



**HAL**  
open science

## Les noms propres chez Zola

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Les noms propres chez Zola. A. Zampolli. Studies in honour of Roberto Busa, S.J., IV-V, Giardini editori e Stampatori, pp.29-57, 1985, *Linguistica Computazionale*, ISSN 0392-6907. hal-01790735

**HAL Id: hal-01790735**

**<https://hal.science/hal-01790735>**

Submitted on 28 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Les noms propres chez Zola

Etienne Brunet, BCL(CNRS), Université de Nice Côte d'Azur

Les noms propres ont la curieuse propriété d'apparaître fort souvent dans les Index et fort rarement dans les statistiques. Bien avant l'avènement de l'ordinateur on s'est soucié de relever leur apparition et leur localisation dans les textes et de proposer au lecteur un *Index nominum*<sup>1</sup>. Tout candidat à la thèse avait cette obligation, aussi contraignante que la bibliographie et la table des matières. Les méthodes liées à l'informatique ont permis de généraliser cette procédure et de l'étendre, par des procédés entièrement automatiques, aux exposés de toute nature, simples articles ou longs rapports. Les logiciels de traitement de textes offrent à l'envi de telles facilités – dont l'intérêt est évident quand on se préoccupe de retrouver non seulement un personnage mais aussi les épisodes où il entre en jeu et les thèmes dont il est porteur.

Inversement les études lexicologiques, en dehors de l'onomastique, attachent généralement peu d'importance à la fréquence des noms propres – qu'on écarte le plus souvent des relevés pour ne comptabiliser que les noms communs. L'étendue d'un texte (le nombre des occurrences) et celle de son vocabulaire (le nombre des vocables) sont établies comme si les noms propres n'existaient pas ou comme si les avait passés au caviar quelque censure, hostile à la personnalisation du discours, ou à la publicité des marques commerciales. Cet ostracisme s'appuie sur des considérations théoriques et pratiques qui ne sont pas sans valeur et qui nous ont conduit à exclure cette catégorie de notre étude générale du vocabulaire de Zola. Les noms propres ont en effet un statut particulier qui les assimile difficilement à la langue et les relègue au ghetto des immigrés. Il y a certes des échanges avec la population des mots communs, soit qu'une particularité physique, une appartenance

---

1. Il en est ainsi de l'édition de la Pléiade, où la tradition impose un Index des personnages et parfois aussi un Index des lieux. En ce qui concerne Zola cet Index se trouve à la fin du tome V, p. 1791-1886.

géographique ou une occupation sociale aient créé un patronyme (exemples : DUPONT, MULLER OU BRUNET), soit qu'inversement un personnage, réel ou fictif, accède au rang de type, parfois même en perdant la majuscule (un HARPAGON, un MECENE). Mais ces faits d'assimilation sont extrêmement lents et en pure synchronie la distinction des deux populations s'impose sans difficulté, les noms propres ayant, épinglée à leur poitrine, la marque distinctive de la majuscule. Cette distinction graphique dans le discours est transposée aussi nettement dans le dictionnaire et si les encyclopédies accueillent les noms propres (du moins les noms réels), les dictionnaires de langue leur refusent l'entrée. Les noms propres seraient donc étrangers à la langue. Englués dans la contingence, impropres à la représentation symbolique, rebelles à la vie communautaire des mots, incapables de faire souche durablement et de produire autre chose que des dérivés étiques, les noms propres semblent jetés par le hasard sur la surface de la langue et emportés par le même hasard avant d'avoir pris racine. Ils sont dans la langue ce que sont les constantes dans le langage mathématique, c'est-à-dire des êtres étroitement limités et dépourvus de la richesse symbolique attachée aux variables, aux opérateurs et aux fonctions. Il n'y a de langage que là où il y a des signes, et quand, tout autour, les mots communs s'agitent et se font mille signes, le nom propre reste planté là, étourdi par ce jeu, comme le pompier de service égaré sur la scène parmi les acteurs.

À ces considérations théoriques s'ajoutent certains embarras pratiques qui rendent malaisé le traitement des noms propres. On rencontre d'abord un problème de définition : le critère graphique (le mot étant l'espace compris entre deux blancs ou plus généralement deux séparateurs), qui est déjà fort discutable quand on l'applique au vocabulaire commun et par exemple au CHEMIN DE FER, est encore moins satisfaisant lorsqu'il est étendu au nom propre. Quand un personnage est désigné par un nom et un prénom (par exemple EUGENE ROUGON) ou par des titres nobiliaires (par exemple VERONIQUE BEULIN-D'ORCHERE, femme de ce même Rougon), le traitement automatique distingue EUGENE de ROUGON comme il distinguerait EUGENE de son frère ARISTIDE, et sa femme est coupée en trois morceaux. On peut ainsi craindre une multiplication des noms propres dans les romans aristocratiques, par simple bouturage<sup>2</sup>.

---

2. On a tout de même veillé à couper les particules et les M., MLE, MME, MMES dont Zola fait un usage modéré (on trouve autant de M. chez Proust (2649) que chez Zola (2663) pour une étendue deux fois moindre ; les chiffres font donc apparaître Proust plus aristocratique que Zola, ce qui ne surprendra personne).

Cette insensibilité de la machine aux relations d'équivalence est aussi grave, quoique moins visible, lorsque disparaît la contiguïté. À quelque distance la même personne peut être désignée par son nom de femme mariée ou son nom de jeune fille, par un surnom, un prénom, une périphrase, ou tout simplement un pronom de rappel. Un décompte exact pourrait mettre au compte de Gervaise ou de Félicité toutes les occurrences de ELLE, de LA ou de LUI qui leur appartiennent, mais comment retrouver tant de bottes d'aiguilles dans tant de bottes de foin (il y a 46 343 ELLE et 83 045 LA) ? Mais même sans avoir des exigences aussi terribles, on pourrait souhaiter moins d'ambiguïté dans ce domaine, Or si le traitement est impuissant à réunir ce qui est disjoint (un EUGENE et un ROUGON), il est tout aussi incapable de disjoindre ce qui est confondu. Quand on obtient un effectif pour ROUGON, s'agit-il de PIERRE, d'EUGENE, de SIDONIE, de PASCAL ou de quelque autre membre de la tribu ? Les noms de famille sont ainsi par essence des noms génériques et la même ambiguïté s'attache à d'autres patronymes, comme MOURET et MACQUART. Mais les prénoms ne sont pas plus purs. Zola les reprend d'un texte à l'autre pour désigner des personnages parfaitement étrangers les uns aux autres<sup>3</sup>. On compte ainsi 3 CLEMENCE (une vendeuse, une bonne, une ouvrière), 3 LOUISE (l'épouse de Lazare Chanteau, une actrice dans *Nana*, une domestique dans *Pot-Bouille*). On distingue aussi 3 HENRIETTE différentes (*Son Excellence*, *l'Assommoir*, la *Débâcle*), 3 ADELE, 3 AUGUSTE (et 2 AUGUSTINE différentes dans *l'Assommoir*, à 100 pages de distance), 3 ALEXANDRE (sans parler d'un empereur et d'un coq qui répondent pareillement à ce nom glorieux), et jusqu'à 7 ROSE et 8 CHARLES (sans compter les rois). Est-il légitime de mettre tous les CHARLES dans le même effectif et toutes les ROSE dans le même bouquet ? Et que dire de la promiscuité qui appareille la Vierge MARIE de la *Faute* et la MARIE Pichon de *Pot-Bouille* ou qui apparente au Rédempteur le picaresque JESUS CHRIST de la *Terre*<sup>4</sup> ?

