hésiter, vivre, insister, placer, mettre, pleuvoir, prier, croire, amener, excuser, regretter, découvrir, plaire, bouillir, remercier, signaler, suffire, boire, dominer, imiter, payer, indiquer, permettre, raconter, tuer, confondre, compter, croiser, marquer.

Substantifs : chose, heure, femme, fille, chien, cigarette, sorte, rat, espèce, an, type, part, point, pluie, dos, café, papier, dieu, genre, pièce, façon, animal, chaise, lunette, bas, face, envie, exemple, minute, paquet, gauche, tas, cas, boule, lettre, journal, reste, verre, truc, foule, droite, guerre, page, galet, cahier, phrase, matière, bête, sein, raison, chair, compte, poche, objet, bar, cage, bière, million, angle, police, carton, billard, cinéma, niveau, numéro, jeu, moitié, comptoir, bouteille, crayon, barreau, cadavre, noyé, promenade, étudiant, plaisir, téléphone, parent, infirmière, système, infirmerie, élément, plancher, mesure, carré, degré, franc, pyjama, coude, merci, chienne, quart, univers, bille, connaissance, revue, écolier, tasse, asile, moyen, croix, poste, imperméable, flot, centimètre, communication, rigueur, rapport, ennui, photographie, conscience, général, bonhomme, paire, accident, raisonnement, parquet, médecin-chef, touriste, grand-chose, flic, cendrier, chocolat, unité, tiroir, baigneur, rectangle, somme, spectateur, mysticisme, samedi, propriétaire, singe, race, sensation, balustrade, rien, détente, salaud, banane, dock, guérite, bourrelet, important, atome, praline, lionne, prunelle, diable, secteur, ivoire, soutien-gorge, terreur, intérêt, laisse, style, télé, notion, personnage, échelon, morgue, choix, culture, milliard, nombril, cerisier, armoire, posture, rosier, bikini, copain, gendarme, badaud, jazz, août, demie, aisselle, musée, pose, roman, tabac, campagne, rouge, accord, vacance, bruissement, crime, peine, ordure, affiche, confiance, fonte, couche, fou, relief, fenêtre, propos, mètre, couverture, danger, amas, dentiste, loup, société, milieu, volonté, millimètre, blanc, amour, halo, mare, dessous, série, patte, fois, monstre, raie, autobus, allumette, neige, lieu, histoire, disque, villa, paysage, besoin, hauteur, réalité, forme, moment, position, urine, billet, extrémité, effet, banc, habitude, nuque, poil, acier, service, atmosphère, charbon, puissance, sens, taille, gorgée, peinture, idée, éternité, dessin.

Adjectifs : plein, bon, sûr, possible, propre, sale, simple, gauche, cher, faux, américain, blond, intelligent, ridicule, rayé, unique, psychologique, général, plastique, parallèle, juste, abstrait, microscopique, déplaisant, intéressant, tenu, divers, mental, soudain, atomique, multiple, entendu, restant, mystique, sacré, quelconque, conséquent, plan, étalé, voûté, postal, total, inutile, différent, nu, certain, creux, moindre, bouillant, humain, incapable, bizarre, joli, vu, éternel, rond, laid, horrible, occupé, jaune, délicat, rose, raide.

Pronoms : je, vous, on, tu, ça, tout, en, autre, un, quoi, toi, celui-là, ce, se, ceci.

Adverbes : pas, plus, peu, tout, oui, non, trop, pourquoi, moins, ainsi, assez, bas, environ, exactement, vaguement, combien, bon, facilement, brusquement, évidemment, parfaitement, certainement, autant, guère, mentalement, immédiatement, terriblement, exprès, visiblement, absolument, particulièrement, heureusement, naturellement, par-là, finalement, petit, plutôt, bien, soudain, tard.

Déterminants : un, deux, quelque, trois, quatre, vingt, ton, cinq, notre, mille, votre, dix, huit, quel, certain, sept, trente, six, un, cinquante, douze, quarante, zéro, quatorze, quinze, onze.

Conjonctions et prépositions : à, que, en, ou, sous, entre, chez, voilà, donc, soit, cependant, vu, après, selon, par, dès, parmi, puisque, si.

Nous trouvons une liste très semblable à celle fournie par Hyperbase s'appuyant sur le même corpus mais qui comporte toutefois quelques différences. Par exemple les dix premiers substantifs issus du logiciel Lexicométrie sont dans l'ordre : *chose, heure, femme, fille, chien, cigarette, sorte, rat, espèce* et *an*, tandis que Hyperbase fournit : *rat, type, chien, billard, espèce, heure, paquet, exemple* et *cigarette*. Comment expliquer cette différence ? Nous savons que le rôle de la lemmatisation est faible pour les substantif et il convient aussi d'ajouter que si l'on compare un peu plus loin dans les listes, on trouve à peu près les mêmes mots, mais dans un ordre un peu différent. En effet, nous trouvons toujours le même portrait de l'œuvre que nous avons déjà commentée dans l'étude précédente. Rien d'étonnant de trouver le nom propre *Adam* en tête de liste car il est très significatif de l'œuvre, lorsque, effectivement 448 des 449 emplois dans le corpus total se trouvent dans

ce livre, où *Adam* est le nom du personnage principal. En revanche, les emplois significatifs de *ça*, *cigarette*, *type* (employé dans son sens familier dans le livre), *truc*, *bouffer* ainsi que de certaines interjections donnent une tonalité particulière à l'œuvre. Il est également à noter la présence des numéros cardinaux (classés sous "déterminants") qui constitue une particularité de cet ouvrage. Le Clézio n'aime pas les chiffres et son œuvre en contient très peu. Dans ce corpus restreint c'est surtout *Adam* dans *Le procès-verbal* qui compte.

Par la même technique nous avons également extrait le vocabulaire spécifique négatif de *Le procès-verbal* en prenant comme corpus de référence la totalité du corpus H :

Vocables significativement sous-employés au seuil de 1% :

(Classement par catégories grammaticales et spécificité décroissante)

Noms propres : Lalla, Nassima, Naja Naja, Jacques, Moguer, Mondo, Surya, Suzanne, Nour, Gaspar, Aamma, Anna, Ananta, Azzar, Lullaby, Hartani, Gaby, Martin, Jon, Maurice, Giribala, Palissades, Tayar, Ma el Ainine, Roch, Suryavati, Gabriel, Pouce, Poussy, Andriamena, Quarantaine, Naman, David, Petite Croix, Sarah, Léon, Daniel, Radicz, Hawa, Liana, Zinna, Juba, NN, Beaumont, Miloz, Plate, Indien, Paoli, Ini, Bartoli, Metcalfe, Abel, Alia, Nadia, Chérif, Cité, Elisabeth, Paris, Véran.

Verbes : briller, entendre, venir, regarder, battre, courir, glisser, marcher, apparaître, entrer, disparaître, traverser, revenir, monter, serrer, brûler, avancer, avoir, éclairer, arriver, souffler, rêver, effacer, trembler, voler, aller, rouler, coucher, éloigner, chercher, vibrer, repartir, être, emplir, rester, bouger, résonner, devenir, sentir, connaître, rencontrer, voyager, fermer, frissonner, emmener, danser, revoir, changer, siffler, rire, jaillir, cacher, parler, endormir, ressentir, frapper, laver, éteindre, tourner, emporter, oublier, plonger, descendre, redescendre, troubler, filer, asseoir, préparer, approcher, réveiller, arrêter, planer, guider.

Substantifs : vent, lumière, pierre, visage, ciel, oiseau, île, vallée, mer, sable, désert, nuage, bruit, nuit, regard, fleuve, vague, nom, grand-père, rocher, dune, cœur, plaine, pays, oeil, intérieur, enfant, étoile, rivière, voix, rêve, pont, fièvre, guerrier, roche, toit, musique, mère, porte, lueur, ravin, lagon, champ, abri, broussaille, grotte, falaise, horizon, chemin, appartement, bateau,

écume, auto, océan, hutte, forêt, cheikh, vitesse, eau, entrée, ouest, récif, vertige, travers, village, lave, montagne, ombre, branche, riz, basalte, lac, troupeau, arbre, souffle, feu, odeur, immeuble, instant, peur, cour, maison, cheveu, navire, hôtel, immigrant, rive, berger, fils, sommet, larme, manteau, frisson, côté, mât, campement, insecte, oncle, tente, trésor, baleine, long, puits, respiration, brume, plateau, faim, douleur, rumeur, aube, baie, sel, volcan, place, force, poussière, ivresse, creux, nord, voyage, bûcher, silence, tempête, ruelle, flamme, roseau, or, chant, joue, tourbillon, cachette, ruine, bras, marée, mouton, palmier, est, cri, cabane, piton, chèvre, tôle, lagune, sirdar, homme, commencement, coolie, lune, corps, rue, parking, froid, lait, tour, piste, patron, coque, feuille, secret, îlot, marécage, ville, bord, source, muraille, silhouette, paille-en-queue, citerne, fleur, avenue, cuivre, poitrine, voyageur, valise, sœur, bleu, guide, fumée.

Adjectifs : grand, sombre, haut, lent, gris, vieux, long, immobile, froid, terrible, noir, brûlant, lointain, lisse, ancien, brillant, sauvage, silencieux, sec, bleu, triste, pareil, doux, invisible, enveloppé, rouge, clair, vêtu, dur, pâle, entouré, lourd, étrange, assis, pur, étroit, maigre, aigu, perdu.

Pronoms : qui, il, cela, ils, personne, y.

Adverbes : si, très, au-dessus, lentement, loin, vite, maintenant, seulement, où, jamais, debout, aussi, ici, fort, peut-être, là, ensuite, parfois, longtemps, puis, autour, quelquefois.

Déterminants : le, son, mon, leur, chaque, autre.

Conjonctions et prépositions : quand, dans, comme, jusque, sans, devant, malgré, sur, près, avec, vers, mais.

Cette liste est en quelque sorte l'image négative de l'ouvrage. La liste comporte les mots que nous trouvons dans les autres livres du corpus mais pas dans celui-ci. En tête de liste des noms propres nous trouvons les protagonistes des autres ouvrages, absents dans *Le procès-verbal*: Lalla de *Désert*, Nassima de *Hasard*, Naja Naja de *Voyages à l'autre côté* etc. Cette liste reflète aussi bien ce que nous avons déjà pu observer auparavant, que l'élément nature, thématique extrêmement importante dans l'œuvre leclézienne, n'est pas présent au début de la production et

n'interviendra qu'avec la rupture dans l'œuvre à la fin des années 1970. Les substantifs *vent, lumière, pierre, ciel, oiseau, île, vallée, mer, sable, désert, nuage, regard et fleuve* sont ici significativement absents.

La technique de segments répétés, employée dans l'analyse suivante, consiste à repérer des suites de formes graphiques revenant à l'identique au moins deux fois dans un texte et permet ainsi d'extraire les phrases les plus significatives d'un texte par rapport aux autres textes dans un corpus⁵³².

Ci-dessous nous avons classé les cinq phrases les plus significatives de *Le procès-verbal*, extraites d'abord en s'appuyant sur les valeurs absolues, ensuite sur les valeurs relatives :

Phrases les plus spécifiques en valeur absolue (avec leurs scores).

42 : CM proces-verbal : Je voulais faire un jeu comme, un, deux, trois, quatre, cinq, six, sept, huit, neuf, dix, onze, douze, tu y es ? treize, quatorze, quinze, seize, dix-sept, dix-huit, tu y es ? dix-neuf, vingt, vingt-et-un, vingt-deux, vingt-trois, vingt-quatre, je compte jusqu'à trente, vingt-cinq, vingt-six, vingt-sept, vingt-huit, vingt-neuf, vingt-neuf et demi, vingt-neuf trois quarts, et, et, 30 ! et puis chercher partout dans la ville.

26 : CM proces-verbal : J'imagine qu'il va falloir passer sous peu devant un tribunal d'hommes ; je leur laisse ces ordures en guise de testament ; sans orgueil, j'espère qu'on me condamnera à quelque chose, afin que je paye de tout mon corps la faute de vivre ; si on m'humilie, si on me fouette, si on me crache au visage, j'aurai enfin une destinée, je croirai enfin en Dieu.

21 : CM proces-verbal : N'oubliez pas avant tout que si je pouvais trouver un travail quelconque à faire, quelque chose de peu absorbant, un truc manuel,

⁵³² Les procédures de repérage des segments répétés sont issues d'un travail réalisé à l'ENS de Fontenay-St. Cloud, cf. A. Salem (1987).

plongeur au restaurant, habilleur à la morgue, ou figurant aux studios, je m'en contenterais.

19 : CM proces-verbal : Nous pensions qu'après les récentes fatigues de ton service, tu allais pouvoir te reposer auprès de nous – nous pensions aller tous ensemble pendant quelque temps à la campagne chez ta tante – nous n'en avons pas trop parlé, évidemment, mais tu semblais fatigué depuis un moment et je savais que tu n'aimes pas faire de projets.

