

# La thématique. Essai de repérage automatique dans l'œuvre d'un écrivain

Margareta Kastberg Sjöblom et Etienne Brunet

Bases, Corpus et Langage (CNRS, INaLF), UFR Lettres , Arts et Sciences Humaines, 98, bd. E. Herriot. B.P. 209, 06204 Nice Cedex 3, France

## Abstract

The purpose of this paper is to outline some of the possibilities offered by using the Hyperbase tool in textual analysis. The object is a statistical investigation of a corpus consisting of 2.000.000 tokens obtained from 30 novels by J.M.G. Le Clézio, a French contemporary author. Certain semantic fields and the thematic context of recurrent words are explored in the investigation. Several different strategies are adopted when exploring the corpus with the Hyperbase tool from a semantic perspective.

**Mots clés:** Lexicométrie, Corrélats, Thématique, Mesure de la distance lexicale, Le Clézio

## 1. Introduction

Jean-Marie Gustave Le Clézio passe 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?

Rappelons d' abord que le *Procès-verbal*, son premier roman, pour lequel il s' est vu attribuer le prix Renaudot, est publié en 1963. Il a été, aussitôt, rattaché à l' école du "Nouveau Roman" par la critique. Depuis, l' auteur s' est imposé, dans la littérature française, avec une œuvre importante et variée qui comprend non seulement des romans et des nouvelles mais aussi des essais de type ethnologique et biographiques et des livres pour enfant. Le succès des romans comme *Désert*, *Le Chercheur d'or*, *Onitsha* et *Étoile errante* tend à éclipser les titres plus anciens et il arrive souvent que la critique oppose deux périodes dans l' œuvre de Le Clézio. Aujourd'hu i ses livres figurent aux programmes des écoles et des universités et se maintiennent en tête de meilleures ventes: on aime sa manière qui mêle une histoire simple, insérée dans le temps, et une aventure mystique à valeur d' éternité.

### 1.1 Le corpus

Notre corpus représente la majeure partie de sa production. Il s' étend sur 35 ans, de 1963 à 1998, en englobant 30 textes complets et en recouvrant différents genres littéraires. Pour permettre l' indexation et l' exploration statistique du corpus, nous avons utilisé le logiciel Hyperbase dans sa version Windows (version 4.0).

Les trente ouvrages du corpus totalisent 49.773 formes différentes et 2.179.273 occurrences <sup>1</sup>. Certains ouvrages ont été écartés dont les éléments graphiques s' intégraient mal à une base purement textuelle.

Précisons que l' exploitation de cette base (et d'un e autre parallèle constituée avec les mêmes données, mais étiquetées) vient de commencer, la saisie, le contrôle et le traitement ayant occupé les premiers mois d'un e recherche qui doit conduire à une thèse. Nous ne nous attacherons ici qu' à quelques aspects, centrés autour de la thématique, en délaissant tout ce qui

---

<sup>1</sup> En annexe n° 1, la répartition des ouvrages représentés dans le corpus.

touche à la syntaxe, à la structure lexicale et même à la stylistique.

Dans cette optique, tous les mots ne sont pas à considérer, même si tous entrent dans le calcul de pondération. Le contenu lexical intéresse assez peu les mots outils, et mêmes les verbes à tout faire. Aussi bien les résultats qui suivent portent-ils généralement sur les substantifs.

## 2. Les spécificités de Le Clézio

Il s'agit ici d'une démarche classique, que le logiciel accomplit en s'appuyant sur Frantext, et plus précisément le corpus littéraire du XXe siècle. 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, auquel l'œil et le cœur de l'écrivain sont si sensibles. Les noms propres écartés, rien d'humain dans cette liste. 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. On est au premier jour de la Création: les éléments viennent d'apparaître. Rien ne bouge: pas de verbe, pas même d'adjectifs ou d'adverbes. Aucune articulation du discours: on dirait que la phrase est faite de substantifs juxtaposés. 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.

The screenshot shows the 'HYPERBASILEC' software interface. It features a menu bar with 'Corpus' (set to 'corpus trié') and 'Deficits'. Below the menu bar, there are navigation arrows and a search box labeled 'chercher'. The main area contains two side-by-side tables of data. The left table lists words with positive reduced deviations, and the right table lists words with negative reduced deviations. Each entry includes a 'N°' (number), an 'écart' (reduced deviation), a 'corpus' count, and the 'texte mot' (text word).

N°	écart	corpus	texte mot
139.78	6819	3033	mer
103.03	8196	2601	lumière
89.07	10986	2759	ciel
88.96	7270	2149	vent
88.77	1062	724	plage
84.79	10706	2614	soleil
84.19	15601	3288	terre
81.30	1124	689	montagnes
80.11	1164	693	vallée
75.99	12487	2649	eau
75.92	567	447	rochers
73.26	7241	1837	bruit
70.62	1740	774	nuages
68.98	2266	881	centre
66.00	326	292	ciment
64.78	2186	818	sable
64.71	759	452	collines
60.81	1647	659	fleuve
60.12	272	243	ravin
59.82	266	239	immeubles
59.56	1579	632	vagues
57.30	2208	742	pierres
56.07	2158	719	poussière
52.32	1767	604	dieux