---

3. Cela est vrai parfois des noms de famille eux-mêmes, par l'effet de quelque négligence de Zola. Ainsi quel rapport y a-t-il entre la ROSALIE PICHON d'*Une Page d'Amour* et le ménage PICHON de *Pot-Bouille* ?

4. À ce mélange du spirituel et du temporel, s'ajoute parfois un recouvrement de l'homme et de l'animal : il arrive que LISE désigne une vache, ALEXANDRE un coq, MATHIEU un cochon (dans la *Faute*) ou un chien (dans la *Joie de vivre*), CESAR un taureau, BERTRAND un chien et NANA une pouliche.

Il serait vain de nier ces difficultés théoriques et pratiques et, faute d'avoir désambiguïsé les quelque 72 000 occurrences de noms propres chez Zola, on ne saurait prétendre à l'exactitude du détail. Mais cette situation est assez répandue là où s'exerce la statistique, c'est-à-dire là où la mesure directe des phénomènes est peu praticable. Et l'on peut tirer des conclusions relativement précises et sûres à partir de données floues et incertaines, quand du moins la loi des grands nombres permet d'espérer la neutralisation des erreurs et des « bruits ». D'ailleurs des brouillages de toutes sortes perturbent aussi le vocabulaire commun, où l'on coupe également les syntagmes insécables et où l'on confond pareillement les homographes. Et de ce point de vue il n'y a pas de différence de nature entre le traitement des noms propres et celui des noms communs.

D'ailleurs Zola prend quelques précautions pour ne pas égarer son lecteur dans un labyrinthe onomastique. Tout d'abord les noms de lieu sont univoques, qu'ils soient réels (PARIS), inventés ou déguisés (PLASSANS). Mais le choix des noms de personnages semble obéir aussi à quelques directives prudentes. Zola évite de galvauder le nom des héros principaux afin que la figure de NANA, de GERVAISE, ou d'ÉTIENNE ne se confonde avec nulle autre. Il en est ainsi de beaucoup de noms ou prénoms réservés : THERESE, CAMILLE, LAURENT, GUILLAUME, SILVERE, MIETTE, PIERRE, FELICITE, RENEE, MAXIME, SACCARD, FLORENT, MARTHE, FAUJAS, ALBINE, SERGE, CLORINDE, HELENE, OCTAVE, DENISE, LAZARE, CLAUDE, CLOTILDE, PASCAL. Même les patronymes sont accaparés par une ou deux figures dominantes : ROUGON par Pierre et Eugène, MOURET par François et Serge, MACQUART par Antoine. Quant aux appellations collectives, à celles qu'on retrouve d'un roman à l'autre pour désigner des comparses différents, Zola leur confère assez souvent une unicité qui est celle de leur emploi et en cela il reprend la tradition de Molière qui donne à ses noms une spécificité sociale, les servantes s'appelant NICOLE ou MARTINE et les jeunes premiers DORANTE, CLEANTE ou LEANDRE. Ainsi les multiples variétés de ROSE que Zola cultive dans les *Rougon-Macquart* partagent une propriété commune : la condition ancillaire. Une ROSE est fille de comptoir dans le *Ventre de Paris*, une autre est servante dans la *Conquête*, une autre paysanne au hameau des Artaud. On trouve aussi une ROSE femme de chambre dans *Germinal*, une ROSE fille de concierge dans la *Débâcle* et une ROSE servante dans le *Docteur Pascal*. Une seule ROSE accède à la condition de bourgeoise : c'est dans *Pot-Bouille* la femme de Campardon qui, à dire vrai, ne s'élève guère au-dessus des autres. Il en est ainsi des CLEMENCE, des AUGUSTE, des ROSALIE ou des ALEXANDRE. De tels noms deviennent alors des prête-noms, des symboles, des formes à contenu variable, et l'on se rapproche alors

du statut des noms communs. Il y a presque une typologie des noms propres chez Zola, qui s'explique par une attention soutenue portée par lui au choix des noms de ses héros, comme le montrent les hésitations et les corrections multiples de ses manuscrits et son obstination à maintenir les noms une fois choisis, en dépit de procès<sup>5</sup>.

Cependant Zola est loin de multiplier les noms propres dans ses romans et la comparaison avec la littérature de son temps laisse apparaître un déficit, qui demeure lorsque la toile de fond s'étend à l'ensemble du corpus du *Trésor de la langue française*, de 1789 à nos jours. Voici ce que disent les chiffres, après que les normes de dépouillement ont été ajustées les unes aux autres dans un sens extensif<sup>6</sup>.

	<i>norme corpus</i>	<i>norme prose époque</i>
étendue de la norme	70 273 552	15 931 175
effectif réel dans la norme	2 234 448	565 974
effectif réel chez Zola	81 795	81 795
effectif théorique chez Zola	91 407	102 129
écart absolu	-9 612	-20 334
écart réduit	-32,46	-70,29

Quel que soit le terme de la comparaison, le déficit des noms propres est patent chez Zola. C'est ce qu'on observe aussi lorsqu'on applique à Chateaubriand ou à Proust la mesure (ou « norme ») de Zola<sup>7</sup>.

	<i>Chateaubriand</i>	<i>Proust</i>
étendue	1 398 984	1 267 069
effectif réel	61 464	42 707
effectif théorique d'après Zola	39 805	36 052
écart absolu	+21 659	+655
écart réduit	+151,52	+48,92

On s'expliquerait aisément la rareté des noms propres si l'on avait affaire à un essai, à quelque spéculation métaphysique détachée du monde réel et située hors du temps, de l'espace et de l'humanité. Tel n'est pourtant pas le cas de Zola qui ne s'écarte guère de l'objet concret et qui préfère montrer des hommes plutôt que des abstractions. Mais il est plusieurs façons de montrer les hommes. Chateaubriand montre les individualités, les héros qui font l'histoire au lieu que Zola est plus soucieux de faire apparaître la société, les êtres et les mouvements

5. On songe notamment au procès que le nom d'un héros de *Pot-Bouille* a suscité.

6. C'est-à-dire sans exclusion des particules et des titres et sans regroupement des noms composés.

7. Comme les corpus de Proust et de Chateaubriand sont moins étendus que celui de Zola, il a fallu renverser la perspective et choisir Zola pour pôle de référence.

collectifs. Quand le premier se soucie du combat des chefs, le second s'intéresse aux mouvements de foule, à cette marée montante où les volontés propres – et les noms propres – sont noyées. Sans doute aussi Zola est-il peu sensible à ce sentiment aristocratique qui attache tant de prix aux noms et aux titres et que Proust partage avec Chateaubriand.