19 : CM proces-verbal : Hier je suis allée au cinéma ; c'était un film bizarre, mais ça m'a donné envie de parler ; je pense que tu gaspilles ton temps à des choses sans intérêt ; tu te gaspilles toi-même ; tu n'aboutiras à rien ; tu as peur de tout ce qui est sentimental ; moi j'ai envie de te raconter une histoire.

Phrases les plus spécifiques en valeur relative (avec leur scores)⁵³³.

0.882 : CM proces-verbal : Vous viviez par deux, par trois, quatre, cinq, six, dix, vingt-neuf, cent quatre vingt - trois, et cetera.

0.769 : CM proces-verbal : Dans quel cas, je vous dirais oui, ou non, ou rien, peu importe.

0.750 : CM proces-verbal : “ Eh bien, ces histoires - toutes ces histoires qu'on” - “attendez !” dit Adam.

0.750 : CM proces-verbal : Vous ne voulez pas prendre un verre, en attendant votre bonne femme ?

0.727 : CM proces-verbal : Vous me raconterez une histoire de vos vacances, de votre enfance.

Le fait de prendre la valeur absolue ou la valeur relative ne change pas tellement ces phrases, nous retrouvons les mots de spécificité significativement positive et

⁵³³ L'indice relatif : Indice = (N liaisons positives - N liaisons négatives)/longueur de la phrase. NB : par construction, l'indice favorise les phrases brèves construites sous forme de slogans.

les phrases témoignent aussi du style particulier du livre caractérisé par son langage familier et ses nombreuses ponctuations expressives.

Toutefois, la fréquence de ces mots significatifs d'une œuvre à l'autre n'évolue pas de la même manière. L'étude de l'évolution lexicale permet de mieux comprendre les mouvements thématiques de l'œuvre.

5.3.5. L'évolution du vocabulaire.

Pour observer l'évolution du vocabulaire de Le Clézio, nous allons nous livrer à l'étude de la corrélation chronologique au niveau endogène, laissant cette fois-ci toute comparaison externe de côté. Il convient de rappeler qu'il ne s'agit, dans cette analyse, que du contenu lexical, les phénomènes des structures lexicales, syntaxiques ou rythmiques où pourtant l'évolution de J.M.G. Le Clézio est très perceptible étant laissés de côté.

En effet, étant donné que les textes qui constituent notre corpus s'échelonnent dans le temps et sont classés en ordre chronologique, le coefficient de corrélation peut être calculé par le logiciel Hyperbase qui compare, pour chaque mot, les valeurs de l'écart réduit au rang de chaque élément et établit le coefficient de Bravais-Pearson⁵³⁴ quand la fréquence d'une forme est suffisante pour permettre le calcul de l'écart réduit.

Le tableau ci-dessous regroupe tous les mots qui atteignent un seuil de 0,35 au nombre de textes considérés. Les deux colonnes reproduisent, d'abord, la liste des mots que Le Clézio emploie de plus en plus et, ensuite, ceux qu'il abandonne progressivement :

⁵³⁴ Pour plus de détails sur ce calcul, cf. Ch. Muller (1973) : p. 157-160.

Mots en progression											
Coeff.	Fréq.	Mot									
+ 0.753	317	arrivée	+ 0.514	30	accompagné	+ 0.451	120	mari	+ 0.419	80	indiennes
+ 0.741	806	première	+ 0.512	35	costume	+ 0.450	590	matin	+ 0.418	46	justement
+ 0.694	345	donné	+ 0.511	65	appelé	+ 0.450	93	étrave	+ 0.417	79	tante
+ 0.685	9489	pour	+ 0.511	54	rejoint	+ 0.449	378	voyage	+ 0.415	93	souvenait
+ 0.682	80	magnifique	+ 0.510	53	accent	+ 0.449	44	cachait	+ 0.415	75	tombes
+ 0.671	2737	jusqu'	+ 0.510	38	rêvait	+ 0.448	1124	porte	+ 0.415	72	
+ 0.666	112	France	+ 0.510	33	jetés	+ 0.448	1103	déjà			compagnie
+ 0.643	501	malgré	+ 0.509	207	saint	+ 0.448	88	proue	+ 0.415	52	apporter
+ 0.637	153	décidé	+ 0.509	118	restée	+ 0.448	83	rêver	+ 0.415	36	bâtiment
+ 0.628	107	auprès	+ 0.508	9004	son	+ 0.447	49	étrangement	+ 0.414	52	annonce
+ 0.623	74	raconté	+ 0.508	88	reçu	+ 0.446	346	filles	+ 0.413	37	bagages
+ 0.614	348	marché	+ 0.507	31	ajouté	+ 0.446	132	san	+ 0.412	154	vécu
+ 0.613	77	rêvé	+ 0.506	53	visite	+ 0.446	100	linge	+ 0.412	93	entendais
+ 0.612	72	apporté	+ 0.505	50	enfermée	+ 0.446	68	huttes	+ 0.412	90	partis
+ 0.611	57	riches	+ 0.503	333	partie	+ 0.446	65	chapeau	+ 0.412	88	jeter
+ 0.610	41	étonnement	+ 0.503	85	gardé	+ 0.446	34	occupait	+ 0.411	109	fatiguée
+ 0.609	379	entrée	+ 0.502	158	grâce	+ 0.444	75	penchée	+ 0.409	161	départ
+ 0.609	180	famille	+ 0.501	177	océan	+ 0.444	30	moustache	+ 0.409	39	
+ 0.606	71	montée	+ 0.501	53	accrochée	+ 0.443	33	cherchais			américains
+ 0.598	59489	la	+ 0.498	209	pensé	+ 0.442	423	mis	+ 0.408	228	
+ 0.593	45	âpre	+ 0.497	130	enfance	+ 0.442	53	mariage			recherche
+ 0.587	137	montré	+ 0.497	58	allongée	+ 0.442	48	écoutais	+ 0.408	207	laquelle
+ 0.587	136	mise	+ 0.495	55	emmené	+ 0.442	40	pleurait	+ 0.408	44	arabe
+ 0.586	133	hôpital	+ 0.492	75	impatience	+ 0.441	154	devenue	+ 0.408	41	conduite
+ 0.585	52	laissée	+ 0.491	200	dernier	+ 0.441	129	serrée	+ 0.407	372	avons
+ 0.582	46	préparer	+ 0.491	98	affaires	+ 0.440	971	cœur	+ 0.407	107	vêtue
+ 0.581	127	soeur	+ 0.488	81	jeté	+ 0.440	66	mauvaise	+ 0.407	51	luxe
+ 0.575	201	mémoire	+ 0.488	39	aidé	+ 0.439	155	ancien	+ 0.406	32276	l'
+ 0.573	36	médecins	+ 0.487	35	emmenait	+ 0.439	42	houle	+ 0.406	46	levée
+ 0.571	102	suivi	+ 0.485	60	jetée	+ 0.438	311	nouvelle	+ 0.405	688	chez
+ 0.569	622	sombre	+ 0.484	634	enfant	+ 0.438	87	paraissait	+ 0.405	447	partir
+ 0.568	51	amoureux	+ 0.484	47	traversée	+ 0.438	79	crainte	+ 0.405	98	rafales
+ 0.562	31	épreuve	+ 0.484	30	battu	+ 0.438	57	préparé	+ 0.405	48	emmenée
+ 0.559	180	changé	+ 0.481	133	revenu	+ 0.437	348	colère	+ 0.404	37	montés
+ 0.558	97	rencontré	+ 0.481	52	clef	+ 0.437	259	arrivé	+ 0.404	31	protéger
+ 0.554	493	passé	+ 0.479	138	thé	+ 0.437	92	frontière	+ 0.403	226	fièvre
+ 0.554	71	ressenti	+ 0.478	35	amitié	+ 0.437	41	commandant	+ 0.403	75	voyageurs
+ 0.552	923	mère	+ 0.476	69	monté	+ 0.436	79	anciennes	+ 0.403	62	
+ 0.552	141	jusque	+ 0.474	235	retourner	+ 0.436	42	leçons			enveloppée
+ 0.552	57	voyant	+ 0.474	48	emmener	+ 0.436	38	pension	+ 0.403	39	projet
+ 0.550	85	couchée	+ 0.472	56	recommencé	+ 0.435	32	autel	+ 0.402	699	disait
+ 0.549	59	endormie	+ 0.472	36	retournés	+ 0.434	259	hôtel	+ 0.402	204	demandé
+ 0.549	32	travaux	+ 0.471	220	premiers	+ 0.434	42	échange	+ 0.402	99	tombée
+ 0.546	41	jadis	+ 0.471	164	Paris	+ 0.434	38	vivres	+ 0.402	43	attiré
+ 0.545	17654	du	+ 0.470	425	sud	+ 0.434	36	déjeuner	+ 0.401	321	robe
+ 0.542	32	errant	+ 0.470	109	appuyée	+ 0.434	32	coiffé	+ 0.401	272	dû
+ 0.540	272	venus	+ 0.470	69	palmes	+ 0.433	95	violent	+ 0.401	33	habillée
+ 0.539	89	montrer	+ 0.469	172	retour	+ 0.433	34	terrifiant	+ 0.399	12621	au
+ 0.537	120	racontait	+ 0.469	163	derniers	+ 0.432	90	pleurer	+ 0.399	157	légende
+ 0.537	64	bijoux	+ 0.469	36	coiffés	+ 0.432	46	nausée	+ 0.398	136	étions
+ 0.537	46	enlevé	+ 0.469	31	haussé	+ 0.431	6839	sa	+ 0.398	46	mât
+ 0.536	39	coupé	+ 0.468	526	parlait	+ 0.431	266	côte	+ 0.398	34	retrait
+ 0.534	74	différent	+ 0.467	81	attente	+ 0.431	39	interdit	+ 0.397	215	îles
+ 0.533	351	premier	+ 0.467	44	musée	+ 0.429	116	dormi	+ 0.397	62	occuper
+ 0.533	161	français	+ 0.466	31899	à	+ 0.429	106	quai	+ 0.397	43	honneur
+ 0.532	88	empêcher	+ 0.466	201	laissé	+ 0.429	59	soin	+ 0.397	41	flanc
+ 0.532	54	touché	+ 0.465	124	prière	+ 0.428	198	donnait	+ 0.396	480	nord
+ 0.530	146	oncle	+ 0.465	108	rive	+ 0.428	106	hutte	+ 0.396	121	cru
+ 0.528	1887	après	+ 0.464	108	étroite	+ 0.428	75	reconnu	+ 0.396	60	
+ 0.528	247	aube	+ 0.464	33	gâteaux	+ 0.426	89				auparavant
+ 0.526	98	rives	+ 0.462	333	village			lendemain	+ 0.396	33	grimace
+ 0.526	65	semaines	+ 0.462	231	venue	+ 0.426	46	rade	+ 0.395	142	tenait
+ 0.526	57	écart	+ 0.462	42	coupée	+ 0.425	139	bébé	+ 0.394	87	habitants
+ 0.523	58	sœurs	+ 0.460	180	commencement	+ 0.425	30	nauffrage	+ 0.394	67	endormi
+ 0.522	62	retournée	+ 0.460	152	indienne	+ 0.424	58	échelle	+ 0.393	425	pont
+ 0.521	146	prise	+ 0.459	85	Juan	+ 0.423	88	attendu	+ 0.393	168	appris
+ 0.519	90	laver	+ 0.458	7128	par	+ 0.423	46	maîtresse	+ 0.393	162	épaule
+ 0.518	1080	nom	+ 0.457	188	expression	+ 0.423	45	allés	+ 0.393	100	quartier
+ 0.518	426	venu	+ 0.456	35	méchanceté	+ 0.422	1582	contre	+ 0.393	89	couru
+ 0.518	39	teint	+ 0.455	121	pénombre	+ 0.422	409	étrange	+ 0.393	51	coussins
+ 0.514	413	juste	+ 0.454	186	âge	+ 0.421	362	navire	+ 0.393	48	apportait
+ 0.514	30	débarqué	+ 0.454	42	études	+ 0.421	174	revenir	+ 0.393	41	quartiers
			+ 0.453	671	soir	+ 0.421	141	maigre	+ 0.392	264	entendu
			+ 0.453	271	parlé	+ 0.421	65	violente	+ 0.392	76	scène
			+ 0.453	113	aide	+ 0.421	36	exil	+ 0.392	64	pêcheurs
			+ 0.452	154	armée	+ 0.420	52	ère	+ 0.392	64	connaissais
			+ 0.452	45	réveillée	+ 0.419	506	bateau	+ 0.392	39	européens