N°	écart	corpus	texte mot
-2.01	6479	325	donné
-2.03	21269	1121	quelques
-2.04	7112	358	vivre
-2.08	6050	301	jeunes
-2.08	32943	1755	vie
-2.10	4629	226	roi
-2.11	20360	1069	porte
-2.18	14681	760	comment
-2.20	14896	771	presque
-2.23	34286	1822	après
-2.27	4565	220	attention
-2.37	7269	360	juste
-2.41	35823	1898	jamais
-2.42	7212	356	voulu
-2.53	13367	680	pendant
-2.66	11455	575	part
-2.68	8416	414	sommes
-2.70	5123	242	perdu
-2.70	5219	247	nez
-2.83	12649	634	disait
-2.85	15746	798	aurait
-2.85	136791	7405	tout
-2.90	14834	748	eu
-3.13	4907	224	parti

Figure 1. Le vocabulaire spécifique de Le Clézio

Le logiciel permet également l'observation du vocabulaire spécifique de chacune des 30 œuvres, c'est-à-dire une comparaison endogène. Cette spécificité est déterminée par le calcul de l'écart réduit pour chaque forme dans chaque partie du corpus. Les textes sont comparés, les uns après les autres, avec le corpus dans son ensemble. Les résultats sont très nets, ces mots reflètent parfaitement le thème de l'ouvrage et nous donnent le profil caractéristique de chaque livre<sup>2</sup>.

### 2.1. L'évolution

<sup>2</sup> En annexe n°2, la liste des mots spécifiques des différents sous-corpus.

L' exploration du champ thématique des éléments naturels montre que la distribution n' en est pas régulière à travers le corpus. Prenons les 14 premiers éléments de la liste n°1 groupons-les dans un tableau, dont la première ligne reproduit le total.

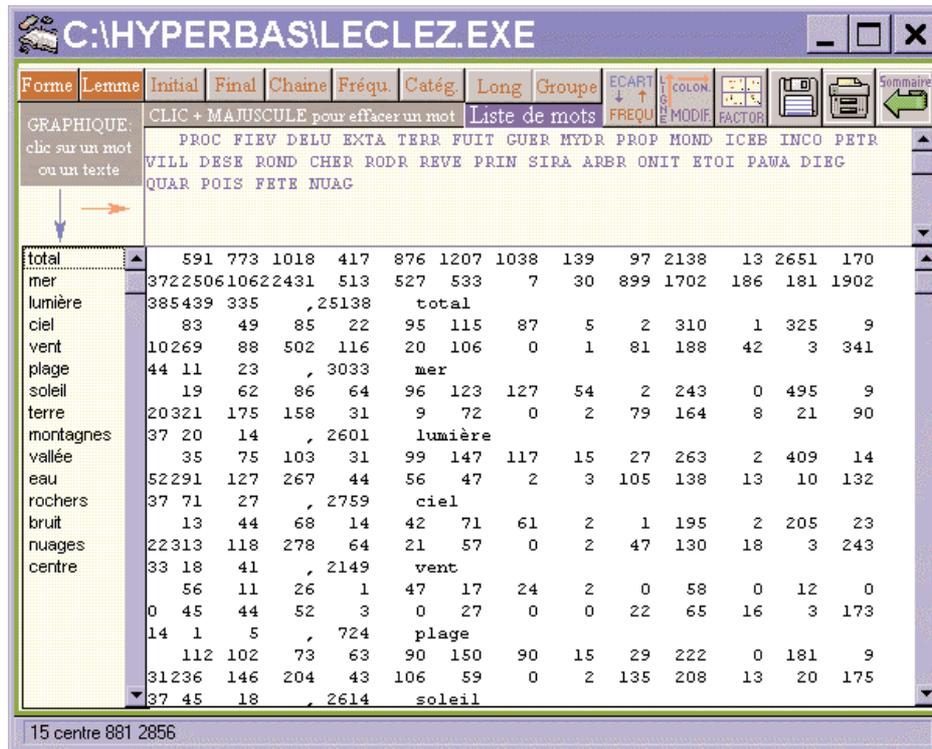


Figure 2. Le détail de quelques éléments spécifiques

La tendance générale n' est pas sans ruptures et reprises comme l' indique le graphique 3 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.