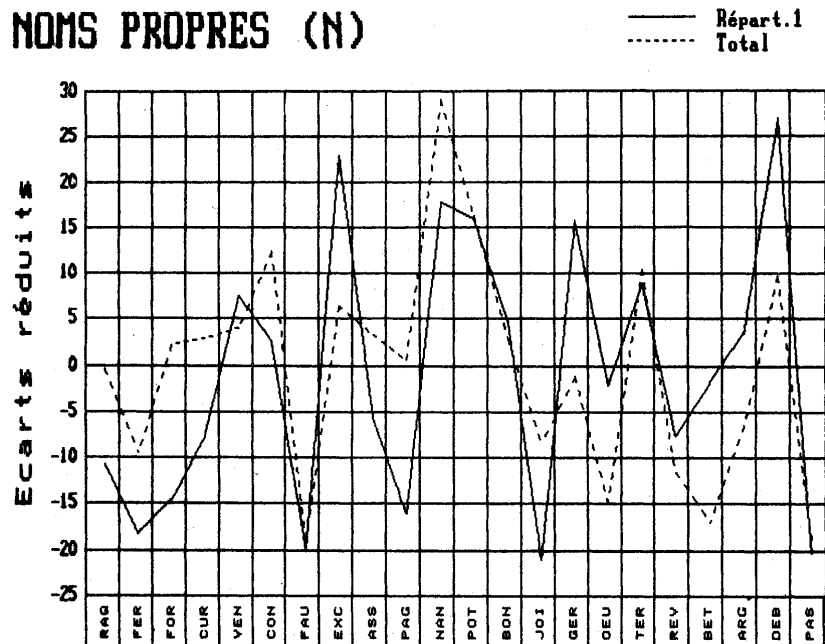
Cependant le déficit des noms propres n'est pas constant chez Zola et les comparaisons internes laissent voir d'amples variations d'un texte à l'autre. On a d'abord constitué les effectifs des noms propres dans chaque texte et on les trouvera dans la colonne de droite du tableau 1 pour les vocables (ligne supérieure) et les occurrences (ligne inférieure). Précisons qu'il s'agit là de données corrigées, d'où les particules et titres ont été expurgés, comme aussi certains éléments de composition faciles à décanter (notamment SAINT et SAINTE). Ainsi corrigé l'effectif des noms propres tombe à 72 014 (pour 2940 vocables). Un calcul d'écart réduit, appliqué aux occurrences permet d'obtenir la courbe supérieure du graphique 2 dont le point culminant coïncide avec la période centrale de *Nana*, *Pot-Bouille* et *Au Bonheur des Dames*, avec des sommets secondaires au niveau de la *Conquête*, de *Son Excellence*, de la *Terre* et de la *Débâcle*. Les noms propres se multiplient chaque fois que l'action s'intensifie et se diversifie et que la scène se remplit, que ce soit dans la chambre-carrefour de *Nana*, dans l'escalier collectif de *Pot-Bouille*, ou l'escalier majestueux d'un grand magasin, dans la ferme âprement disputée des Buteau ou sur les routes encombrées de la *Débâcle*. Ils se raréfient au contraire dans la campagne aride de la Haute Provence (*La Faute de l'abbé Mouret*), dans la solitude d'un petit port breton (*La joie de vivre*), dans la fuite obsessionnelle du criminel (*Thérèse Raquin*, la *Bête humaine*), dans l'ombre de la cathédrale (*Le Rêve*) ou dans la retraite volontaire du savant (le *Docteur Pascal*). Il ne s'agit pas seulement de l'opposition entre la foule et le désert mais peut-être aussi d'une question de rythme narratif, de technique de découpage. Quand les noms propres abondent c'est que les plans se succèdent plus rapidement. Quand ils s'espacent c'est que les séquences s'allongent et que les personnages restent plus longtemps dans le champ. Il ne faut pas oublier en effet que les pronoms de rappel se substituent aux noms propres quand l'ambiguïté n'est pas menaçante, c'est-à-dire quand le même personnage se maintient au devant de la scène. Et cela expliquerait peut-être la position moyenne de *Germinal* qui ne manque pas de personnages mais qui les met en branle sur un rythme lent d'oratorio.

Tableau 1. Les classes de répartition des noms propres  
(Ligne 1 : vocables – Ligne 2 : occurrences)

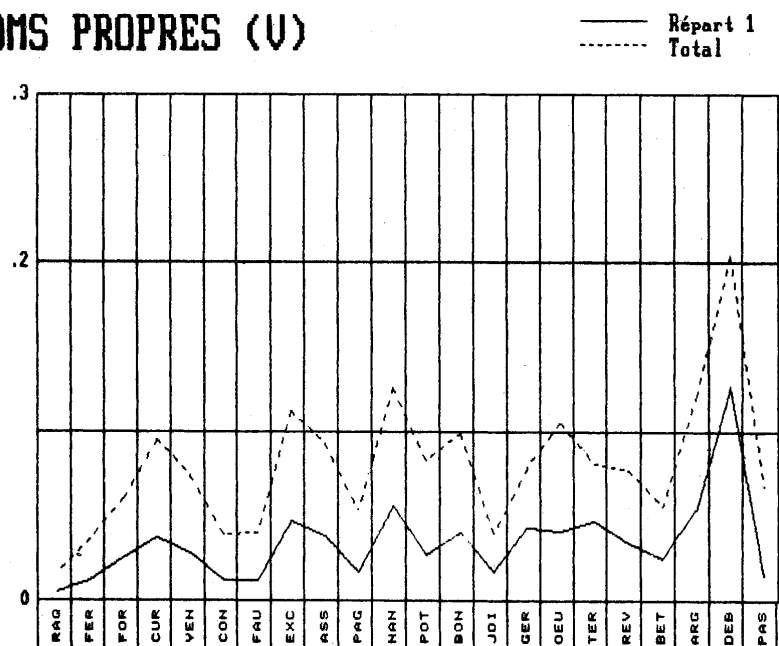
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	20	22	
RAQ	10	3	3	5	5	3	1	0	1	2	0	1	0	0	0	1	0	0	2	37
	357	281	44	5	464	12	459	0	1	13	0	17	0	0	0	28	0	0	23	1704
FER	28	6	13	4	5	6	1	3	2	5	2	1	3	0	1	2	1	2	2	87
	379	14	207	4	15	514	2	45	3	29	3	248	5	0	1	3	506	18	97	2093
FOR	59	18	10	12	7	6	3	7	2	6	3	3	2	2	2	2	0	2	2	148
	627	696	101	142	281	92	12	550	12	121	13	19	8	21	196	176	0	40	53	3160
CUR	90	27	24	19	11	13	5	5	6	7	9	4	1	2	2	3	1	2	2	233
	747	305	490	265	628	87	34	57	8	13	24	26	5	4	2	29	8	19	125	2876
VEN	67	25	14	12	7	9	8	7	6	3	7	5	3	1	1	3	1	2	2	183
	1267	147	774	47	31	168	334	49	31	26	21	54	7	3	5	14	1	15	81	3075
CON	28	14	14	6	4	3	2	5	1	5	3	2	0	1	2	1	0	2	2	95
	1147	525	535	184	97	38	44	602	6	202	4	19	0	2	14	132	0	4	72	3627
FAU	27	15	12	7	7	2	5	5	2	4	3	0	1	1	1	1	0	1	2	96
	451	495	288	60	145	32	49	177	2	78	5	0	60	3	1	10	0	2	158	2016
EXC	114	37	28	21	16	6	6	9	6	9	6	5	3	1	2	3	1	2	2	277
	1953	397	60	38	43	13	12	811	23	27	24	24	24	3	2	31	1	59	93	3638
ASS	92	35	18	12	11	15	5	5	5	7	5	4	3	1	2	3	1	2	2	228
	1279	1326	269	31	798	159	8	20	265	18	27	15	5	2	2	24	1	9	65	4323
PAG	41	17	12	10	10	7	2	4	4	6	4	1	3	2	2	3	1	2	2	133
	472	230	261	167	41	526	4	608	26	10	12	9	17	125	24	17	3	3	138	2693
NAN	136	56	28	19	11	12	9	6	4	9	7	4	1	1	1	2	1	2	2	311
	1934	1222	1224	72	48	124	65	22	13	323	99	28	1	7	1	14	4	27	120	5348
POT	65	30	25	19	10	6	8	5	4	11	7	2	3	1	1	2	1	2	2	204
	1811	513	204	448	229	31	616	257	32	70	31	4	98	1	6	9	5	9	67	4441
BON	96	42	33	15	11	10	5	1	4	4	8	2	2	2	1	3	1	2	2	244
	1553	940	403	64	172	141	21	404	11	8	45	3	10	73	1	6	6	21	129	4011
JOI	41	12	10	5	8	3	6	1	3	1	2	0	0	0	1	1	1	1	2	98
	427	669	83	55	16	778	18	2	4	503	18	0	0	0	2	1	1	3	62	2642
GER	104	23	13	8	6	8	2	3	4	4	6	3	1	2	1	1	1	2	2	194
	2155	841	197	25	23	539	236	22	7	12	26	5	17	70	2	1	14	6	43	4241
OEU	100	41	26	24	13	8	5	3	7	7	9	3	2	2	2	2	1	2	2	259
	1171	158	72	246	46	536	10	10	24	16	46	44	3	6	9	58	4	18	134	2611
TER	114	19	13	10	4	5	3	3	2	7	3	2	2	1	2	1	2	1	2	198
	1886	1051	442	102	7	335	3	8	38	490	11	12	10	353	1	123	2	15	58	4947
REV	82	30	16	13	9	5	5	3	2	4	4	4	2	2	1	3	1	2	2	190
	441	373	180	34	12	7	12	4	2	36	9	7	52	21	1	4	1	17	57	1270
BET	61	16	7	11	6	8	1	1	4	5	5	3	0	1	0	3	1	2	2	137
	1141	333	40	122	11	12	90	3	94	11	35	348	0	24	0	11	1	6	100	2382
ARG	132	49	26	22	12	15	4	5	4	9	6	3	3	2	2	3	1	2	2	302
	1480	313	84	121	624	246	13	50	45	137	33	15	24	7	11	13	3	38	131	3388
DEB	313	50	31	26	22	14	7	8	5	7	6	4	2	1	2	2	1	2	2	505
	2859	838	348	108	301	79	49	29	15	28	48	96	12	432	2	17	2	79	194	5536
PAS	34	43	17	12	10	10	5	7	3	8	5	3	2	1	2	2	0	2	2	168
	426	98	81	57	552	350	17	163	7	24	6	11	4	10	15	103	0	14	54	1992
TOT	1834	304	131	73	41	29	14	12	9	13	10	5	3	2	2	3	1	2	2	2940
	25963	11765	6387	2397	4584	4819	2108	3893	669	2195	540	1004	362	1167	298	824	563	422	2054	72014