+ 0.392 37 portée	+ 0.370 65 horreur	- 0.700 56 vibrations	- 0.608 32 rondes
+ 0.392 35 chapelle	+ 0.370 59 continent	- 0.700 46 noeud	- 0.607 588 bouche
+ 0.392 34 chaux	+ 0.370 46 Angleterre	- 0.699 80 plaque	- 0.607 65
+ 0.391 825 depuis	+ 0.370 34 explique	- 0.699 43 allumette	imperméable
+ 0.391 87 éclairé	+ 0.370 33 enthousiasme	- 0.697 30 croûte	- 0.607 60 cerveau
+ 0.391 54 appelaient	+ 0.369 834 vu	- 0.693 182 reflets	- 0.607 48 surfaces
+ 0.391 34 jus	+ 0.369 86 aimais	- 0.685 146 minuscules	- 0.606 57 minute
+ 0.390 20049 elle	+ 0.369 71 dangereux	- 0.683 68 bille	- 0.605 785 intérieur
+ 0.390 112 forte	+ 0.368 194 hiver	- 0.680 974 choses	- 0.604 212 taches
+ 0.390 80 mâts	+ 0.368 152 tourné	- 0.680 202 voilà	- 0.603 76
+ 0.390 73 passée	+ 0.368 39 tribus	- 0.671 204 mouvements	caoutchouc
+ 0.390 70 espagnol	+ 0.367 178 senti	- 0.671 69 fixes	- 0.603 35 craie
+ 0.389 105 époque	+ 0.367 151	- 0.668 3537 ;	- 0.603 33 écraser
+ 0.388 1363 maison	crépuscule	- 0.666 241 soudain	- 0.602 63 2
+ 0.388 488 longue	+ 0.367 46 osé	- 0.666 37 tremblant	- 0.601 381 face
+ 0.388 323 jeunes	+ 0.366 303 ouest	- 0.664 34 miettes	- 0.601 314 genre
+ 0.388 200 cour	+ 0.366 240 parti	- 0.663 305 espèce	- 0.601 188 odeurs
+ 0.388 33 jupe	+ 0.366 194 garde	- 0.662 456 vit	- 0.598 32 milliards
+ 0.387 128 lait	+ 0.366 64 posée	- 0.662 113 boule	- 0.597 453 droite
+ 0.387 125 arrivés	+ 0.366 55 répétait	- 0.661 44 index	- 0.597 34 passants
+ 0.387 68 rituel	+ 0.363 2001 visage	- 0.661 36 réguliers	- 0.595 150 plafond
+ 0.386 431 autrefois	+ 0.363 64 corde	- 0.660 224 sûr	- 0.595 70 obscure
+ 0.386 173 arrêté	+ 0.363 49 entourent	- 0.660 107 rond	- 0.595 65 repos
+ 0.386 50 ressentait	+ 0.363 39 mêlait	- 0.659 58 couches	- 0.594 354 cigarette
+ 0.386 46 restais	+ 0.362 437 sommes	- 0.659 50 nappe	- 0.594 99 phrases
+ 0.386 32 dîner	+ 0.361 100 essayé	- 0.657 92 bulles	- 0.594 63 habiter
+ 0.385 55 couchant	+ 0.361 79 cents	- 0.657 78 ongles	- 0.594 34 avidement
+ 0.385 43 précieux	+ 0.361 33 couper	- 0.657 36 espaces	- 0.593 62 parts
+ 0.385 31 remonté	+ 0.360 103 rejoindre	- 0.656 754 partout	- 0.593 33 égout
+ 0.384 403 commencé	+ 0.360 84 regardais	- 0.656 38 particules	- 0.592 189 arrêt
+ 0.384 89 née	+ 0.360 82 rires	- 0.654 836 sol	- 0.592 113 miroirs
+ 0.384 45 tribu	+ 0.360 34 méchant	- 0.654 45 imperceptiblement	- 0.592 55 poussières
+ 0.384 34 bouclés	+ 0.359 91 Sarah	- 0.653 91 boules	- 0.590 194 exemple
+ 0.383 379 voulu	+ 0.359 56 chance	- 0.651 486 enfin	- 0.590 67 obscur
+ 0.383 149 voiles	+ 0.358 39 italienne	- 0.648 40 centimètres	- 0.589 167 masse
+ 0.383 98 premières	+ 0.358 32 reviendrait	- 0.647 117 liquide	- 0.588 362 jaune
+ 0.382 449 surtout	+ 0.358 31 amer	- 0.645 215 chair	- 0.587 386 plein
+ 0.382 53 payer	+ 0.357 141 su	- 0.645 129 goutte	- 0.587 92 dedans
+ 0.382 49 ajoute	+ 0.357 46 versant	- 0.645 111 absolument	- 0.585 82 Voyez
+ 0.382 35 brisée	+ 0.357 32 commun	- 0.643 1115 seul	- 0.584 420 gauche
+ 0.382 31 biscuits	+ 0.357 31 bracelets	- 0.643 232 seconde	- 0.581 47 tonnes
+ 0.381 89 chantait	+ 0.356 554 bois	- 0.640 36 obscures	- 0.580 74 essence
+ 0.381 45 José	+ 0.356 461 midi	- 0.639 316 trottoir	- 0.579 5525 ?
+ 0.381 41 fortune	+ 0.356 230 lisse	- 0.638 862 blanc	- 0.579 453
+ 0.379 145 lorsqu'	+ 0.356 69 mexicaine	- 0.637 76 muscles	doucement
+ 0.379 112 serrait	+ 0.356 57 usé	- 0.636 32 intensément	- 0.579 97 fixe
+ 0.379 76 revenue	+ 0.356 56 cinquante	- 0.634 90 poumons	- 0.579 68 flot
+ 0.379 50 bêche	+ 0.356 36 emmêlés	- 0.634 75 continuellement	- 0.577 54 lueurs
+ 0.378 55 Inde	+ 0.355 284 port	- 0.633 119 rides	- 0.576 195 morceau
+ 0.378 39 brouhaha	+ 0.355 124 déesse	- 0.631 108 phrase	- 0.576 87 béton
+ 0.378 37 sienne	+ 0.354 129 serré	- 0.630 9412 se	- 0.575 71 nappes
+ 0.377 188 Américque	+ 0.354 66 jeunesse	- 0.630 1976 autres	- 0.574 34 semelles
+ 0.377 58 imaginé	+ 0.354 33 dauphins	- 0.630 60 antennes	- 0.574 32 liquides
+ 0.377 35 cahiers	+ 0.353 87 estuaire	- 0.624 94 crayon	- 0.573 283 couleurs
+ 0.376 286 assise	+ 0.353 37 instinct	- 0.624 83 bloc	- 0.570 3590 rien
+ 0.376 209 voile	+ 0.353 33 planche	- 0.622 48 faces	- 0.570 64 roc
+ 0.376 203 indien	+ 0.353 32 récits	- 0.622 43 fonte	- 0.570 64 éteindre
+ 0.376 108 docteur	+ 0.352 108 honte	- 0.621 45 ronds	- 0.570 64 arracher
+ 0.376 49 entrés	+ 0.352 40 emmène	- 0.620 42 mouche	- 0.569 192 électrique
+ 0.376 34 faute	+ 0.352 39 guetter	- 0.620 35 automobiles	- 0.569 57 veines
+ 0.375 206 resté	+ 0.351 52 châte	- 0.617 121 gaz	- 0.569 37 gouffres
+ 0.375 132 rôle	+ 0.351 50 Nice	- 0.616 227 tas	- 0.566 567 petits
+ 0.375 71 crié	+ 0.351 34 déluge	- 0.616 180 mètres	- 0.565 126 drôles
+ 0.375 47 foulard	+ 0.351 30 entendu	- 0.616 103 sons	- 0.564 145 minutes
+ 0.375 39 allions		- 0.616 57 indéfiniment	- 0.564 45 écrasé
+ 0.374 212 larmes	Mots en régression	- 0.615 1583 toutes	- 0.564 42 molle
+ 0.374 147 sentais		- 0.615 88 ceci	- 0.564 36 raides
+ 0.374 34 travailleurs	Coeff. Fréq. Mot	- 0.615 71 marbre	- 0.563 69 10
+ 0.373 2501 ont		- 0.613 133 sale	- 0.563 36 terne
+ 0.373 51 arabes	- 0.754 407 papier	- 0.612 30 affiches	- 0.562 94 étendu
+ 0.373 41 esclavage	- 0.746 207 pattes	- 0.610 7833 tout	- 0.562 61 cendrier
+ 0.372 139 voyais	- 0.745 91 minuscule	- 0.610 1201 non	- 0.562 39 organes
+ 0.372 82 moderne	- 0.733 150 rayons	- 0.610 585 bruits	- 0.560 78 bizarres
+ 0.371 83 nomades	- 0.733 91 narines	- 0.610 63 allumettes	- 0.560 46 écran
+ 0.371 31 modernes	- 0.720 188 secondes	- 0.610 43 sphère	- 0.559 9221 plus
+ 0.370 20669 dans	- 0.710 124 unes	- 0.609 340 verre	- 0.559 276 surface
+ 0.370 169 tempête	- 0.706 666 oui	- 0.608 45 explosions	- 0.559 196 millions