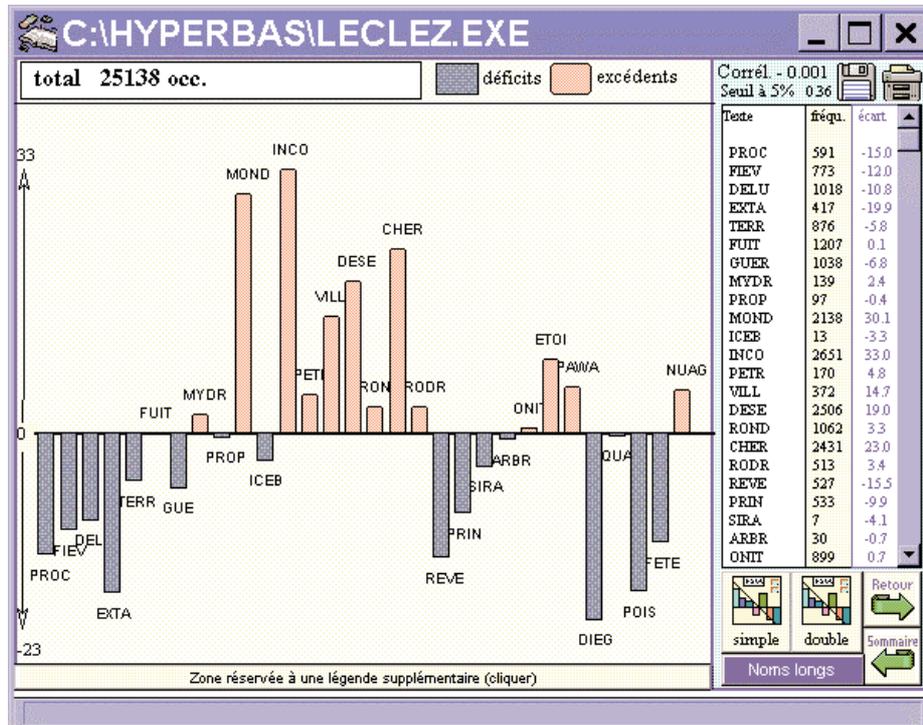


Figure 3. L'évolution du vocabulaire spécifique

En particulier la répartition chronologique des emplois du mot *mer* qui est le premier de la

liste, résume à elle seule l' évolution.

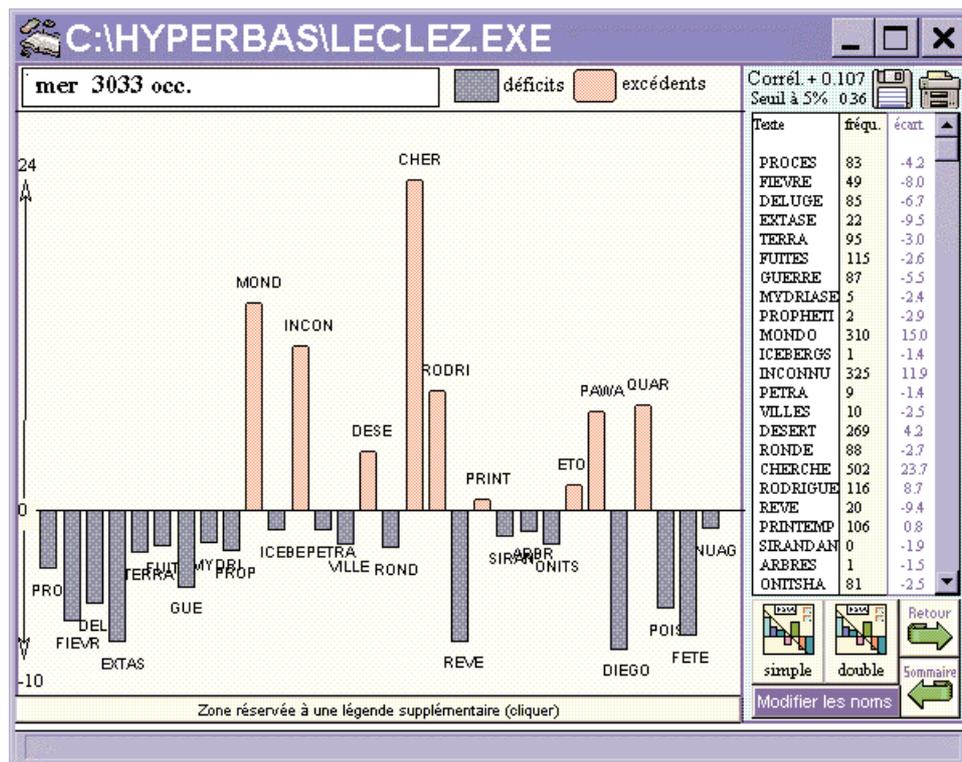


Figure 4. Le mot mer

On peut examiner à la loupe chacune des lignes (comme ci-dessus le mot mer), mais aussi chacune des colonnes - c' est-à-dire chacun des textes - dont le profil se dessine (comme ci-dessous *La Quarantaine*) à travers le choix qui est fait à l' intérieur de la liste.