### NOMS PROPRES (N)



### NOMS PROPRES (V)



Graphique 2. Courbes de répartition des noms propres  
(N : occurrences - V : vocables)

La variété des personnages et des lieux est plus facile à mesurer si l'on prend en compte les vocables et non plus les occurrences et si l'on considère la courbe inférieure du nombre de vocables appartenant aux noms propres. On compte ainsi le nombre relatif de personnages (ou lieux) différents, cités dans chaque texte. Cette courbe ne recouvre pas exactement la précédente. Certes l'encombrement est toujours visible dans les lieux publics ou privés où l'on voit Nana et Eugène Rougon et c'est le même entassement dans les grands magasins et la même cohue des soldats en déroute. Mais on observe une distinction plus systématique entre Paris et la province. La foule est parisienne et la solitude provinciale. Quand l'action prend pour cadre les rues, les maisons et les salons de la capitale, les personnages sont plus nombreux et plus variés et aux titres déjà cités d'ajoutent la *Curée*, le *Ventre de Paris*, l'*Œuvre*, et l'*Argent*. Mais deux romans restent en retrait, *Une Page d'amour* et *Le Rêve*, sans doute parce qu'ils mettent en scène une histoire d'amour et que l'amour cherche la solitude. En revanche, tous les romans de la province (à l'exception de la *Débâcle*, qui s'achève, il est vrai, à Paris) se situent dans les creux de la courbe, de la *Fortune* au *Docteur Pascal*.

Mais cette distinction balzacienne, entre les scènes de la vie de province et les scènes de la vie parisienne, se combine avec une évolution assez visible sur le graphique. La scène zolienne tend à se diversifier et la courbe des noms propres est montante pour les vocables. Si l'on fait pour chaque texte le rapport  $N/V$  des noms propres, et qu'on calcule une fréquence moyenne de chaque nom cité, on obtient la série suivante :

46, 24, 21, 12, 17, 38, 21, 13, 19, 20, 17, 22, 16, 27, 22, 10, 25, 7, 17, 11, 11, 12.

Comme il y a plus de monde autour du gâteau, chacun voit sa part diminuer. La corrélation chronologique est certaine, le coefficient s'élevant à -0,54, soit une probabilité de 0,01.

La galerie des personnages et des lieux tend aussi à se spécialiser. Zola a toujours affirmé l'unité organique du Cycle des *Rougon-Macquart*. Il semble pourtant que le lien se relâche en cours de rédaction et que les derniers romans, le *Docteur Pascal* excepté, tendent à se détacher de la série. C'est du moins ce que l'étude de la répartition des noms propres laisse entendre. Le tableau 1 détaille pour chaque texte le nombre de vocables (et d'occurrences) que ce texte a en commun avec un autre texte, ou avec 2, 3, 4 autres. C'est ce qu'on appelle les classes de répartition qui dans le cas présent s'échelonnent de 1 à 22. À vrai dire il eût été logique d'exclure les deux premiers textes du corpus qui

n'appartiennent pas au cycle des *Rougon-Macquart*. Nous les avons gardés cependant pour éprouver les vertus de la méthode.

Une première remarque s'impose, devant les résultats du tableau 1 : la cohésion du cycle n'apparaît pas très forte puisque les deux tiers des noms propres n'apparaissent que dans un seul roman à l'exclusion de tous les autres (1834 sur 2490 vocables). Même si on raisonne au niveau des occurrences, la part privative de chaque texte est considérable (25 963 sur 72 014). On ne rencontre que 2 noms propres communs (si l'on peut dire) à tous les textes et il n'y en a aucun dans la classe de répartition 21, non plus que dans les classes 18 et 19<sup>8</sup>. Au total on ne trouve que 20 mots qui soient présents dans au moins la moitié des textes (c'est-à-dire de la classe 12 à la classe 22). C'est dire que l'arbre généalogique des *Rougon-Macquart* est loin de suffire à l'intrigue. Au reste cet arbre même est fait de tronçons dont beaucoup apparaissent isolés. Cette spécificité des noms propres éclate dans le tableau 3, qui restitue le début de la liste alphabétique. La plupart des noms sont réservés à un texte unique, même lorsque la fréquence est élevée (par exemple BECU 109 occurrences, ARCHANGIAS 73, BAZEILLES 89, BACHELARD 122, etc.) et ALBINE qui a 327 emplois n'apparaît qu'une fois en dehors de la *Faute*.

La seconde observation – qui peut être dégagée de la figure 2 – a trait à l'évolution de Zola. La courbe en pointillés, relative à la classe des noms propres exclusifs (de répartition 1), n'est pas tout-à-fait parallèle à celle de l'ensemble, toutes classes réunies. Elle est en-deçà dans la première moitié chronologique et au-delà dans la deuxième – ce qu'on peut montrer en calculant le pourcentage, pour chaque texte, de la répartition 1 sur l'ensemble des noms propres :

Texte	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	Ra	Fé	Fo	Cu	Ve	Co	Fa	Ex	As	Pa	Na	Po	Bo	Jo	Ge	Oe	Te	Re	Be	Ar	Dé	Pa
% (N)	21	16	20	26	41	32	22	54	30	16	36	41	39	16	51	45	36	35	48	44	52	21
% (V)	27	32	40	39	37	29	28	41	40	31	44	32	39	42	54	39	58	43	45	44	62	20

Le coefficient de corrélation ( $r = 0,47$ ) ne laisse guère de doute ( $p = 0,03$ ), comme celui qu'on tire du même rapport au niveau des vocables ( $r = 0,43$ ,  $p = 0,05$ ). Il y a donc un mouvement excentrique qui s'accroît au fil des années et qui tend à projeter les textes en dehors du cycle.