- 0.558	70	centaines	- 0.516	227	objets	- 0.481	123	bouts	- 0.454	97	réverbères
- 0.557	79	carrés	- 0.516	95	trottoirs	- 0.481	74	assit	- 0.453	188	cigarettes
- 0.556	110	fourmis	- 0.516	61	douloureux	- 0.480	139	roues	- 0.453	31	jeta
- 0.556	53	parcelles	- 0.515	253	importe	- 0.480	35	photographies	- 0.452	540	peine
- 0.556	53	espèces	- 0.515	239	poids	- 0.479	168	galets	- 0.452	190	larges
- 0.556	40	solide	- 0.515	31	excréments	- 0.479	48	rangées	- 0.452	84	mortes
- 0.555	90	flotte	- 0.514	55	plâtre	- 0.479	44	6	- 0.452	44	
- 0.555	49	droites	- 0.513	3168	puis	- 0.479	41	éclater			évidemment
- 0.554	99	glace	- 0.513	41	verticale	- 0.479	38	balle	- 0.452	42	douter
- 0.553	131	bouches	- 0.512	1566	...	- 0.478	588	savoir	- 0.452	33	tentacules
- 0.553	114	marcha	- 0.512	95	éclair	- 0.478	511	jambes	- 0.451	253	pleine
- 0.553	60	terrains	- 0.512	39	respirant	- 0.477	1015	vite	- 0.451	166	siècles
- 0.552	75	brusquement	- 0.511	285	arrêter	- 0.477	71	hein	- 0.451	54	opaque
- 0.552	67	ampoule	- 0.511	89	précis	- 0.476	364	fenêtres	- 0.450	588	heures
- 0.550	335	regarda	- 0.511	65	régulièrement	- 0.476	109	montent	- 0.450	190	avançait
- 0.549	36	luisant	- 0.511	55	sales	- 0.476	106	oh	- 0.450	187	infini
- 0.548	35	mobiles	- 0.511	42	répéta	- 0.474	514	cesse	- 0.450	114	heureux
- 0.548	34	rigueur	- 0.510	307	soi	- 0.474	305	assez	- 0.450	44	cubes
- 0.547	220	quelle	- 0.510	32	cognant	- 0.474	132	tranquille	- 0.449	566	part
- 0.547	160	ah	- 0.509	112	avancent	- 0.474	112	pleines	- 0.449	78	tourna
- 0.545	78	néon	- 0.509	34	inlassablement	- 0.474	106	lourds	- 0.449	44	nylon
- 0.544	457	voitures	- 0.508	75	chiffres	- 0.474	62	lumineux	- 0.448	4186	être
- 0.544	89	passa	- 0.507	2139	corps	- 0.473	189	ombres	- 0.448	129	voulez
- 0.543	352	petites	- 0.507	1850	air	- 0.473	141	journaux	- 0.448	70	fallu
- 0.543	184	pleins	- 0.507	610	espace	- 0.473	59	places	- 0.448	40	suffisait
- 0.543	171	sortes	- 0.507	42	mètre	- 0.473	54	substance	- 0.448	37	chiffre
- 0.543	82	gouffre	- 0.507	30	monta	- 0.473	46	attendit	- 0.447	231	dur
- 0.541	34	obstacles	- 0.506	98	rapidement	- 0.473	44	bribes	- 0.447	112	poche
- 0.540	46	vitrine	- 0.506	57	incendie	- 0.472	37	rétines	- 0.447	32	clos
- 0.539	35	araignées	- 0.506	56	ordures	- 0.471	268	milliers	- 0.447	30	reprit
- 0.538	847	faut	- 0.505	34	mou	- 0.471	82	tournant	- 0.446	257	lunettes
- 0.538	487	quoi	- 0.504	52	parquet	- 0.470	2633	ça	- 0.446	42	ventres
- 0.538	109	ondes	- 0.503	117	pneus	- 0.469	3936	bien	- 0.446	40	étages
- 0.538	92	peaux	- 0.501	122	exactement	- 0.469	118	sortit	- 0.445	280	terrible
- 0.538	58	nœuds	- 0.501	93	leva	- 0.469	78	bête	- 0.445	154	avion
- 0.538	41	hésita	- 0.501	46	sourit	- 0.469	65	rage	- 0.445	118	type
- 0.537	51	flottent	- 0.500	54	encre	- 0.469	40	écrase	- 0.445	88	légèrement
- 0.537	41	surgir	- 0.500	40	bulle	- 0.468	30	demie	- 0.445	38	séries
- 0.537	34	gras	- 0.500	31	intéresse	- 0.467	903	murs	- 0.444	161	vitres
- 0.537	31	fondu	- 0.499	96	bière	- 0.467	129	sûrement	- 0.444	147	arrêta
- 0.536	3590	vous	- 0.499	31	normal	- 0.466	90	ouvertes	- 0.444	48	presser
- 0.535	156	pensées	- 0.498	289	matière	- 0.466	51	vibrait	- 0.443	929	centre
- 0.534	87	durs	- 0.498	113	phares	- 0.465	42	grandir	- 0.443	95	invisibles
- 0.534	45	traversa	- 0.497	9158	on	- 0.464	13571	pas	- 0.443	94	action
- 0.534	39	écoula	- 0.497	1551	quelque	- 0.463	868	derrière	- 0.443	68	aurai
- 0.533	93	paquet	- 0.495	320	foule	- 0.463	458	tellement	- 0.443	56	000
- 0.533	46	flots	- 0.494	127	but	- 0.463	290	ciment	- 0.443	50	
- 0.533	38	aimerais	- 0.494	77	tranquille	- 0.463	88	entendit			brutalement
- 0.532	62	flottait	- 0.494	77	ponts	- 0.463	85	cafés	- 0.442	32	carapaces
- 0.532	44	gravier	- 0.494	40	vibrant	- 0.463	38	heureusement	- 0.441	59	
- 0.531	267	gouttes	- 0.494	30	faiblement	- 0.462	1581	chose			parfaitement
- 0.530	32	monstres	- 0.492	70	flaques	- 0.462	56	rebord	- 0.441	38	vivantes
- 0.528	31	klaxons	- 0.491	4477	ou	- 0.462	35	écoutez	- 0.441	36	lézards
- 0.527	5805	sans	- 0.491	188	feuille	- 0.462	33	labyrinthe	- 0.441	31	répandait
- 0.527	307	métal	- 0.491	45	relief	- 0.461	141	cercles	- 0.440	131	épais
- 0.527	49	alluma	- 0.491	45	flotter	- 0.461	125	électriques	- 0.440	32	mégot
- 0.527	47	cavernes	- 0.490	108	grains	- 0.461	58	écartant	- 0.439	268	écrire
- 0.527	45	pencha	- 0.490	76	poils	- 0.460	82	parois	- 0.439	104	bougent
- 0.527	42	soleils	- 0.490	53	entraîlles	- 0.460	36	trainaient	- 0.439	46	infinie
- 0.526	96	danger	- 0.489	1516	dire	- 0.459	41	mâchoires	- 0.439	40	échos
- 0.525	111	facilement	- 0.489	59	explosion	- 0.459	37	troubles	- 0.439	35	quatrième
- 0.524	535	mouvement	- 0.489	58	durer	- 0.458	2458	tous	- 0.438	126	savez
- 0.524	151	plaques	- 0.488	77	glaces	- 0.458	143	machine	- 0.438	58	comprenez
- 0.524	45	carrosseries	- 0.488	53	3	- 0.458	58	nerfs	- 0.438	40	posés
- 0.524	40	balustrade	- 0.488	51	souple	- 0.457	318	calme	- 0.438	30	portières
- 0.523	937	peau	- 0.487	241	mit	- 0.457	147	passaient	- 0.437	115	spectacle
- 0.523	324	cailloux	- 0.486	82	angles	- 0.457	77	entra	- 0.437	110	caillou
- 0.523	69	continua	- 0.485	122	fou	- 0.457	46	orbites	- 0.437	68	durs
- 0.522	109	1	- 0.485	76	sourd	- 0.456	8400	y	- 0.437	62	hauteur
- 0.522	106	effort	- 0.485	36	dressées	- 0.456	84	tables	- 0.437	45	profonds
- 0.520	469	beau	- 0.484	75	humaines	- 0.456	66	vitrites	- 0.436	40	verrait
- 0.520	65	série	- 0.483	904	vide	- 0.455	398	blancs	- 0.435	271	oeil
- 0.520	59	monnaie	- 0.483	57	fût	- 0.455	32	obscur	- 0.435	136	grosse
- 0.518	187	dessin	- 0.482	160	mouches	- 0.454	391	fallait	- 0.435	82	numéro
- 0.518	103	murailles	- 0.482	79	merci	- 0.454	272	avez	- 0.435	55	5
- 0.518	45	cellules	- 0.482	34	mécanique	- 0.454	231	possible	- 0.435	39	sauter
- 0.516	269	trous	- 0.481	619	veux	- 0.454	104	bougeait	- 0.434	216	immobiles

- 0.434	207	donc	- 0.420	50	rythmes	- 0.402	43	carrefours	- 0.380	55	boutons
- 0.434	95	talons	- 0.420	34	cassés	- 0.402	31	rocs	- 0.379	351	table
- 0.434	79	etc	- 0.420	30	vertigineux	- 0.401	1349	fond	- 0.379	168	
- 0.433	625	chaleur	- 0.419	6903	mais	- 0.401	259	uns			impossible
- 0.433	55	vase	- 0.419	171	autobus	- 0.401	54	traversaient	- 0.379	157	sentit
- 0.433	44	rails	- 0.419	33	aile	- 0.401	34	15	- 0.379	106	kilomètres
- 0.432	169	miroir	- 0.418	321	présent	- 0.400	954	milieu	- 0.379	79	pâles
- 0.432	125	facile	- 0.418	56	fixés	- 0.400	69	prisonnier	- 0.379	74	fermées
- 0.432	74	carrefour	- 0.418	51	humides	- 0.400	60	vieillesse	- 0.378	231	gestes
- 0.431	477	serait	- 0.418	35	douleurs	- 0.400	43	paysages	- 0.377	489	veut
- 0.431	430	as	- 0.416	163	aussitôt	- 0.399	246	bientôt	- 0.377	100	tourbillons
- 0.431	207	vitesse	- 0.416	42	trainant	- 0.399	216	quel	- 0.377	99	roulent
- 0.431	118	rapide	- 0.415	1107	mots	- 0.399	171	morceaux	- 0.377	30	systèmes
- 0.431	42	terriblement	- 0.415	130	éclats	- 0.399	57	Allô	- 0.376	34	boulevards
- 0.431	33	habité	- 0.415	64	creuse	- 0.399	55	froids	- 0.375	102	voulaient
- 0.431	30	gel	- 0.415	59	oubli	- 0.398	148	joie	- 0.375	99	environ
- 0.430	496	heure	- 0.414	136	orange	- 0.398	101	faux	- 0.375	70	colonnes
- 0.430	244	lettres	- 0.414	72	lame	- 0.398	47	perdait	- 0.374	257	paysage
- 0.430	239	nez	- 0.414	57	fissures	- 0.397	43	étale	- 0.374	78	transparent
- 0.430	75	rideaux	- 0.414	49	voulut	- 0.396	3591	!	- 0.373	170	veulent
- 0.429	127	nuque	- 0.414	33	immobilité	- 0.396	50	pare	- 0.373	38	incompréhensibles
- 0.429	36	12	- 0.413	696	vraiment	- 0.396	47	plats	- 0.372	872	sais
- 0.428	125	boîtes	- 0.413	430	point	- 0.396	30	hé	- 0.371	139	parcelles
- 0.428	71	désirs	- 0.413	387	fer	- 0.395	60	poteaux	- 0.371	105	commença
- 0.428	60	devenaient	- 0.413	147	os	- 0.395	40	arraché	- 0.371	98	vitre
- 0.428	40	trombes	- 0.413	139	complètement	- 0.395	30	fausses	- 0.371	67	insecte
- 0.428	37	mirent	- 0.413	103	têtes	- 0.394	1844	faire	- 0.371	60	gagner
- 0.428	33	enfoncés	- 0.413	77	atmosphère	- 0.394	100	glacé	- 0.370	17301	en
- 0.428	32	aimez	- 0.413	59	crevasses	- 0.394	60	dense	- 0.370	387	mêmes
- 0.428	31	cirque	- 0.413	32	tasse	- 0.394	44	plis	- 0.370	254	eut
- 0.428	30	bitume	- 0.412	68	infiniment	- 0.394	32	intelligent	- 0.370	75	vis
- 0.427	225	conscience	- 0.412	45	disparaissaient	- 0.393	10758	n'	- 0.370	40	vies
- 0.427	203	tache	- 0.412	30	tuyaux	- 0.393	32	cadavre	- 0.369	80	limite
- 0.427	151	chaussée	- 0.411	217	invisible	- 0.392	104	vivants	- 0.369	72	arrêtent
- 0.427	125	demanda	- 0.411	185	dents	- 0.392	51	caverne	- 0.369	69	bougeaient
- 0.427	82	crâne	- 0.411	184	cercle	- 0.392	47	halo	- 0.369	65	barres
- 0.427	54	expliquer	- 0.411	33	grain	- 0.391	140	faudrait	- 0.368	2057	deux
- 0.427	33	écarter	- 0.411	32	griffes	- 0.389	173	dessous	- 0.368	57	pans
- 0.426	57	lutter	- 0.411	32	agir	- 0.388	288	dure	- 0.368	47	gonflé
- 0.425	12634	ne	- 0.410	174	plat	- 0.388	169	étincelles	- 0.368	33	recommençait
- 0.425	59	inconnue	- 0.409	786	comment	- 0.388	60	dites	- 0.368	32	suivit
- 0.424	193	faits	- 0.409	468	droit	- 0.388	47	écartent	- 0.367	2849	temps
- 0.424	111	angle	- 0.409	77	lourdes	- 0.387	221	lumières	- 0.367	232	fit
- 0.424	106	ch	- 0.409	31	lenteur	- 0.387	131	acier	- 0.366	191	paupières
- 0.424	32	referme	- 0.408	361	pensée	- 0.387	84	aimer	- 0.366	66	sein
- 0.424	32	courbes	- 0.408	246	étendue	- 0.387	72	clairs	- 0.366	51	cachent
- 0.423	239	paix	- 0.407	146	dessins	- 0.387	32	inscrit	- 0.366	45	quantités
- 0.423	216	bouge	- 0.407	119	essayer	- 0.386	363	bon	- 0.366	35	retard
- 0.423	100	dirait	- 0.407	68	fume	- 0.386	180	dis	- 0.365	1284	femme
- 0.423	49	cependant	- 0.407	63	ronces	- 0.386	75	parcelle	- 0.365	439	immense
- 0.423	36	parfaite	- 0.407	55	extérieur	- 0.386	34	inconnues	- 0.365	120	magasins
- 0.423	33	féroce	- 0.406	298	hasard	- 0.386	31	souterrains	- 0.365	79	brouillard
- 0.422	428	rouges	- 0.406	78	bouteille	- 0.385	56	rayon	- 0.365	61	télévision
- 0.422	396	faisant	- 0.406	53	gonfle	- 0.385	53	répondit	- 0.365	61	coulent
- 0.422	273	mot	- 0.406	50	flottaient	- 0.385	40	gucule	- 0.365	46	asphalte
- 0.422	86	membres	- 0.406	49	4	- 0.384	339	ailleurs	- 0.365	38	prie
- 0.421	75	alla	- 0.406	36	fabrique	- 0.383	80	faits	- 0.364	285	voiture
- 0.421	44	trucs	- 0.405	62	essaya	- 0.383	70	deviennent			
- 0.420	503	immobile	- 0.405	41	avança	- 0.382	181	haine			
- 0.420	253	gros	- 0.403	202	points	- 0.382	88	brindilles			
- 0.420	202	trou	- 0.403	36	baigne	- 0.381	41	bruns			

Nous retrouvons ici l'emploi croissant des participes passés que nous avons pu constater auparavant. Cette tendance à la hausse pourrait refléter non seulement une tendance croissante à l'emploi du passé composé mais aussi traduire une sorte d'accomplissement de l'action de l'écriture avec des mots comme *arrivée, donné, décidé, raconté, marché, apporté, rêvé, montré, rencontré* etc. Nous l'avions déjà constaté

dans l'étude des temps verbaux⁵³⁵ mais la charge sémantique de ces participes renforce encore les résultats de l'étude syntaxique. Deux interprétations s'imposent : premièrement l'écrivain d'âge mûr, ayant atteint son but, se projette peut-être moins dans l'avenir et a plus souvent recours au passé ; deuxièmement, l'emploi croissant du passé composé traduit une tendance vers un style plus simple avec plus d'oralité au fur et à mesure que l'œuvre progresse et que la reconnaissance et le succès de Le Clézio se confirment. Les quelques formes en passé simple que l'on trouve dans liste de corrélation négative semblent confirmer l'abandon du style plus soutenu du début de l'œuvre⁵³⁶. Il convient de constater à ce propos qu'il y a une augmentation générale et progressive des verbes dans l'œuvre et que la liste de corrélations négatives en contient très peu.