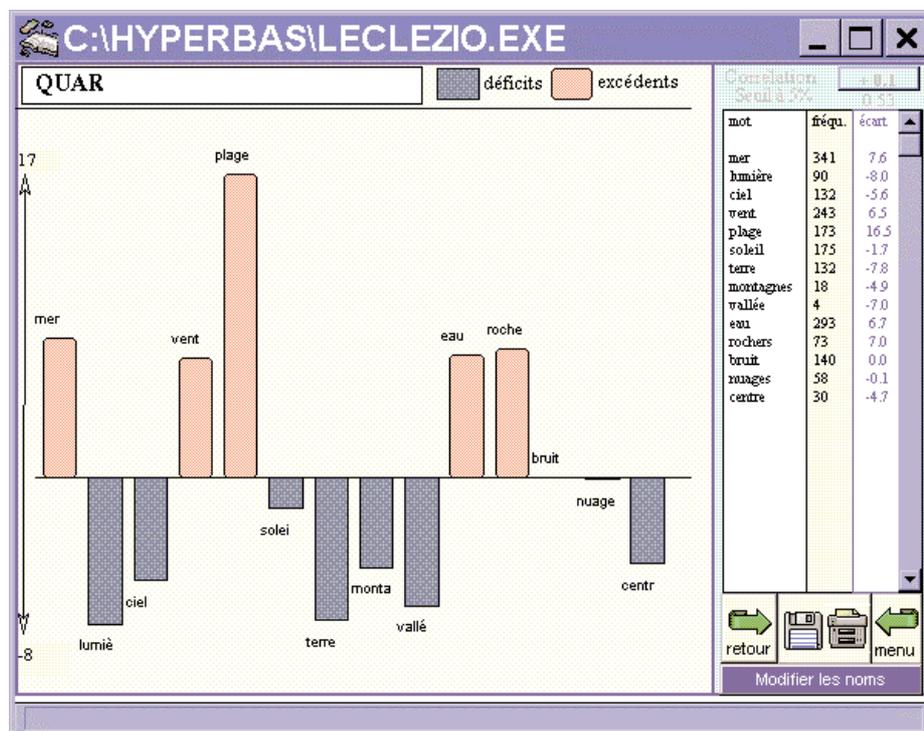
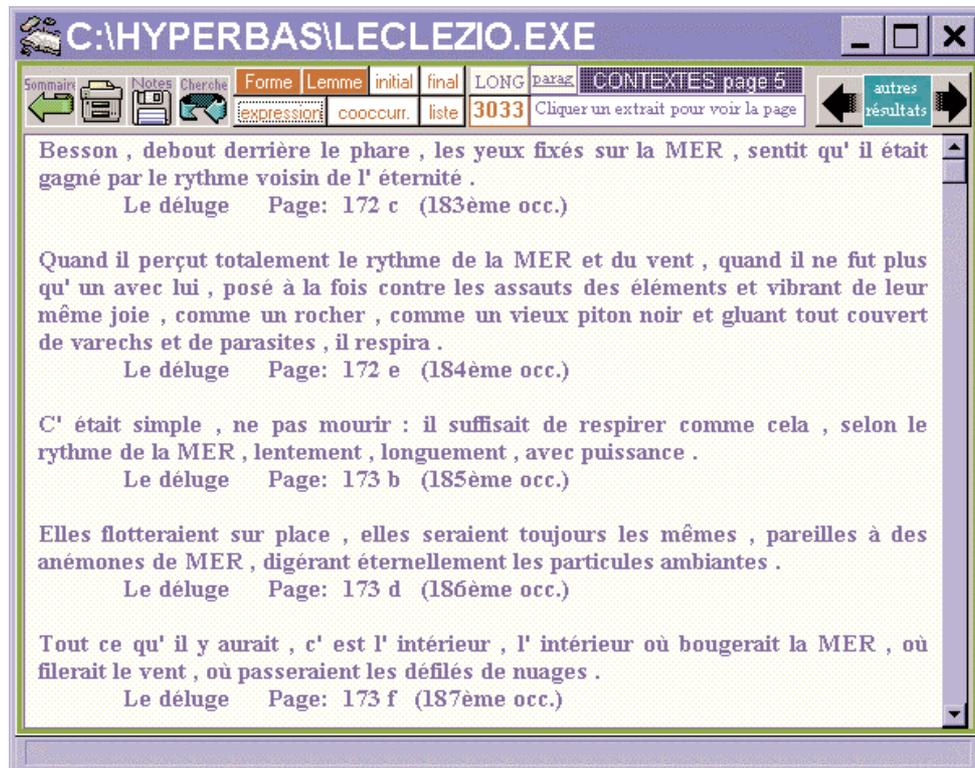


Figure 5. La nature dans la Quarantaine

## 2.2. Les contextes

Si l'on se méfie des nombres et des figures, rien n'empêche de contrôler dans le texte l'emploi des mots. Chaque occurrence d'un e forme est montrée dans son contexte (par défaut le paragraphe mais il est possible de d'étendre ou de diminuer la longueur des extraits). Pour



l'exemple on livre ci-dessous quelques-uns des contextes obtenus pour la mer.

Figure 6. Quelques contextes du mot mer

### 3. La fonction thématique

On doit pourtant reconnaître que l'œil humain se fatiguera à examiner un à un les 3033 extraits relevés pour ce seul mot. Y a-t-il un moyen mécanique de pallier les défaillances de l'attention? Le logiciel utilisé s'y emploie avec la fonction *THÈME* qui est disponible lorsque les extraits sont rassemblés. 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. Mais la procédure ne se réduit pas ici à deux mots confrontés, mais à 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 pôle. On peut changer de pôle aussi souvent qu'on veut, si bien que la procédure se trouve multipliable et réversible.

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 les deux listes, hiérarchique à gauche, alphabétique à droite, juxtaposées dans la figure 7.