---

8. Si on limite l'enquête aux vingt textes du cycle des *Rougon-Macquart*, les chiffres doivent être ainsi corrigés :

répartition :	20	19	18	17	16	15	14
effectif :	2	2	0	0	2	2	4

Tableau 3. Liste alphabétique des noms propres (extrait)

	F	R		F	R		F	R		F	R
ABAGAR	1	1	ABBE	1	1	ABD	1	1	ABEL	2	2
ABELARD	1	1	ABISAIG	12	1	ABOUKIR	1	1	ABRAHAM	6	4
ABYSSINIE	2	1	ACHILLE	19	4	ACRE	6	2	ADAM	2	1
ADELAIDE	61	2	ADELE	109	5	ADELIN	16	1	ADOLPHE	10	1
ADRESSE	1	1	ADRIATIQUE	2	2	AFRIQUE	37	5	AGAR	2	1
AGATHE	3	1	AGATHON	1	1	AGNES	42	2	AGRA	1	1
AIGRE	38	1	AISNE	12	1	ALBERT	43	2	ALBI	1	1
ALBINE	327	2	ALBOISE	4	1	ALENCON	15	4	ALEP	7	1
ALEXANDRE	41	7	ALEXANDRIE	4	1	ALEXIS	1	1	ALFRED	1	1
ALGER	3	1	ALGERIE	14	5	ALICE	33	2	ALLEMAGNE	50	7
ALMA	1	1	ALPES	4	3	ALPHONSE	12	2	ALSACE	11	3
ALTKIRCH	2	1	ALZIRE	49	2	AMADIEU	12	1	AMANDA	3	1
AMBROISE	7	3	AMELIE	9	4	AMERIQUE	25	11	AMIENS	6	4
AMOURS	1	1	AMSTERDAM	8	3	ANACREON	1	1	ANASTASIE	4	1
ANATOLE	2	1	ANDELYS	1	1	ANDRE	6	4	ANDROGLES	1	1
ANDROMEDE	1	1	ANGE	1	1	ANGECOURT	1	1	ANGELE	99	5
ANGELINE	11	1	ANGELIQUE	242	2	ANGERS	13	2	ANGLARS	4	1
ANOLETTERRE	34	12	ANGORA	10	1	ANGOULEME	7	3	ANJOU	5	4
ANNA	5	2	ANNE	25	6	ANNOUCHKA	2	1	ANSELME	1	1
ANTIN	9	4	ANTIOCHE	3	1	ANTIOPE	1	1	ANTOINE	102	10
ANTONIA	19	1	ANZIN	3	1	APIA	1	1	APOLLON	2	1
AQUIN	2	1	ARABIE	2	2	ARARAT	1	1	ARC	2	2
			ARCHANGIAS	73	1	ARCHE	2	1	ARCOLE	6	2
ARCUEIL	2	1	ARDENNES	13	2	ARGENTEUIL	5	3	ARGONNE	3	1
ARISTIDE	124	4	ARLES	1	1	ARMENIE	1	1	ARNE	1	1
ARRAS	2	2	ARROMANCHES	30	1	ARTAUD	57	2	ARTHENAY	1	1
ARTHUR	1	1	ASIE	15	2	ASNIERES	8	5	ASPASIE	1	1
ASSAS	5	1	ASTORG	4	1	ATALANTE	1	1	ATHENES	1	1
ATTIGNY	1	1	AUBERIVE	2	1	AUBERTOT	12	1	AUBIN	1	1
AUBLETS	4	1	AUBRY	2	1	AUBUSSON	2	1	AUDEMER	1	1
AUDIERNE	1	1	AUGEREAU	1	1	AUGUSTA	1	1	AUGUSTE	301	8
AUGUSTIN	46	5	AUGUSTINE	75	2	AULNAY	3	1	AUMALE	1	1
AUNEAU	2	1	AURELIE	149	3	AURIGNY	29	1	AUSTERLITZ	14	8
AUSTRALIE	1	1	AUSTREBERTHE	1	1	AUTEUIL	5	1	AUTHE	2	1
AUTRECOURT	2	1	AUTRICHE	16	4	AUVERGNAT	1	1	AUVERGNE	6	4
AUXERROIS	1	1	AVIGNON	1	1	AVRON	1	1	BABEL	3	2
BABET	11	1	BABYLONE	4	2	BAC	10	1	BACHELARD	122	1
BADE	8	4	BADEUIL	3	1	BADINGUE	7	1	BADINGUET	10	3
BAGDAD	4	1	BAGNEUX	2	1	BAGNOLET	1	1	BAILLEHACHE	45	1
BAILLY	1	1	BAKOUNINE	1	1	BALAN	26	1	BALBI	18	1
BALBINE	5	1	BALE	1	1	BALLAY	3	1	BALTHAZAR	15	1
BALZAC	5	3	BAMBOUSSE	35	1	BANBAN	29	1	BAPTISTE	22	3
BAPTISTIN	3	1	BAQUET	5	1	BAR	2	2	BARATTE	5	1
BARBARIE	2	2	BARBE	8	4	BARBEVILLE	1	1	BARENTIN	50	1
BARILLOT	18	1	BARREUX	1	1	BART	33	1	BARTHELEMY	1	1
BASILE	1	1	BASSES	1	1	BASTIAN	4	1	BASTIEN	1	1
BASTILLE	7	4	BATAILLE	20	1	BATARD	8	1	BATIGNOLLES	47	11
BAUDEQUIN	12	2	BAUDU	147	3	BAUGE	40	1	BAVIERE	3	1
BAVOUX	2	1	BAYBEL	5	1	BAYEUX	9	1	BAYON	1	1
BAZAINE	37	1	BAZEILLES	89	1	BAZUCHES	22	1	BAZOUGE	32	1
BEAGE	4	2	BEAU	1	1	BEAUCE	68	2	BEAUCHAMP	3	1
BEAUCLAIR	2	1	BEAUDEQUIN	1	1	BEAUDOIN	53	1	BEAUDOUIN	4	2
BEAUGENCY	1	1	BEAUGNIES	3	1	BEAUMONT	61	3	BEAURECUEIL	1	1
BEAURIVAGE	4	1	BEAUVILLIERS	45	1	BEBERT	51	1	BEC SALE	28	1
BECKER	1	1	BECON	1	1	BECOT	15	1	BECOURT	6	1
BECU	109	1	BEDORE	6	1	BEETHOVEN	7	3	BEJUIN	77	1
BEL	2	2	BELFORT	15	1	BELGIQUE	34	6	BELHOMME	2	1
BELLARMIN	1	1	BELLE	1	1	BELLECHASSE	2	1	BELLEVILLE	13	4
BELLEVUE	6	2	BELLOMBRE	7	1	BELLOQUE	1	1	BENARD	1	1
BENGALE	3	3	BENNECOURT	18	1	BERANGER	3	2	BERAUD	31	2
BERCY	4	3	BERGASSE	1	1	BERGERET	1	1	BERLIERE	2	1
BERLIN	39	5	BERLINGOT	1	1	BERLIOZ	3	2	BERLOQUE	1	1
BERNARD	4	2	BERNAY	1	1	BERNHAIM	4	1	BEROND	1	1
BERTHE	386	4	BERTHIER	32	3	BERTHOU	1	1	BERTRAND	2	1
BESACE	10	1	BESANCON	24	3	BESNUS	3	1	BESSIERE	8	1
BESSIN	3	1	BETHINVILLE	2	1	BETHLEEM	2	2	BETHUNE	3	2
BEULIN	22	1	BEUVILLE	5	1	BEUZEVILLE	2	1	BEYROUT	20	1
BIARRITZ	3	2	BIBI	45	2	BICETRE	5	3	BIEVRE	1	1
BIGNON	2	2	BIJARD	33	1	BIJOU	10	1	BILLAULT	1	1
BILLECOQ	15	1	BINEAU	7	1	BISKRA	1	1	BISMARCK	51	3