Un autre phénomène que l'on pourrait attribuer à l'âge et à la maturité de l'écrivain est l'importance croissante de la *famille* avec la *sœur*, la *mère*, l'*oncle*, les *sœurs*, l'*enfant*, le *bébé* et le *mariage* en forte corrélation positive. Remarquons au passage que ce sont les membres féminins qui s'accroissent le plus en nombre.

Peut-être serait-il abusif d'attribuer la même explication à la présence croissante des mots comme *hôpital*, *médecins*, *couchée*, *endormie*, *allongée*, *tombe* etc. ? Néanmoins ces vocables ont un fort coefficient de corrélation positif... L'emploi croissant de *France*, *français* et *Paris* n'est sûrement pas le reflet d'un patriotisme accru chez notre écrivain, mais manifeste une tendance à faire installer la fiction en France, au moins partiellement, dans les derniers livres du corpus.

L'observation la plus intéressante est peut-être celle qui traduit une tendance vers l'abstraction et l'abandon des choses matérielles. Le Clézio semble de plus en plus s'éloigner du postulat inscrit sur la quatrième de couverture de *L'extase matérielle* : "La beauté de la vie, l'énergie de la vie ne sont pas de l'esprit, mais de la matière."

⁵³⁵ Cf. p. 394-398.

⁵³⁶ Cf. p. 402-403.

C'est en effet dans la liste de corrélation négative que nous trouvons la matière et les choses matérielles comme *papier, allumette, bille, choses, miettes, boule, couches, nappes, trottoir, antennes, crayon, bière*, etc., tandis que parmi les mots que l'écrivain utilise de plus en plus nous trouvons des mots abstraits qui souvent traduisent le sentiment et la subjectivité (*étonnement, impatience, amitié, méchanceté, crainte, enthousiasme*, etc.).

Les choses matérielles qui apparaissent dans le début de l'œuvre sont souvent de petite taille. Le Clézio aime aussi décrire les détails et le minuscule durant la première période de sa production, notamment dans *Le procès-verbal* où l'existence d'Adam Pollo semble étroitement guidée par un certain pédantisme et où le moindre détail contribue au délire du jeune homme. Des formes telles que *minuscule, minuscules, miettes, particules, centimètres*, etc. abondent dans ce livre. Dans *L'extase matérielle* il écrit ceci par rapport à l'importance qu'il donne aux détails⁵³⁷ :

“Des détails, toujours davantage de détails. Je n'en serai jamais rassasié. Le lavabo blanc où les deux robinets de cuivre, en fuyant goutte à goutte, ont laissé deux traînées indigo. Merveilleuses coulées inscrites dans la faïence. Je pleure avec votre bleu magique, je crie dans la couleur paradisiaque, je ne parle plus, je ne dis plus rien, je ne veux plus rien d'autre que vous ! Et les gouttes rondes tombent, tombent sans arrêt sur les taches bleues, ajoutant chaque jour une fraction minuscule de couleur infinie.

Je ne suis jamais autant ému que par les choses microscopiques. C'est en elles que je disparaissais le mieux. Ce sont elles qui me révèlent le plus exactement la vérité de la nature solide.”

Lorsque Pierre Lhoste lui rappelle cette dernière phrase en 1971⁵³⁸, l'écrivain lui répond ceci : “Oui parce que je suis myope. J'ai tendance à regarder les choses de très près, à voir dans chaque détail un infini.” Un peu plus loin il déclare à propos

⁵³⁷ *L'extase matérielle*, p. 113.

⁵³⁸ P. Lhoste (1971) : p. 21-22.

de la forme de sa mémoire⁵³⁹ : “Il me semble que je retiens davantage les détails, je veux dire, je laisse volontiers s’échapper des vues d’ensemble pour retenir des caractéristiques minuscules.” Parfois le minuscule devient microscopique⁵⁴⁰ :

“Mon corps est voué aux maladies, aux frémissements, aux attaques. Sans que je le sache, des millions de bactéries et de microbes sont en moi. Ils essaient de détruire. C’est cela : je sens mon organisme qui lutte pour vivre. Chaque cellule se défend avec acharnement, livre son combat microscopique avec une rage que je ne pourrai jamais imaginer.”

Un autre thème que l’on associe souvent au minuscule est en effet celui des insectes que Le Clézio exploite au début de sa carrière, pour, par la suite, l’abandonner de plus en plus. Parmi les mots ayant le coefficient de corrélation négative le plus important nous trouvons *pattes, mouche, araignées, bête, tentacules, lézards, carapaces* etc. qui témoignent de l’attention déclinante de l’écrivain pour les insectes, très présents dans le début de l’œuvre.

Toujours à ses débuts, Le Clézio attache aussi beaucoup d’importance aux différentes parties du corps, ce que reflètent des mots comme *narines, ongles, chair, muscles, poumons, bouche, cerveau, veines, organes* et *peaux*. En 1967, Le Clézio écrit dans *L’extase matérielle*⁵⁴¹ :

“C’est tellement fragile, un corps humain. Tellement compliqué, et chacune de ces complications tellement nécessaire à l’ensemble. Comment tout cela tient-il ensemble, comment tout cela arrive-t-il à persister ? Voilà ce qui devrait nous rendre à la fois fiers et inquiets ; car, cela est sûr, si notre chair avait la lâcheté de notre esprit, nous tomberions en pourriture dans la minute même. Mais notre chair est

⁵³⁹ P. Lhoste (1971) : p.116.

⁵⁴⁰ *L’extase matérielle*, p. 46.

⁵⁴¹ *Ibid.*, p. 47.

forte ; elle se bat. Elle a une volonté implacable, qui n'est pas celle de la raison, mais du courant même de la vie. Une volonté capable de lutter contre la mort pendant soixante-dix, quatre-vingts ans sans faiblir ! La beauté de la vie, l'énergie de la vie ne sont pas de l'esprit, mais de la matière. Je ne connais que cela : mon corps, mon corps. Ma vie électrique, ma vie chimique. Quant au reste, mes espèces de pensées, mes envies, ma conscience, cela vaut-il vraiment la peine qu'on en parle ?

Les mots correspondant aux différentes parties du corps ont toutefois un très fort coefficient de corrélation négatif, témoignant d'un abandon d'emploi successif. Le tableau ci-dessous illustre clairement la distribution de ces mots, à travers l'œuvre de Le Clézio :

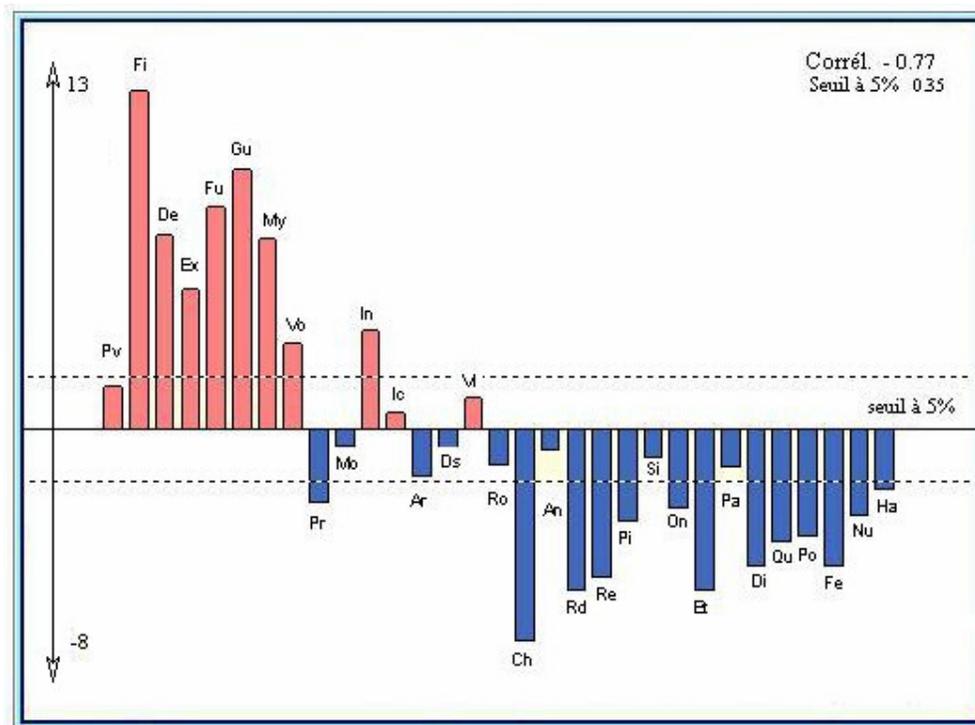


Figure n°158 : La distribution des parties du corps à travers le corpus A (écarts réduits).

En effet, le corps tout comme la matière semble prendre de moins en moins d'importance dans l'œuvre de Le Clézio et laisser place à l'abstraction et aux

sentiments avec des mots comme *magnifique, rêvé, étonnement, mémoire, amoureux* etc. que nous trouvons en haut de la liste des corrélations positives.

De la même façon, Le Clézio s'éloigne de plus en plus de la thématique de la ville et de la vie moderne qui était l'espace par excellence de ses premiers livres, des mots tels que *trottoir, essence, électrique, asphalte, carrosseries, voitures*, etc. témoignent de cet abandon ayant des coefficients de corrélation fortement négatifs.

Nous avons remarqué à ce propos que la liste de corrélation négative est deux fois plus longue que celle de corrélation positive, ce qui est assez exceptionnel. En principe les deux listes sont symétriques, ou plus exactement on observe une dissymétrie inverse de celle que nous constatons chez Le Clézio : il y a généralement plus de naissances verbales que de décès. Cela amène à constater que l'écriture de Le Clézio évolue différemment. S'agit-il de la rupture avec le vocabulaire et l'écriture du nouveau roman ou un abandon brutal de certaines réalités ? Nous avons pu constater dans nos analyses sur la structure du vocabulaire et les apports lexicaux qu'il n'y a pas un apport nouveau significatif de vocabulaire (sauf dans certains textes ethnologiques, mais ceux-ci sont distribués tout au long de l'œuvre et donc pèsent peu sur le coefficient de corrélation, alors que les textes du nouveau roman sont tous à un bout de la chronologie, ce qui influence grandement le coefficient) mais qu'il s'agirait plutôt d'une exploitation plus intense du vocabulaire courant.

5.3.6. Conclusion.

L'étude des spécificités – positives et négatives – permet de dégager les caractéristiques d'une écriture et d'une œuvre. Nous avons vu à l'étude exogène les thématiques caractéristiques de Le Clézio par rapport aux autres écrivains de son époque, ainsi que les particularités endogènes et l'évolution chronologique de son vocabulaire.

Il s'avère que certaines thématiques sont très importantes, la nature – terrestre et marine –, les couleurs ; d'autres comme le milieu urbain, les parties du corps, les insectes et le minuscule sont très présentes au début de l'œuvre mais perdent de l'importance au fur et à mesure que l'œuvre progresse.

Les champs thématiques majeurs dans l'œuvre et qui sont caractéristiques de l'écriture leclézienne ont souvent été l'objet d'excellentes analyses littéraires et critiques. Nous nous proposons de compléter ces études avec la vue objective et impartiale que l'ordinateur et l'analyse statistique nous fournissent.

5.4. Les structures thématiques.

“Les mots veulent dire tout de suite, sans tarder, toute la beauté terrestre. Ils ont beaucoup de hâte, ils sont impatients de dire tout cela, de raconter tout cela. Les vallées entre les montagnes, les arbres immobiles dans l’air du soir, les nuages, la brume au-dessus des rivières, les villages ocre, les villages rouges, les chemins, les restanques, les prés d’herbe verte. Les mots ne veulent pas dire les sentiments, les passions, ou les obsessions. [...] Rien d’autre ne résonne. L’horizon vaste est la seule limite du langage, qui n’exprime rien que la réalité.”

L’inconnu sur la terre, page 109.

Dans l’étude des spécificités, nous avons pu avec l’aide de l’ordinateur extraire les mots importants par leur fréquence ou bien par leur absence dans l’écriture leclézienne.

Nous avons ainsi pu mettre en exergue des thèmes caractéristiques de l’œuvre, sans pour autant avoir effectué un regroupement aussi précis que celui que nous avons pu faire lorsqu’il s’agissait par exemple des classes grammaticales.