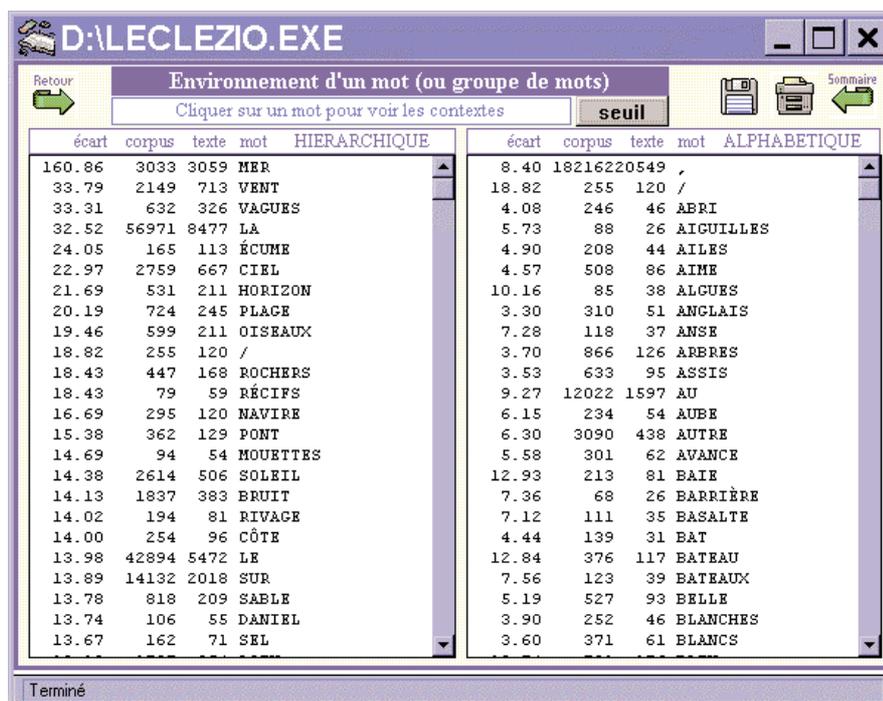


Figure 7. L'environnement du mot mer.

Un clic sur un mot de la liste des spécificités renvoie directement aux contextes, de sorte qu'on peut mesurer instantanément si la relation des "corrélats" tient 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, ou encore à un véritable lien sémantique, au partage de sèmes communs par quoi on peut définir un thème.

Une fois le thème isolé - dans ce cas les mots qui gravitent autour du pôle *mer* - nous pouvons illustrer, par une graphique, l'évolution chronologique de la constellation lexicale qui entoure le pôle. La tendance du thème suit et complète celle du pôle (figure 4). Elle est en accord aussi avec celle du vocabulaire spécifique (figure 3). Les déficits de la période initiale font place à des excédents, du moins lorsque le genre romanesque est seul en cause..

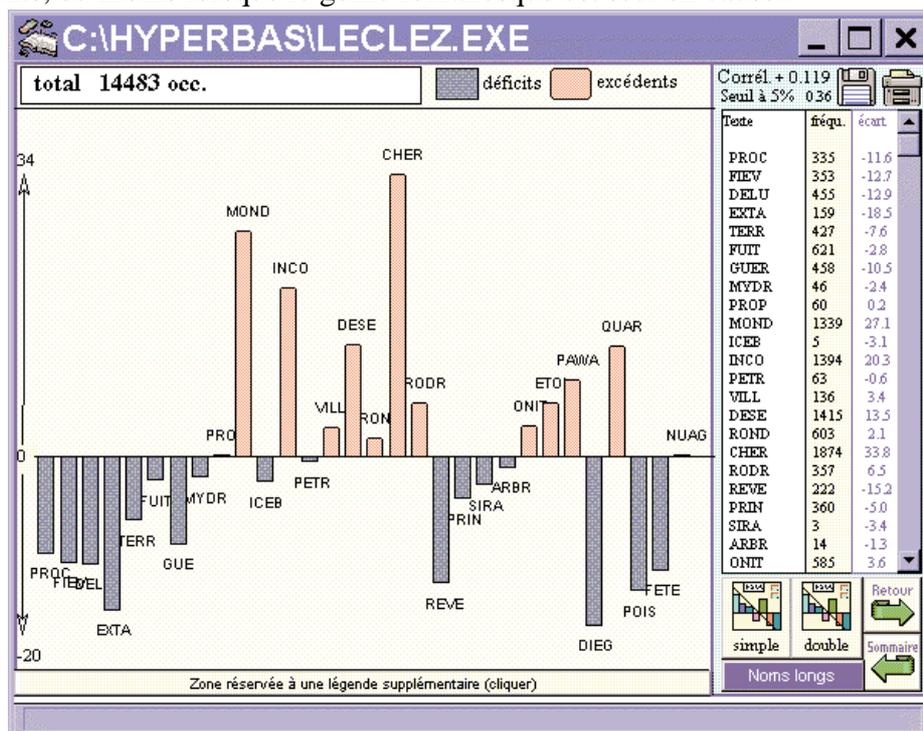


Figure 8. L'évolution d'ensemble du thème

#### 4. La distance lexicale. La connexion thématique des textes

On ne veut plus cette fois isoler un thème, grâce à une sélection automatique ou raisonnée de mots privilégiés. Il s'agit de considérer le vocabulaire intégral de chacun des textes du corpus et de repérer ceux qui partagent des thèmes semblables. Mais on ne se préoccupe plus de fréquence. Pour un mot donné seul 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 deux textes s' il est commun aux deux et à augmenter la distance s' il est privatif et ne se rencontre que dans un seul. La collection des données est assez lourde parce qu' il faut considérer tous les mots sans exception et que pour chacun on doit prendre en compte tous les appariements de textes deux à deux (le nombre des confrontations pour  $n$  textes étant égal à  $n * (n - 1) / 2$ , et ici à 435). Elle est réalisée dans la phase d' indexation et le résultat auquel on aboutit est délivré par le bouton DISTANCE de la page STRUCTURE.