ETC

Est-ce la fatigue, la saturation ou la crainte de produire ces effets sur le lecteur ? En tous cas, Zola renouvelle son personnel et son cadre quand il s'approche de la fin de l'entreprise, et particulièrement dans les grandes fresques : *Germinal*, *la Terre* et *la Débâcle*. Dans ces trois textes plus de

la moitié des noms propres sont exclusifs, soit respectivement 54 % des vocables, 58 % et 62 %<sup>9</sup>. Il faut cependant mettre à part le dernier titre de la série, le *Docteur Pascal*, qui est le moins excentrique et cultive au contraire les biens communaux. La part privative est la plus faible de toute la série (20 %). C'est que Zola tenait depuis longtemps ce livre (ou le sujet de ce livre) en réserve, et qu'il le destinait à servir de conclusion au cycle entier. Et l'on trouve effectivement dans la bouche du docteur le bilan approfondi de toute la série et le rappel des titres antérieurs. Et c'est pourquoi le *Docteur Pascal* a des attaches privilégiées avec tous les autres romans du cycle. C'est le carrefour où se croisent tous les fils de l'intrigue. Et c'est pourquoi l'effectif de la répartition 2 est supérieur à celui de classe 1 (43 contre 34) – ce qui ne s'observe dans aucun autre texte.

Mais la distinction entre la part exclusive et la part commune est sans doute trop abrupte. Il est des degrés entre le privé et le public et précisément les 22 classes de répartition représentent toutes les graduations du semi-privé ou du semi-public. Et c'est l'ensemble des 22 classes dans les 22 textes (c'est-à-dire la matrice carrée du tableau 1) qu'il faut soumettre au calcul. Dans de tels cas l'analyse factorielle fournit l'outil approprié. Nous avons appliqué cette méthode aux vocables et reproduit les résultats dans la figure 5. Mais comme au-delà de la classe 6 les effectifs deviennent faibles, on a procédé à quelques regroupements, les classes 7 et 8 constituant la 7<sup>e</sup> colonne, les classes 9, 10 et 11 la 8<sup>e</sup>, et respectivement les classes 12 à 15 et 16 à 20 les colonnes 9 et 10. On obtient alors 10 classes moins inégales dont les effectifs sont représentés dans la partie supérieure du tableau 4 et dont l'analyse factorielle apparaît en italique dans le graphique 5. Mais le soupçon nous est venu que l'équilibre des lignes et celui des colonnes n'étaient peut-être pas suffisamment respectés. Une contraction supplémentaire, en 6 colonnes seulement, a été réalisée, qui regroupe respectivement les colonnes 4 et 5, 6 et 7 et enfin 8, 9 et 10. Et pour faire taire tout scrupule nous avons pondéré les colonnes et les lignes, de façon à opérer sur des profils et à annuler l'effet de taille. On lira dans la moitié inférieure du tableau 4 la nouvelle disposition des données et le détail des résultats fournis par le programme d'analyse<sup>10</sup>.

---

9. Les pourcentages seraient certainement plus forts, si l'on pouvait distinguer les homonymes.

10. Il s'agit de l'analyse de correspondance du professeur Benzécri et du logiciel ADDAD qui est très répandu en France.

Tableau 4. Analyse factorielle de la répartition

NOM(J)	1	2	3	4	5	6	7	8	9	10	
FJ(J)	1834.	608.	393.	292.	205.	174.	194.	321.	157.	149.	4327.
RAQ	10.	3.	3.	5.	5.	3.	1.	3.	1.	3.	37.
FER	28.	6.	13.	4.	5.	6.	4.	9.	5.	7.	87.
FOR	59.	18.	10.	12.	7.	6.	10.	11.	9.	6.	148.
CUR	90.	27.	24.	19.	11.	13.	10.	22.	9.	8.	233.
VEN	67.	25.	14.	12.	7.	9.	15.	16.	10.	8.	183.
CON	28.	14.	14.	6.	4.	3.	7.	9.	5.	5.	95.
FAU	27.	15.	12.	7.	7.	2.	10.	9.	3.	4.	96.
EXC	114.	37.	28.	21.	16.	6.	15.	21.	11.	8.	277.
ASS	92.	35.	18.	12.	11.	15.	10.	17.	10.	8.	228.
PAG	41.	17.	12.	10.	10.	7.	6.	14.	8.	8.	133.
NAN	136.	56.	28.	19.	11.	12.	15.	20.	7.	7.	311.
POT	65.	30.	25.	19.	10.	6.	13.	22.	7.	7.	204.
BON	96.	42.	33.	15.	11.	10.	6.	16.	7.	8.	244.
JOI	41.	12.	10.	5.	8.	3.	7.	6.	1.	5.	98.
GER	104.	23.	13.	8.	6.	8.	5.	14.	7.	6.	194.
OEU	100.	41.	26.	24.	13.	8.	8.	23.	9.	7.	259.
TER	114.	19.	13.	10.	4.	5.	6.	12.	8.	7.	198.
REV	82.	30.	16.	13.	9.	5.	8.	10.	9.	8.	190.
BET	61.	16.	7.	11.	6.	8.	2.	14.	4.	8.	137.
ARG	132.	49.	26.	22.	12.	15.	9.	19.	10.	8.	302.
DEB	313.	50.	31.	26.	22.	14.	15.	18.	9.	7.	505.
PAS	34.	43.	17.	12.	10.	10.	12.	16.	8.	6.	168.

DONNEES REPARTIES EN 10 CLASSES

1	2	3	4	5	6	11	QLT	POID	INR	1EF	COR	CTR	2EF	COR	CTR	
999.	999.	999.	1000.	999.	999.	5995	RAQ	995	40	152	-234	236	86	346	511	338
RAQ	25.	21.	31.	80.	42.	241	FER	1000	55	113	-182	262	71	85	57	28
FER	42.	25.	79.	43.	63.	328	FOR	978	47	18	-2	0	0	91	356	28
FOR	49.	42.	34.	50.	56.	283	CUR	884	46	5	-36	193	2	55	425	10
CUR	45.	38.	49.	48.	49.	277	VEN	929	51	35	-53	68	6	4	0	0
VEN	46.	49.	39.	42.	70.	303	CON	996	55	48	-143	384	44	-95	173	36
CON	38.	53.	77.	43.	57.	330	FAU	849	56	31	-159	754	56	-52	81	11
FAU	36.	56.	65.	59.	67.	334	EXC	994	44	7	41	160	3	12	14	0
EXC	48.	44.	48.	49.	37.	266	ASS	921	45	14	38	75	2	-51	146	9
ASS	47.	50.	37.	37.	53.	267	PAG	778	52	31	-131	479	35	95	250	33
PAG	39.	46.	47.	61.	52.	314	NAN	971	41	31	126	344	25	-157	542	72
NAN	48.	56.	40.	34.	40.	29	POT	977	51	14	-102	617	21	-25	38	2
POT	39.	51.	61.	55.	48.	306	BON	984	43	42	43	30	3	-166	454	84
BON	43.	53.	60.	37.	30.	257	JOI	229	45	12	18	21	1	21	27	1
JOI	48.	40.	48.	49.	49.	34	GER	990	37	54	284	894	116	18	4	1
GER	60.	37.	30.	26.	31.	221	OEU	761	44	20	30	32	2	-19	14	1
OEU	44.	51.	46.	52.	30.	41	TER	987	35	85	346	836	165	82	47	17
TER	64.	30.	30.	25.	26.	37	REV	836	42	15	106	528	19	-45	99	6
REV	49.	51.	39.	42.	33.	39	BET	767	42	41	99	161	16	155	394	70
BET	52.	38.	24.	46.	36.	54	ARG	994	41	16	115	551	21	-75	239	17
ARG	48.	50.	38.	39.	37.	32	DEB	942	32	113	441	887	241	83	31	15
DEB	64.	29.	26.	31.	25.	17	PAS	996	56	107	-172	253	65	-235	472	220
PAS	25.	89.	51.	52.	68.	53										