“Malheureusement, écrit Étienne Brunet⁵⁴², le vaste domaine des mots sémantiques n’a pas été loti si précisément. Aucune théorie ne s’est imposée suffisamment pour définir des règles communément admises en matière de découpage.”

Les études sur la reconnaissance des contextes thématiques, automatiques ou semi-automatiques, notamment celle d’Éveline Martin⁵⁴³, ont montré les grandes difficultés auxquelles le chercheur est confronté à cause de la polysémie, de l’appartenance d’un mot à différentes classes et des choix arbitraires qui s’ensuivent dans le traitement. Le découpage en “univers-registres” tels que

⁵⁴² É. Brunet (1983) : p. 211-212.

⁵⁴³ É. Martin (1993).

“sociopolitique”, “ludique”, “professionnel”, “esthétique-culturel”, “affectif”, etc.,⁵⁴⁴ s’adapte mal à un corpus littéraire et relève, comme le souligne souvent François Rastier⁵⁴⁵, d’une grande part de subjectivité.

Nous avons également remarqué la difficulté que crée le tri automatique effectué par l’ordinateur dans le cas des mots polysémiques ou des mots correspondant à des notions appartenant à plusieurs catégories prédéfinies. Nous nous sommes trouvée devant de nombreux choix discutables et parfois même des aberrations flagrantes. Le tri semi-automatique ou manuel à son tour est une opération très coûteuse en temps et comporte inévitablement des erreurs aussi. Il faut également tenir compte de l’être humain qui ne va pas, comme le fait l’ordinateur dans sa logique parfois stupide mais toujours cohérente, trier chaque fois le même élément dans la même catégorie, phénomène de fluctuation qui s’accroît lorsque le corpus est grand et que le travail s’étend dans le temps.

Dans les travaux littéraires nous sommes confrontée à ces analyses très subjectives et souvent arbitraires. En effet, comme le disent Dominique Labbé et Pierre Hubert⁵⁴⁶ : “La notion de ‘champ lexical’ est classique en lexicologie : elle décrit la manière dont est organisé le lexique de la langue. Elle est plus rarement appliquée à l’analyse du vocabulaire d’un auteur ou alors il s’agit de reconstitutions plus ou moins ‘artisanales’ à l’aide des tables de concordances.”

Jean-Marie Viprey remarque également les choix arbitraires et souvent très conventionnels qui découlent de ces analyses ‘artisanales’⁵⁴⁷.

“De cette manière, on finira inmanquablement par intégrer *noir* au thème de la *mort*, *blanc* à celui de la *virginité*, *noir* et *blanc* ensemble aux thèmes de la virginité, du racisme, de l’éthique, du cinéma, etc. ; c’est-à-

⁵⁴⁴ Cf. É. Martin (1993) : p. 30-35.

⁵⁴⁵ Cf. p. 426-427.

⁵⁴⁶ D. Labbé et P. Hubert, “La structure du vocabulaire du Général De Gaulle.” in *JADT 1995, III Giornate internazionali di Analisi Statistica des Dati Testuali*, S. Bolasco, L. Lebart, A. Salem (éds.), Università degli Studi di Roma, La Sapienza³ – Università degli Studi di Salerno, CISU, Rome, 1995, p. 165.

⁵⁴⁷ J.-M. Viprey (1997) : p. 296.

dire en fait par projeter sur la trame parfois arachnéenne du texte l'appareil mathématique chaque jour appesanti des associations thématiques.”

C'est pourquoi nous avons opté pour une méthode qui dans l'esprit des isotopies⁵⁴⁸ crée des regroupements autour d'un mot-pôle s'appuyant sur la récurrence des mots qui l'entourent pour ainsi tenir compte des structures thématiques de l'auteur même et non pas d'une norme externe quelconque.

“Que peut tirer la statistique lexicale de l'hypothèse selon laquelle le vocabulaire d'un auteur est organisé à la manière d'un champ lexical ? écrit Dominique Labbé⁵⁴⁹. En premier lieu cela suggère que les mots entretiennent entre eux des relations (d'association, de substitution, d'opposition voire d'exclusion) plus ou moins fortes et que l'ensemble des places et des relations constitue la structure du vocabulaire de l'auteur.”

En effet, il s'agit de répondre à des questions simples : dans un corpus donné, quels sont les mots qui sont employés ensemble – on supposera qu'ils appartiennent à un même univers ou une même structure – et quels sont ceux qui s'excluent mutuellement ? Par exemple, quand le narrateur dit 'je', quels verbes préfère-t-il et lesquels évite-t-il ?

En ce qui concerne l'œuvre de Le Clézio, un classement sémantique “traditionnel” s'adapterait encore plus mal à des classements arbitraires et souvent très conventionnels. Car l'écriture leclézienne est caractérisée par une grande ouverture d'esprit et elle est également très influencée par d'autres cultures, notamment l'amérindienne, à laquelle une étude de structures thématiques à partir de l'usage que fait l'auteur même de mots et de thèmes semble bien plus adaptée.

⁵⁴⁸ Sans pour autant toujours utiliser sa terminologie spécifique.

⁵⁴⁹ D. Labbé (1990) : p. 166.

Dans un travail littéraire toutefois, l'étude de structures thématiques ne pourrait s'appuyer uniquement sur la simple récurrence de vocables, comme le suggère souvent la statistique lexicale. On conçoit bien de quelle pauvreté serait une étude thématique qui se limiterait à l'énumération sèche des résultats de cooccurrences ; une activité interprétative se développe nécessairement autour des différentes structures s'appuyant sur des méthodes d'analyse plus littéraires.

L'œuvre de Le Clézio est vaste et notre corpus de taille importante. Faute de place, nous ne pouvons que nous attacher ici qu'à quelques aspects, centrés autour de quelques thématiques, caractéristiques de l'œuvre, émanant de notre étude des spécificités de l'auteur.

Nous nous intéresserons d'abord à la nature et plus spécifiquement à la mer et à la terre omniprésentes dans l'œuvre, ainsi qu'aux insectes et d'autres animaux.

L'œil et le regard ont une importance capitale chez Le Clézio, non seulement par la fréquence et la récurrence des structures thématiques mais aussi par la signification si personnelle dont les dote l'écrivain.

5.4.1. La nature chez Le Clézio.

“La conscience est dans la mer et le ciel, dans la beauté terrestre. L'intelligence est dans l'harmonie de ces couleurs, de ces formes, dans l'immensité des perspectives. L'imaginaire est dans cette matière. Et le langage, celui que je parle, n'est pas ma réponse au monde, mais ce que toute la terre m'envoie, comme porté par un vent qui viendrait du centre du ciel, du cœur de l'espace.”

L'inconnu sur la terre, page 136.

Nous avons déjà pu constater l'importance de la nature et de ses éléments chez Le Clézio dans l'étude des spécificités. Le mot le plus significativement suremployé

du corpus était le mot *mer* suivi par *lumière, vent, plage, ciel, soleil, terre, montagne, eau, nuages, vallée*, etc. Les éléments naturels que sont la mer, le vent et la lumière ont largement retenu l'attention des commentateurs de Le Clézio. Leur omniprésence a pu être interprétée, comme l'exprime Sophie Bertocchi⁵⁵⁰, “dans le sens d’une communion entre l’homme et la nature, symbiose reposant sur un contact sensoriel, charnel.”

L’exploration statistique du champ thématique des éléments naturels montre que la distribution n’en est pas régulière à travers le corpus. Nous avons pris les 15 premiers substantifs de la liste des spécificités positives (*mer, lumière, vent, plage, ciel, soleil, terre, montagnes, eau, nuages, vallée, rochers, fleuve, sable et colline*⁵⁵¹) pour, avec le secours d’Hyperbase, constituer une liste qui reproduit aussi le total des fréquences. Le graphique ci-dessous illustre la distribution du groupe de ces 15 mots dans le corpus :

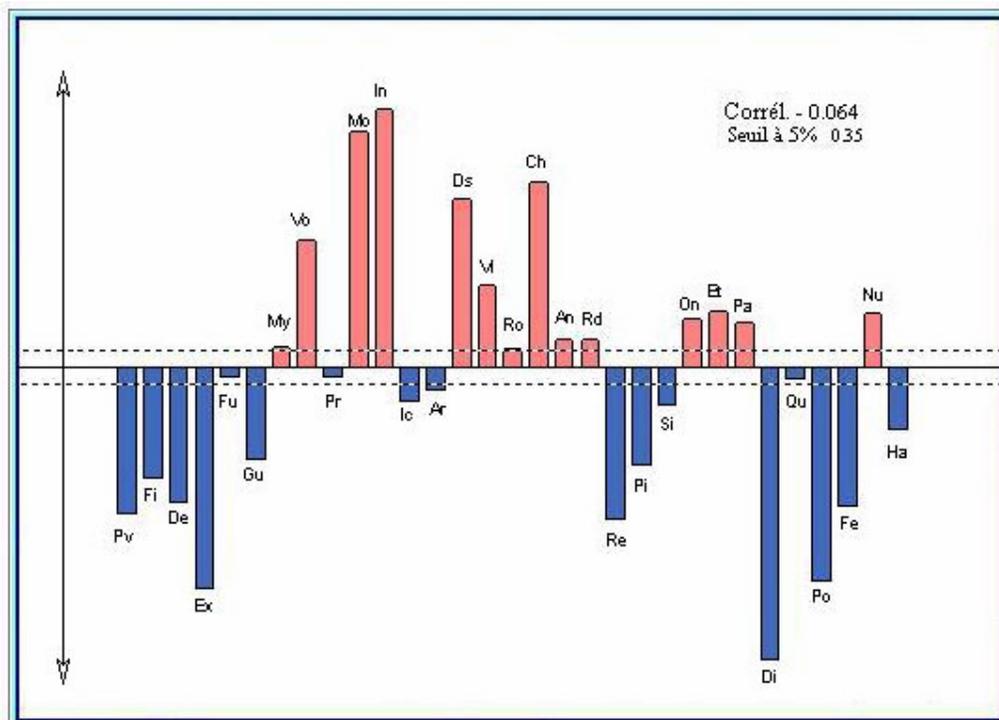


Figure n°159 : L'évolution du vocabulaire spécifique dans le corpus A.

⁵⁵⁰ S. Jollin-Bertocchi (2001) : p. 149.

⁵⁵¹ Il y a évidemment beaucoup d'autres termes qui expriment la nature dans cette liste.

La tendance générale n'est pas sans ruptures et reprises, comme l'indique le graphique qui de gauche à droite s'oriente selon la chronologie : déficitaire dans la période initiale du "nouveau roman", la nature est excédentaire dans les romans qui suivent *Mondo* sauf lorsque le genre, notamment celui des essais, s'y oppose. Mais cet excédent s'affaiblit dans les derniers romans.

La même analyse s'appuyant sur le corpus purement romanesque, le corpus E, met en exergue l'évolution à l'intérieur de ce genre :

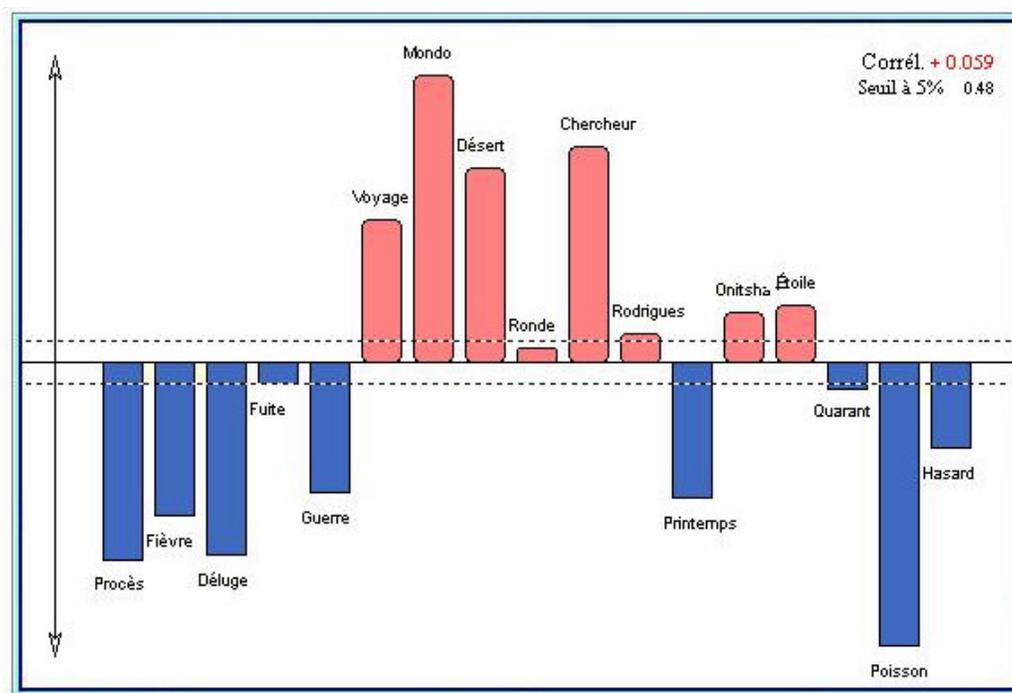


Figure n°160 : L'évolution du vocabulaire spécifique dans le corpus E.