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  $b - ab$  recouvrant les parties privatives).  $C'$  est cette distance que montre le tableau dans sa partie supérieure, les éléments du calcul (parties communes et privatives) étant détaillés dans la suite (du moins lorsque la place est suffisante). Comme ce tableau a près de 1000 éléments, il est assez imperméable à l' interprétation, sauf à isoler un texte en montrant son profil parmi les autres textes,  $c'$  est-à-dire la distance variable qu' il établit avec tous les autres. Pour une vision synthétique des multiples accords bilatéraux qui lient les membres du réseau, le plus commode est de recourir à une carte, à la représentation graphique et quasi géographique qu' en donne l' analyse factorielle:

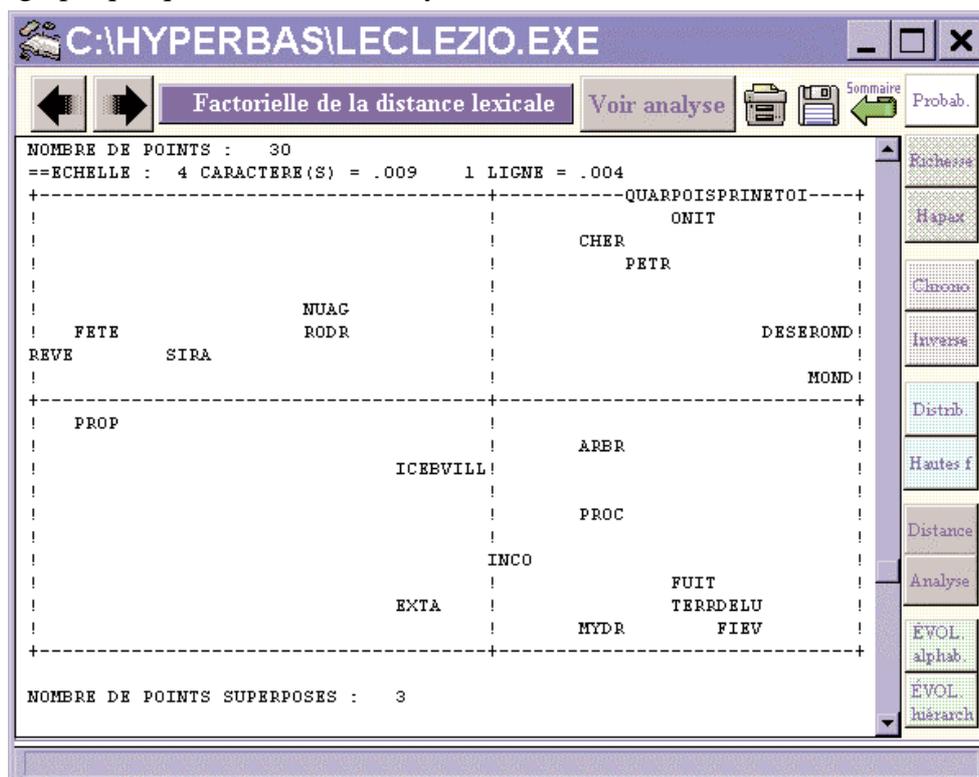


Figure 9. Analyse factorielle de la connexion lexicale

Pour qui connaît l' œuvre de Le Clézio, l' interprétation est aisée. Le premier facteur est surtout 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 rend compte de la diversité de l' écrivain. Le second facteur parcourt la chronologie du bas vers le haut du graphique; il rend compte de l' évolution de l' écrivain. Mais pourquoi parler ici de diversité et d' évolution thématique, puisque tous les mots sont considérés

qu' ils aient ou non une charge sémantique? En réalité le calcul étant indifférent aux faits de fréquence, les mots grammaticaux ne comptent pas plus que les autres. Leur influence est même réduite à néant, puisqu' ils se trouvent toujours dans la partie commune, à l' intersection de deux textes. Ils ne participent donc aucunement à la différenciation, laquelle repose en réalité sur les mots plus rares, sur ceux dont la compréhension - on dit maintenant l' intension - l' emporte sur l' extension. Ils sont plus chargés de traits sémantiques et ce sont eux qui donnent au texte sa coloration thématique.

## Références

- Brunet E. (1981). *Le vocabulaire français de 1789 à nos jours*. Champion-Slatkine. Paris-Genève.  
 Brunet E. (1988). *Le vocabulaire de Victor Hugo*. Champion-Slatkine. Paris-Genève.  
 Brunet E. (1994). Le CD-ROM Rabelais. *Travaux du cercle linguistique de Nice*. Vol n° 16 : 43-79.  
 Brunet E. (1999). *Hyperbase. version 4.0*. (Mac et Windows), INaLF. Bases. Corpus et Langage. Nice.  
*Mots chiffrés et déchiffrés. Mélanges offerts à Etienne Brunet*. (1998) Honoré Champion. Paris.  
 Lebart L. et Salem A. (1994). *Statistique textuelle*. Dunod. Paris.  
 Muller Ch. (1973). *Initiation aux méthodes de la statistique linguistique*. Hachette. Paris.  
 Muller Ch. (1977). *Principes et méthodes de statistique lexicale*. Hachette. Paris.

## Annexe n° 1

No	Titre	Occurrences	Vocables	Code
1	Le procès-verbal	93025	10020	Pv
2	La fièvre	101735	10128	Fi
3	Le déluge	122373	12074	De
4	L'extase matérielle	90633	8926	Ex
5	Terra amata	92116	9282	Te
6	Le livre des fuites	104456	9964	Fu
7	La guerre	110609	9582	Gu
8	Mydriase	9877	1878	My
9	Les prophéties de Chilim Balam	8784	1911	Pr
10	Mondo et autres histoires	99118	6629	Mo
11	Vers les icebergs	2716	780	Ic
12	L'inconnu sur la terre	124463	8457	In
13	Petra	10226	1866	Pe
14	Trois villes saintes	15358	2459	Vi
15	Désert	150902	8337	Ds
16	La ronde et autres histoires	83442	6425	Ro
17	Le chercheur d'or	134529	9232	Ch
18	Voyage à Rodrigues	38338	4896	Rd
19	Le rêve mexicain	87465	9275	Re
20	Printemps et autres histoires	70244	6149	Pi
21	Sirandanes	2516	765	Si
22	Voyages au pays des arbres	2978	662	Ar
23	Onitsha	76097	7199	On
24	Étoile errante	118824	7748	Et
25	Pawana	10616	1903	Pa
26	Diego et Frida	73341	8897	Di
27	La quarantaine	165642	11186	Qu
28	Poisson d'or	86164	7582	Po
29	La fête chantée	71039	8959	Fe
30	Gens des nuages	21647	3845	Nu
	TOTAL	2179273	49773	