EN 6 CLASSES															
						11	QLT	POID	INR	1EF	COR	CTR	2EF	COR	CTR
						5995.0	1000					1000			1000
J1															
						1	QLT	POID	INR	1EF	COR	CTR	2EF	COR	CTR
						1	998	167	328	340	950	748	66	36	51
						2	976	167	168	43	30	12	-210	711	521
						3	993	167	142	-87	146	50	-83	132	82
						4	1000	167	164	-88	130	51	151	374	267
						5	749	167	95	-110	351	80	-4	1	0
						6	766	167	103	-94	238	59	82	177	79
						5995.0	1000					1000			1000

VAL PROPRE	POURCENT	CUMUL	HISTOGRAMME DES VALEURS PROPRES DE LA MATRICE
0.02577788	41.624	41.624	*****
0.01424652	23.004	64.627	*****
0.01097370	17.719	82.347	*****
0.00761385	12.294	94.641	*****
0.00331901	5.359	100.000	*****

On verra que la qualité de représentation est excellente pour toutes les lignes et toutes les colonnes puisqu'elle dépasse 90 % pour les deux tiers des variables et qu'elle ne descend pas au-dessous de 75 %. Les

deux premiers facteurs expliquent respectivement 42 % et 23 % de la variance et ce sont eux qu'on a représentés dans le graphique 5 (en caractères romains). Observons tout d'abord que les deux analyses se recouvrent et que la différence entre les regroupements en 10 ou 6 classes a peu d'effet dans les résultats, au moins dans le cas présent.

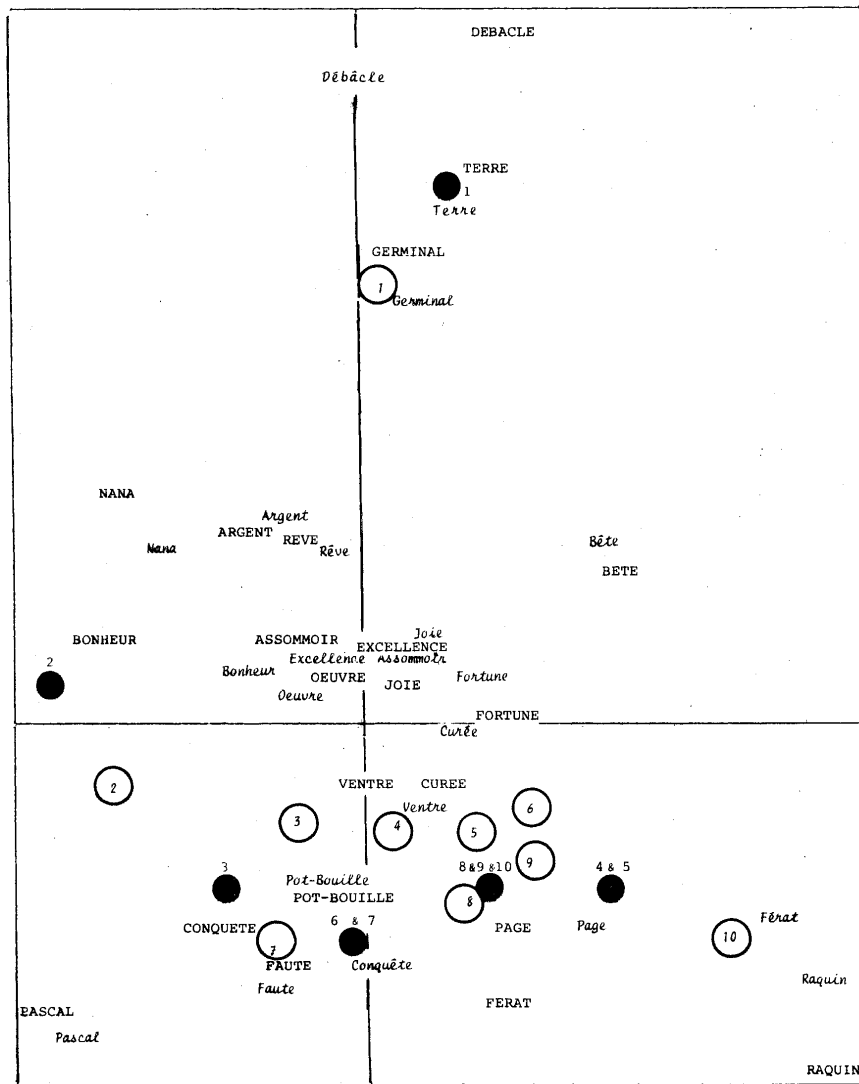


Figure 5. Analyse factorielle des classes de répartition (en italique les données groupées en dix classes)

On ne sera pas surpris que l'analyse reconnaisse des données sérielles et qu'elle distribue dans l'ordre les classes de répartition, en suivant un dessin en forme de croissant, bien connu des statisticiens. La classe 1 occupe un quadrant, la classe 2 un autre, la classe 3 le troisième et les autres classes celui qui reste. Ainsi fixé le jeu des quatre coins, les textes prennent position selon la chronologie : les 7 premiers titres campent en bas dans la zone commune, les 10 derniers (sauf un) en haut dans la zone privée et les 5 titres de la période centrale flottent non loin de la ligne de démarcation. Il y a donc une évolution chronologique qui pousse Zola à abandonner les lieux ou personnages déjà connus du lecteur pour lui en proposer d'autres. Le cycle des *Rougon-Macquart* perd donc progressivement de sa force d'attraction et l'on retrouve au haut du graphique les trois blocs erratiques que nous avons déjà rencontrés et qui échappent à la gravitation : *Germinal*, la *Terre* et la *Débâcle*. À l'opposé les deux romans extérieurs au cycle des *Rougon-Macquart*, *Thérèse Raquin* et *Madeleine Férat* sont rejetés à l'autre bout de la chaîne, là où l'on partage le strict minimum, c'est-à-dire les mots de large répartition qui s'imposent à tout sujet contemporain et qui pour la plupart sont relatifs au cadre géographique (comme PARIS, ou la FRANCE). La méthode est donc consolidée puisqu'elle confirme l'exterritorialité évidente des premiers romans de notre corpus. Entre les deux extrêmes, dans la portion renflée du croissant, on attachera une particulière attention à la place du *Docteur Pascal* qui échappe complètement à l'influence chronologique pour occuper le coin gauche inférieur là où se concentrent les classes intermédiaires de répartition. Ainsi se trouve précisée la fonction unificatrice du dernier texte des *Rougon-Macquart* qui a conclu avec tous les autres des accords bilatéraux. Finale de la symphonie, le *Docteur Pascal* donne un écho prolongé aux thèmes déjà développés et aux noms déjà entendus.