Le graphique rend bien compte du début déficitaire – quand le milieu urbain était le décor privilégié de l'auteur –, d'une période significativement excédentaire au milieu de l'œuvre et d'une période du nouveau déficitaire vers la fin. Nous retrouvons en effet la rupture thématique de l'œuvre au niveau de *Voyages de l'autre côté* et l'unité de sa partie centrale que nous avons pu observer dans l'étude de la distance lexicale. Cet histogramme pourrait même presque résumer les différentes

études que nous avons faites dans les domaines structuraux et stylistiques. Il semble que l'introduction de la nature dans l'œuvre et sa forte présence dans une période centrale soit au cœur de l'écriture leclézienne. L'écriture de cette période constitue un hymne à la nature, où la présence même de la nature conditionne et modifie en fait tout le processus créatif, non seulement les thématiques mais également la structure ainsi que le rythme du récit. Laissons Le Clézio lui-même exprimer cette relation⁵⁵² :

“Tout est rythme. Comprendre la beauté, c'est parvenir à faire coïncider son rythme propre avec celui de la nature. Chaque chose, chaque être a une indication particulière. Il porte en lui son chant. Il faut être en accord avec lui jusqu'à se confondre. Et ce ne peut être une démarche de l'intelligence individuelle, mais de l'intelligence universelle. Atteindre les autres, se précipiter en eux, retourner en eux ; il s'agit de mimétisme.”

Divisons pour instant la nature en deux : l'élément terrestre et l'élément marin avec le mot *mer* par excellence.

5.4.2. La mer.

“J'aime la mer, c'est d'elle que vient la beauté réelle. Elle satisfait mon désir, car elle m'enseigne la force de la vie. D'où vient sa plénitude ? Elle est au fond de l'imaginaire : mer des rêves, mer immense comme le ciel, cercle de l'horizon qui vous étroit comme l'angoisse.”

L'inconnu sur la terre, page 159.

Lors d'un entretien au cours duquel Pierre Lhoste lui demande ce que lui évoque le mot *mer* Le Clézio répond⁵⁵³ : “La mer c'est justement la poésie. La mer ça doit

⁵⁵² *L'extase matérielle*, p. 130.

⁵⁵³ P. Lhoste (1971) : p. 41-42.

être ce bassin inépuisable vers lequel les hommes sont allés depuis des siècles, sur lequel ils se sont penchés.”

Le milieu maritime est en effet un décor de prédilection des récits lecléziens. Déjà *Le procès-verbal* se déroule dans une ville au bord de mer que l'on s'accorde à identifier comme étant Nice et le port d'attache du bateau *Azzar* sur lequel se déroule la fiction du dernier livre du corpus, *Hasard*, est bien Villefranche-sur-mer, à quelques kilomètres seulement de Nice.

La répartition chronologique des emplois du mot *mer*, qui est le premier de la liste de spécificités, résume à elle seule l'évolution de la structure thématique de la nature :

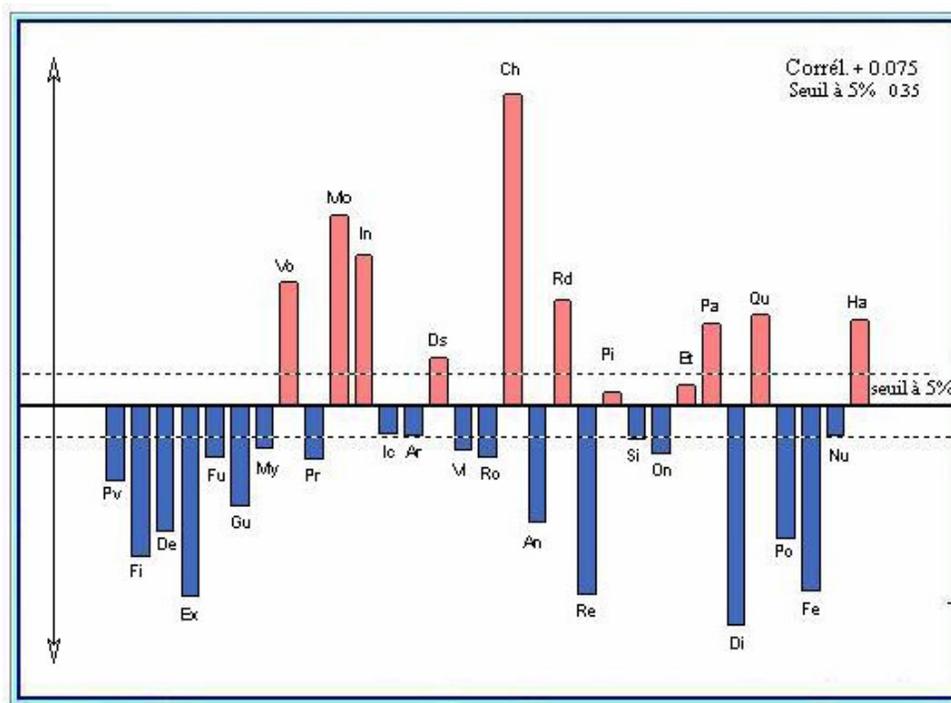


Figure n°161 : Le mot mer dans le corpus A.

Considérons *la mer* comme un thème leclézien ; nous ne le ferons pas à partir d'une liste projective de lexèmes (*océan, vague, bateau, etc....*) mais en prenant pour base une hiérarchie endogène de vocables, avec l'aide du logiciel Hyperbase.

Ce logiciel permet en effet d'effectuer une recherche de contextes automatique où chaque occurrence d'une forme est montrée dans son contexte (ici le paragraphe). Par exemple voici ci-dessous quelques-uns des 3367 contextes obtenus pour la mer⁵⁵⁴.

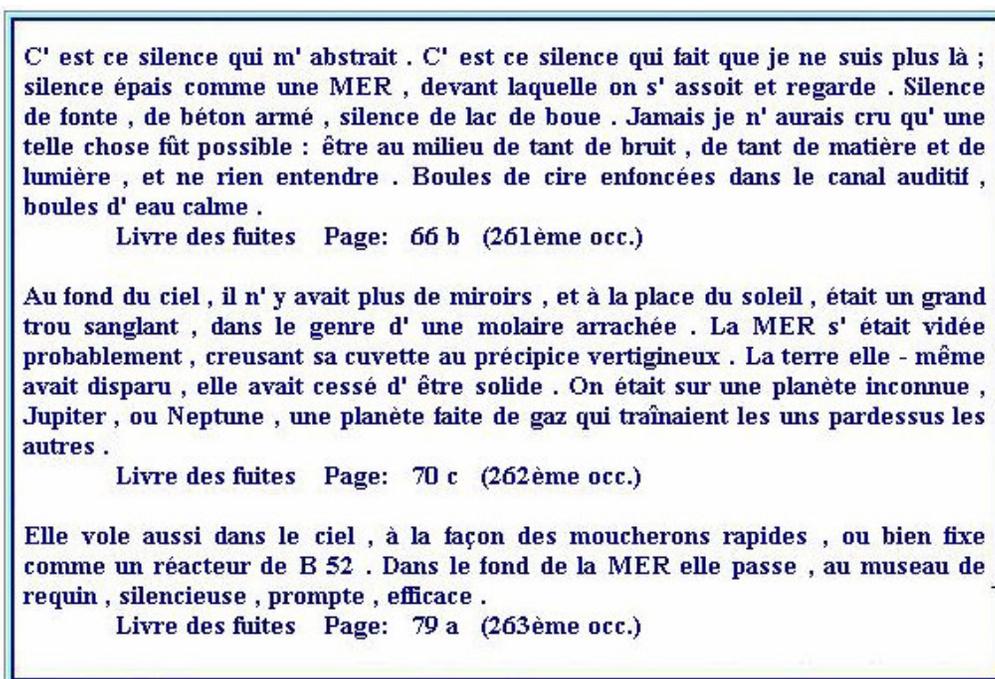


Figure n°162 : Quelques contextes du mot mer dans le corpus A.

On doit pourtant reconnaître que l'œil humain se fatiguera à examiner un à un les 3367 extraits relevés pour ce seul mot. Le logiciel Hyperbase s'y emploie avec une fonction d'extraction de thèmes. Il s'agit d'un calcul de spécificité particulier, puisqu'on ne recherche plus une relation entre un mot et un texte, mais une relation privilégiée entre les mots eux-mêmes – ce que mesure aussi le calcul de corrélation, quand deux séries sont juxtaposées dans le même graphique. La procédure ne se réduit cependant pas ici à deux mots confrontés, mais à

⁵⁵⁴ La liste exhaustive en annexe n° 7.

l'ensemble indéfini de tous les mots qui peuvent se trouver dans l'entourage d'un mot (ou d'un groupe de mots), qu'on définit comme étant le pôle⁵⁵⁵.

En confrontant le mot *mer* à son entourage, ici le paragraphe, on obtient un fichier discontinu de 300 000 mots dont on fera un sous-corpus constitué par les mots qui gravitent autour du pôle. Reste à comparer ce sous-ensemble au corpus Le Clézio qui est 7 fois plus important. Les calculs reposent une fois de plus sur la loi normale et l'écart réduit. Ils produisent en fin de compte la liste hiérarchique ci-dessous.

écart	corpus	texte	mot	écart	corpus	texte	mot	écart	corpus	texte	mot
166.81	3367	3403	MER	12.91	105	53	DANIEL	9.68	776	170	BORD
34.86	2344	787	VENT	12.75	1961	393	BRUIT	9.63	88	38	PROUE
33.17	59489	9093	LA	12.66	506	145	BATEAU	9.46	293	83	ENTENDS
33.01	701	351	VAGUES	12.50	93	48	ÉTRAVE	9.30	871	182	NUAGES
23.25	598	244	HORIZON	12.30	1904	378	LOIN	9.28	53	27	PÊCHE
23.06	178	116	ÉCUME	12.11	222	81	DUNES	9.25	116	44	PLACES
23.00	2936	714	CIEL	12.00	588	156	ÎLE	9.18	117	44	MARÉE
19.30	704	238	OISEAUX	11.94	70	39	ZETA	9.12	122	45	DENIS
19.16	782	254	PLAGE	11.91	2737	497	JUSQU'	9.01	622	139	SOMBRE
17.88	85	61	RÉCIFS	11.91	836	200	BLEU	8.99	220	66	HAUTE
17.54	462	169	ROCHERS	11.80	100	48	ALGUES	8.93	66	30	CORAIL
15.62	362	133	NAVIRE	11.67	734	180	MONTAGNES	8.88	4624	699	NOUS
14.80	2793	553	SOLEIL	10.82	73	37	TIMONIER	8.67	928	185	NAJA
14.67	1387	324	DESSUS	10.62	2851	492	EAU	8.67	151	50	CRÉPUSCULE
14.33	198	85	SEL	10.62	157	59	LAGON	8.55	153	50	LULLABY
14.14	44660	5860	LE	10.61	149	57	VOILES	8.52	246	69	ÉTENDUE
14.11	425	138	PONT	10.31	49	28	RIVAGES	8.47	284	76	PORT
13.66	225	89	BÂIE	10.30	3874	628	VERS	8.47	63	28	VACOAS
13.65	198	82	RIVAGE	10.30	271	83	PLATE	8.42	663	141	SOUFFLE
13.52	102	54	MOUETTES	10.28	243	77	BLEUE	8.37	319	82	ROCHER
13.45	14736	2136	SUR	10.25	473	122	COLLINES	8.35	87	34	ESTUAIRE
13.37	255	95	VAGUE	10.20	139	53	BATEAUX	8.19	169	52	TEMPÊTE
13.35	215	85	ÎLES	10.10	12621	1747	AU	7.96	64	27	SILLAGE
13.17	889	221	S&BLE	9.82	230	72	LISSE	7.93	810	160	COULEUR

Figure n°163 : L'environnement du mot mer.