# JADT 2000 : 5es Journées Internationales d'Analyse Statistique des Données Textuelles

différents sous-corpus				Ecart Corpus Texte Mot				Ecart Corpus Texte Mot				Ecart Corpus Texte Mot			
texte n°1 PROCES				48.11	418	64	peuple	33.05	74	73	timonier	88.11	63	49	lagune
Ecart	Corpus	Texte	Mot	32.19	232	32	parole	23.76	188	90	capitaine	80.82	55	42	baleines
27.78	55	44	rat	31.83	604	52	dieux	23.75	3033	502	mer	51.52	37	22	chaloupe
16.04	27	18	billard	25.46	259	27	livres	23.25	91	59	corsaire	29.91	57	16	navires
14.45	289	62	chien	23.38	53	11	calendrier	21.94	208	89	pirogue	19.76	25	7	carcasses
14.38	126	38	savez	23.10	29	8	prophéties	21.48	295	107	navire	15.54	79	10	géants
14.25	48	22	trucs	18.45	115	13	prêtres	21.40	517	149	rivière	15.09	31	6	chasseurs
14.16	123	37	type	18.16	245	19	univers	20.26	243	91	ravin	13.69	65	8	harpon
13.95	38	19	noyé	17.68	804	35	jours	20.16	118	60	anse	13.30	254	16	côte
12.35	117	32	demanda	16.50	383	22	mois	19.60	274	95	entends	12.76	213	14	baie
12.33	111	31	cas	texte n°10 MONDO				texte n°18 RODRIGUES				texte n°26 DIEGO			
11.24	100	27	paquet	80.78	313	312	mondo	44.31	1114	214	père	56.50	171	139	révolution
texte n°2 FIEVRE				26.45	798	192	petite	40.19	166	71	trésor	50.71	250	153	art
20.96	43	3	tic	25.09	36	33	fronde	35.00	243	76	ravin	45.95	104	88	peintre
18.07	43	27	allô	24.35	38	33	bouc	34.96	118	52	anse	43.53	76	71	révolutionnaire
14.87	41	22	mâchoire	22.46	30	27	directrice	32.60	1957	224	grand	42.07	153	99	peinture
14.33	343	72	espèce	22.15	308	95	croix	29.36	64	32	quête	36.72	49	48	fresques
14.03	57	25	comprenez	21.48	26	24	lanière	25.96	39	22	documents	28.95	37	33	communiste
12.58	610	94	sorte	19.82	182	64	plaine	25.61	40	22	commandeur	28.26	136	64	peint
11.55	324	59	trottoir	18.80	89	41	troupeau	24.63	135	40	plan	27.88	52	38	peindre
10.47	28	13	assure	18.57	45	28	gitan	22.65	91	30	corsaire	26.80	77	45	portrait
10.43	443	67	gauche	texte n°11 ICEBERGS				texte n°19 REVE				texte n°27 QUARANTAINE			
10.43	41	16	mademoiselle	26.81	70	8	poèmes	47.52	362	192	indiens	38.07	136	128	palissades
texte n°3 DELUGE				19.33	75	6	poésie	44.73	604	240	dieux	35.78	124	115	quarantaine
14.79	375	87	regarda	17.88	337	12	langue	38.71	100	80	barbares	30.25	156	112	lagon
14.62	27	19	rousse	17.81	1080	22	mots	36.95	114	82	aztèques	29.08	78	74	immigrants
12.39	54	24	parquet	11.00	57	3	poème	36.17	95	73	rites	27.70	160	105	volcan
11.80	96	32	trottoirs	10.96	379	8	langage	31.75	126	75	conquérants	25.80	59	57	récif
11.72	245	56	mit	10.34	29	2	tremblement	30.56	128	73	mayas	24.00	213	109	baie
11.68	68	26	infiniment	9.10	37	2	pois	29.63	74	53	culte	23.56	71	58	flot
10.70	33	16	étalait	8.86	39	2	poète	28.58	738	182	Dieu	23.40	511	179	île
10.57	324	62	trottoir	8.85	549	8	savoir	28.49	175	81	conquête	22.46	261	116	plate
10.39	97	29	rapidement	texte n°12 INCONNU				texte n°20 PRINTEMPS				texte n°28 POISSON			
9.80	229	47	fit	33.35	459	192	beauté	29.56	59	42	colonel	28.74	34	34	fondouk
texte n°4 EXTASE				29.27	2601	495	lumière	25.50	115	52	amie	26.56	35	32	princesses
28.08	312	112	matière	22.07	577	156	espace	20.84	27	20	opéra	15.10	55	24	disais
24.06	313	98	esprit	21.42	86	51	vertu	19.06	138	44	appartement	14.66	34	18	garage
23.93	235	83	conscience	20.63	2759	409	ciel	18.75	896	128	mère	14.58	139	39	bébé
22.87	124	56	néant	20.55	1755	300	vie	17.94	264	60	souviens	14.06	87	29	aimais
22.60	268	85	réalité	17.34	379	100	langage	16.94	43	21	lycée	13.81	59	23	comprendais
22.48	1178	203	mort	14.84	774	140	nuages	16.90	256	56	voulais	13.64	634	92	disait
21.04	1755	249	vie	14.72	420	94	visages	14.01	28	14	alcôve	13.57	183	43	cour
20.07	32	24	lucidité	14.20	135	46	fleur	13.19	90	25	madame	13.56	148	38	sentais
19.35	291	78	vivant	texte n°13 PETRA				texte n°21SIRANDANES				texte n°29 FETE			
18.78	198	61	infini	Ecart	Corpus	Texte	Mot	33.95	60	9	langues	41.31	604	200	dieux
texte n°5 TERRA				32.18	72	19	esprits	23.44	337	15	langue	35.67	135	78	seigneurs
28.66	1367	271	main	32.12	65	18	chèvre	20.00	34	4	langages	34.63	160	83	relation
25.69	97	55	index	30.04	100	21	tombeau	17.49	379	12	langage	30.09	49	39	amérindien
25.43	116	60	fermée	27.41	70	16	valise	15.93	53	4	créole	29.07	175	74	conquête
24.45	808	174	oui	25.13	104	18	guide	13.25	43	3	mépris	28.99	45	36	Amérindiens
24.29	32	29	annuaire	24.92	74	15	étrangère	12.38	49	3	amérindiennes	28.49	195	77	lac
23.86	79	46	tendus	20.16	157	18	falaise	11.25	159	5	français	25.30	128	55	empire
20.39	41	28	éléphant	17.72	32	7	muette	10.57	30	2	commun	23.94	202	6	seigneur
19.94	79	39	profil	16.15	166	15	trésor	10.03	478	8	pouvoir	23.41	738	137	Dieu
19.90	32	24	arrondie	texte n°14 VILLES				texte n°22 ARBRES				texte n°30 NUAGES			
16.13	36	21	lézard	36.67	308	56	croix	59.37	25	11	chêne	29.25	100	30	tombeau
texte n°6 FUITES				24.74	866	67	arbres	51.34	866	57	arbres	24.09	575	63	désert
17.98	95	42	fuite	23.71	135	24	fleur	26.73	554	24	garçon	22.09	301	41	rocher
17.09	63	32	flûte	22.37	260	32	forêt	26.26	260	16	forêt	21.85	130	26	cheikh
14.59	64	28	romans	17.31	202	22	seigneur	25.78	39	6	clairière	19.96	693	59	vallée
10.58	1974	195	homme	15.60	3288	98	terre	25.76	53	7	sifflant	17.90	83	17	nomades
10.42	154	35	autobus	14.78	49	9	vengeance	25.15	1016	31	petit	17.14	39	11	tribus
10.22	431	66	voitures	13.75	299	22	fleurs	20.18	286	13	branches	14.66	52	11	arabes
10.02	162	35	moteur	13.63	232	19	parole	16.47	42	4	siffllements	13.80	58	11	troupeaux
9.98	875	105	murs	13.42	238	19	puits	15.93	174	8	racines	13.33	26	7	saints
9.25	38	14	cubes	texte n°15 DESERT				texte n°23 ONITSHA							
9.11	142	30	acier	32.82	130	104	cheikh	46.26	659	241	fleuve				
texte n°7 GUERRE				31.57	575	232	désert	26.20	71	43	reine				
34.09	250	131	monsieur	24.60	118	76	manteau	24.83	208	73	pirogue				
27.93	838	220	filles	23.06	216	101	dunes	21.74	47	29	épave				
22.34	1177	228	jeune	21.50	126	70	cité	16.55	31	18	club				
19.84	786	162	guerre	19.60	101	57	guerrier	16.37	35	19	embarcadère				
17.80	431	103	voitures	18.78	818	193	sable	15.27	86	29	varangue				
16.65	106	43	goudron	17.92	26	25	figuier	14.97	112	33	coque				
15.55	1080	167	mots	17.80	2386	386	hommes	14.70	362	64	pont				
13.91	285	66	voiture	17.58	40	31	chameaux	13.21	25	13	membrures				
13.57	401	80	tellement	texte n°16 RONDE				texte n°24 ETOILE							
13.57	37	20	esplanade	39.22	203	115	pouce	31.58	150	96	maman				
texte n°8 MYDRIASE				22.27	79	41	mobile	26.33	68	53	italiens				
21.57	3003	93	yeux	20.08	63	33	aurora	23.90	104	61	berger				
18.28	51	9	pupilles	17.76	82	34	villa	23.77	191	85	camp				
13.57	687	27	voit	16.28	319	68	sent	23.66	53	42	juifs				
13.52	41	6	astre	15.70	66	27	dame	21.36	107	56	torrent				
12.72	100	9	voient	15.18	33	18	vélocoteur	20.89	1351	248	enfants				
12.63	577	23	espace	15.07	189	47	frère	20.82	25	25	carabiniers				
12.32	2601	54	lumière	15.01	53	23	loup	19.60	73	42	piano				
11.99	89	8	lampes	14.32	104	32	guide	19.25	51	34	valises				
10.28	48	5	froids					19.24	38	29	fugitifs				
10.05	50	5	orbites					18.51	239	78	soldats				