A part les “corrélats” qui tiennent à la syntaxe (c’est le cas ici pour l’article féminin singulier *la* que la mer privilégie à l’exclusion des autres articles), à la phraséologie, aux expressions toutes faites, nous trouvons les véritables liens sémantiques, le partage de sèmes communs par quoi on peut définir un thème. Le thème de la

⁵⁵⁵ Cf. M. Kastberg Sjöblom, É. Brunet, “La thématique. Essai de repérage automatique dans l’œuvre d’un écrivain”, in *JADT 2000, 5èmes Journées internationales d’Analyse statistique des Données Textuelles*, M. Rajman et J.-C. Chappelier (éds.), Ecole polytechnique fédérale de Lausanne, Lausanne, 2000, p. 457-465.

mer chez Le Clézio peut ainsi être défini à partir de l'usage de l'écrivain lui-même et non plus par une norme externe. En haut de liste se trouvent des mots que l'on associe souvent avec la mer : *vent, vagues, horizon, écume, ciel, oiseaux, plage*, etc. Il s'agit encore d'un univers où l'être humain et la civilisation moderne n'ont pas de place, cette mer, souvent de l'hémisphère sud (il n'y a point de glace ou des icebergs dans cette liste) semble être chez Le Clézio bien plus un univers que l'on contemple qu'un espace de navigation et de transport.

Une fois le thème isolé – c'est-à-dire les mots qui gravitent autour du pôle *mer* – nous pouvons illustrer, par un graphique, l'évolution chronologique de la constellation lexicale qui entoure le pôle.

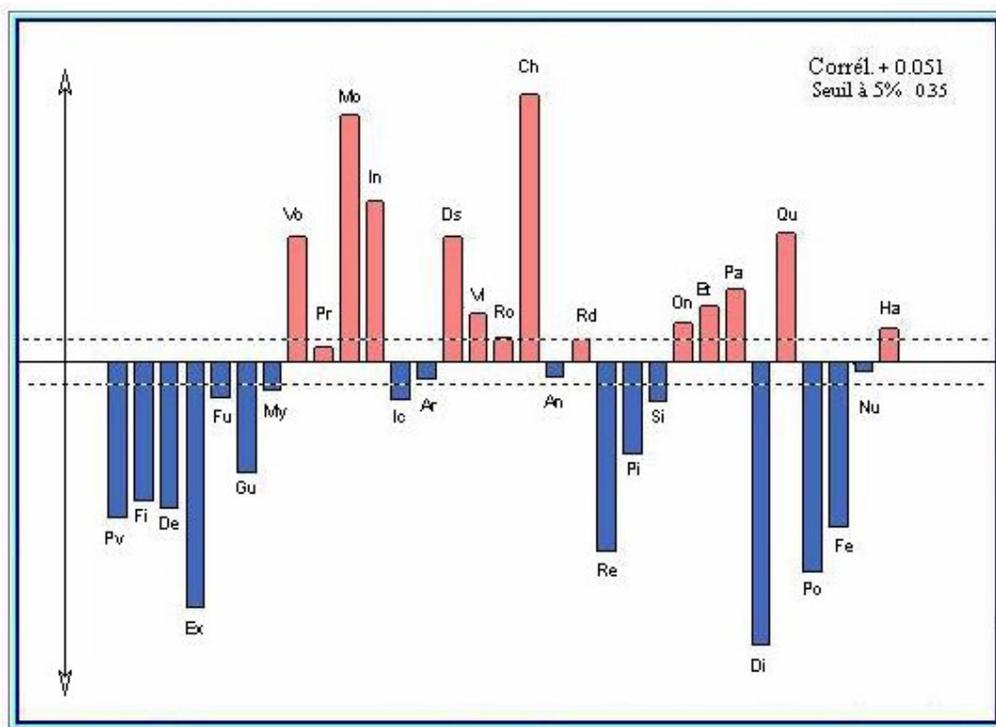


Figure n°164 : L'évolution d'ensemble du thème mer dans le corpus A.

La tendance du thème suit et complète celle du mot-pôle *mer* et elle est aussi en accord avec celle du vocabulaire spécifique. Les déficits de la période initiale font place à des excédents, du moins lorsque le genre romanesque est seul en cause.

Les différents éléments qui forment cette structure ne sont évidemment pas distribués de la même façon dans les différents ouvrages. C'est peut-être l'analyse factorielle qui rend le mieux compte des multiples liens entre les différents termes constitutifs du thème et les diverses œuvres :

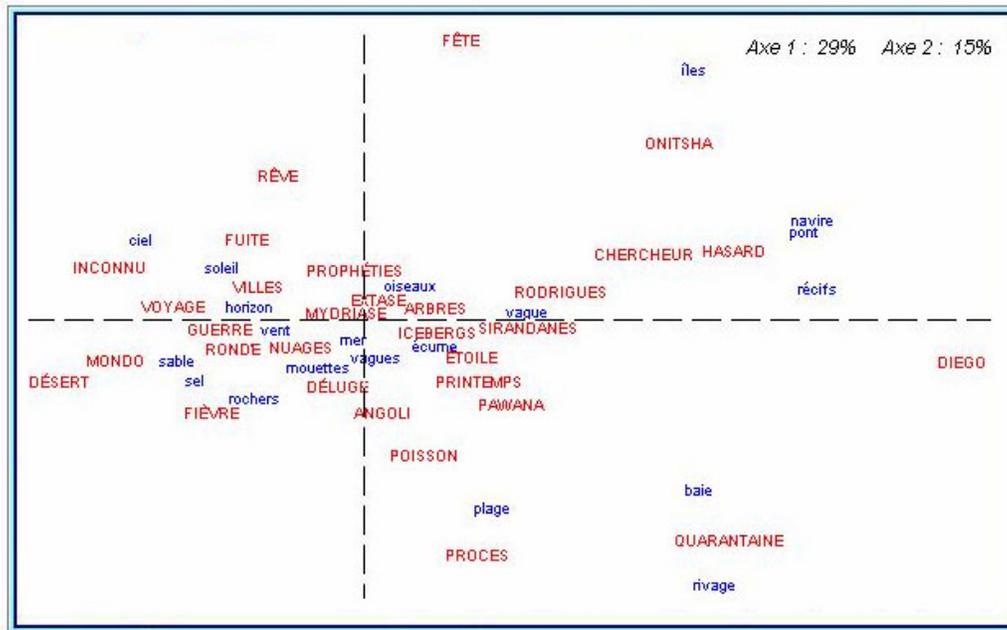


Figure n° 165 : Analyse factorielle de la structure thématique de mer dans le corpus A.

Dans cette analyse, nous ne retrouvons ni notre opposition générique habituelle ni la chronologie de l'œuvre. Il semble que d'autres facteurs prévalent dans ce cas bien que les axes ne soient pas très significatifs avec leurs 29% et 15%. Nous trouvons plutôt les décors et des éléments de l'action des différents livres. Sur le côté droit inférieur du tableau nous trouvons l'élément aquatique avec *Le procès-verbal*, qui se déroule en grande partie au bord de la Méditerranée. Dans *La quarantaine*, les personnages se trouvent dans la baie de l'île Plate en attendant de pouvoir poursuivre leur voyage à l'île Maurice ; dans *Le chercheur d'or*, *Hasard* et *Onitsha* il est question de traversées maritimes d'où l'importance du *navire*. Sur le côté gauche du graphique nous trouvons dans la partie supérieure les romans qui ont plus affaire à l'élément terrestre, dans la partie inférieure on retrouve la bordure entre *mer* et *terre* avec le vent, le sable, les rochers et les mouettes, qui se

trouvent également dans les ouvrages qui se déroulent dans le désert comme *Désert* et *Gens des nuages*.

Si nous examinons à la loupe chacune des différentes œuvres du corpus, nous constatons des constellations différentes. Ci-dessous se dessine le profil thématique spécifique de *Mondo et autres histoires* :

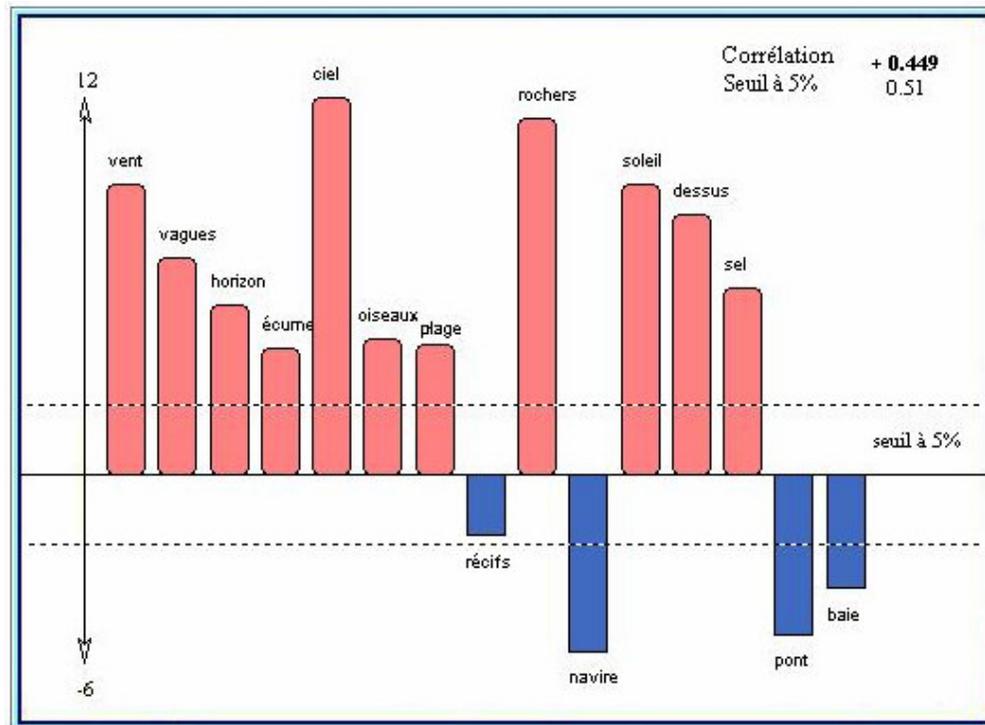


Figure n°166 : La structure thématique de la mer dans *Mondo et autres histoires*.

La figure reflète bien que dans ce recueil il est peu question de navigation et de bateau. Dans *Mondo* la mer est un décor, proche et omniprésent, mais elle n'a pas le même rôle que dans les romans qui se déroulent dans l'océan indien comme *Le chercheur d'or* ou *La quarantaine* où il est souvent question de vie maritime⁵⁵⁶ :

“La-bas, dit le timonier, l'eau est si transparente que vous glissez sur elle dans votre pirogue, sans la voir, comme si vous étiez en train de voler

⁵⁵⁶ *Le chercheur d'or*, p. 122-123.

au-dessus des fonds... J'ai cru que j'arrivais au paradis.”

Dans *Mondo* en revanche, comme souvent chez Le Clézio, il s'agit plutôt d'une action qui se déroule près de la mer et d'une contemplation. Toutefois dans l'œuvre leclézienne, ce ne sont pas seulement les personnages qui regardent et qui vivent près de la mer, la mer même est très vivante. Dans *Désert* l'auteur écrit ceci⁵⁵⁷ :

“Peut-être que c'est la mer qui regarde comme cela sans cesse, regard profond des vagues de l'eau, regard éblouissant des vagues de l'eau, regard éblouissant des vagues des dunes de sable et de sel ? [...] Le regard vient de tous les côtés à la fois.”

Cette relation de l'eau et du regard est bien documentée dans les études littéraires. Gaston Bachelard⁵⁵⁸ disait déjà que : “[...] si le regard des choses est un peu doux, un peu grave, un peu pensif, c'est un regard de l'eau.” L'eau joue en effet souvent, dans les interprétations littéraires, ce rôle inattendu. L'œil véritable de la terre, c'est l'eau. Dans la nature, c'est encore l'eau qui voit, c'est encore l'eau qui rêve. Paul Claudel écrit dans *L'oiseau noir dans le soleil levant* que “L'eau ainsi est le regard de la terre, son appareil à regarder le temps.”

Que la *mer* et le *ciel* délimités par l'*horizon* soient très liés n'a rien d'étonnant, mais cette dialectique met en exergue l'importance de la contemplation et du regard.

Nous pensions peut-être trouver aussi une dialectique commune entre la *mer* et la *terre* comme l'a fait Jean-Marie Viprey dans son étude sur le vocabulaire de Baudelaire. Il a effet constaté une valeur commune entre la terre, et plus exactement le désert, et la mer, et il cite l'avant-dernier quatrain de *La voix*⁵⁵⁹ :

⁵⁵⁷ *Désert*, p. 157.

⁵⁵⁸ G. Bachelard (1942) : p. 42.

⁵⁵⁹ J.-M. Viprey (1997) : p. 311.

“Et c’est depuis ce temps que, pareil aux prophètes,
J’aime si tendrement le désert et la mer ;
Que je ris dans les deuils et pleure dans les fêtes,
Et trouve un goût suave au vin le plus amer ;”

Chez Le Clézio nous n’avons pourtant pas retrouvé cette même dialectique. La terre et le désert - thèmes chers à notre écrivain - ne sont que rarement cooccurrents de celui de la mer, bien que l’écrivain aborde cette dialectique. Ils sont seulement juxtaposés au sein d’un espace qu’unifie la lumière du ciel et du soleil.