Reste à identifier les noms propres dont l'étude qui précède n'envisageait que les classes et les effectifs. Le tableau 6 dresse la liste des 300 unités les plus fréquentes, classée par ordre décroissant. C'est le mot PARIS qui arrive en tête avec 1507 occurrences disséminées dans tous les textes du corpus (répartition 22). Suivent 26 noms ou prénoms de personnages avant qu'on rencontre l'autre pôle géographique du cycle, PLASSANS, qui n'a que 451 occurrences et qui n'apparaît que dans 16 textes. Plus loin encore (rang 51) on trouve la FRANCE (276 emplois), MONTSOU (rang 90, fréquence 160) et SEDAN (rang 95, fréquence 155).



Tableau 6. Liste des noms propres par fréquence

	F	R		F	R		F	R		F	R
PARIS	1507	22	BOCHE	219	2	GILQUIN	111	1	BEJUIN	77	1
ROUGON	1299	8	MIGNON	212	3	SAGET	110	1	MARGUERITE	76	5
MOURRET	1091	8	HENRI	211	14	JORY	110	1	PALOQUE	76	1
NANA	1033	3	GEORGES	189	10	MAHOUDEAU	110	1	SARRIETTE	76	1
JEAN	956	14	MARTINE	186	1	ADELE	109	5	ROME	75	10
SACCARD	875	5	KAHN	186	1	BECU	109	1	JOUVE	75	2
PAULINE	831	6	CHRISTINE	185	4	RASSENEUR	109	1	AUGUSTINE	75	2
GERVAISE	791	5	VERONIQUE	185	2	VALERIE	109	1	NOIRAUDE	75	1
JACQUES	674	12	DELAHERCHE	181	1	OLYMPÉ	108	4	MALIGNON	74	1
HELENE	666	8	MAHEUDE	180	2	HOUREDEQUIN	108	1	BONNEMORT	74	1
COUPEAU	656	2	RAQUIN	175	1	SILVINE	106	1	ESTELLE	73	3
CLAUDE	639	6	DEBERLE	173	1	HAVRE	105	7	CADINE	73	1
ETIENNE	575	6	TRUBLLOT	169	1	WEISS	105	1	MATHIEU	72	3
MADELEINE	563	17	TEUSE	165	1	DESFORGES	104	1	ROCHAS	72	1
GUILLAUME	553	6	ZOE	164	2	HONORE	102	12	ARCHANGIAS	72	1
DIEU	547	22	GENEVIEVE	162	3	ANTOINE	102	10	MARSEILLE	71	15
LAZARE	539	10	CAMPARDON	162	1	GRANOUX	100	1	CECILE	71	3
OCTAVE	539	7	RASTOIL	162	1	VOREUX	100	1	JANTROU	71	1
DENISE	529	2	MONTSOU	160	2	ANGELE	99	5	VENUS	70	7
MAURICE	513	2	FAGEROLLES	160	1	PHILIPPE	98	11	SIMONNE	70	2
BUTEAU	496	2	DELHOMME	158	1	ROUEN	98	9	ROUDIER	70	1
ROSE	483	10	BORDENAVE	157	1	THEOPHILE	98	2	JUZEUR	70	1
LAURENT	477	5	CHAVAL	156	1	CHOUTEAU	98	2	MICHAUD	70	1
THERESE	470	7	SEDAN	155	3	BLANCHE	97	6	BOURRETTE	70	1
FLORENT	461	3	GAVARD	155	1	GUNDERMANN	97	1	HEDOUIN	69	2
FAUJAS	459	2	DESIREE	152	5	MACQUERON	97	1	DELOCHE	69	1
JEANNE	453	6	CONDAMIN	152	1	PARADOU	95	2	NAPOLEON	68	12
PLASSANS	451	16	LEVAQUE	152	1	MARJOLIN	94	1	LEON	68	7
CLOTILDE	439	5	AURELIE	149	3	LAURE	93	4	LARSONNEAU	68	2
MARTHE	418	3	LERAT	149	2	FANNY	93	3	BEAUCE	68	2
LISA	403	7	DELESTANG	148	1	BUSCH	93	1	BOURRAS	67	1
BERTHE	386	4	HENNEBEAU	148	1	MARSY	93	1	DAIGREMONT	67	1
MAXIME	383	5	BAUDU	147	3	MARTY	92	1	LECOEUR	67	1
FELICITE	381	5	LOUIS	146	20	MAZAUD	92	1	LEBIGRE	66	1
CHANTEAU	378	2	VANDEUVRES	146	2	MONTMARTRE	91	11	HURET	66	1
FASCAL	369	6	VIRGINIE	145	2	LUCY	91	1	BONGRAND	66	1
RENEE	366	3	GOUJET	144	1	PECQUEUX	90	1	CAZENOVE	65	1
JOSSERAND	351	1	SEINE	143	16	POISSON	90	1	NENESSE	65	1
FRANCOISE	340	6	POIZAT	139	1	GREGOIRE	90	1	PORQUIER	65	1
ALBINE	327	2	HUTIN	136	2	SIDONIE	89	4	PEPE	64	1
LOUISE	319	6	SATURNIN	135	4	ROUQUETTE	89	1	MES BOTTES	64	1
SERGE	319	3	JULIETTE	135	2	FAVIER	89	1	GRIVET	63	1
SILVERE	313	2	CLEMENCE	131	3	DAGUENET	89	1	GILBERTE	63	1
HENRIETTE	305	5	FALOISE	130	2	BAZEILLES	89	1	VIORNE	62	6
AUGUSTE	301	8	SATHI	129	1	FOUCHARD	88	1	VUILLET	62	1
JESUS	300	10	BOURDONCLE	127	1	BONNEVILLE	87	2	CLOYES	62	1
LANTIER	292	9	FRANCOIS	126	12	MISARD	87	1	DENIZET	62	1
LISE	282	3	ARISTIDE	124	4	MEUSE	86	2	ESCORAILLES	62	1
CAMILLE	282	2	RAMBAUD	124	3	HORTENSE	86	2	PIERRON	62	1
CLORINDE	281	2	JEANLIN	124	1	GOURD	86	1	SICARDOT	61	4
FRANCE	276	20	FLORE	122	4	BOVES	85	1	BEAUMONT	61	3
MACQUART	276	8	STEINER	122	2	DUBUCHE	83	1	ADELAIDE	61	2
FOUAN	273	2	CORREUR	122	1	PROSPER	83	1	PEQUEUR	61	1
CATHERINE	268	7	BACHELARD	122	1	RAMOND	83	1	MEHUDIN	61	1
MAHEU	268	2	BOUCHARD	122	1	LUCIEN	82	3	HUGON	61	1
CAROLINE	263	6	EUGENE	121	8	VABRE	82	2	LISON	61	1
LORILLEUX	257	1	JULES	120	9	LOUBET	82	1	MUCHE	60	1
QUENU	251	3	ROGNES	120	2	LAPOLLE	82	1	GASPARINE	60	1
ROUBAUD	251	1	NORMANDE	120	1	GAGNIERE	82	1	VALLAGNOSC	60	1
MUFFAT	249	2	HUBERTINE	120	1	ITALIE	81	13	TROUILLE	60	1
FAUCHERY	249	1	CLARISSE	119	3	LUCIE	81	3	SUZANNE	59	3
SANDOZ	246	1	FELICIE	118	3	HAMELIN	80	2	MOUQUETTE	59	1
ANGELIQUE	242	2	FONTAN	117	1	CLARA	80	2	MAFFRE	58	1
DUVEYRIER	241	1	HUBERT	117	1	GUBULIN	80	1	REMILLY	58	1
MLETTE	240	2	LABORDETTE	117	1	NEGREL	80	1	BOUTHEMONT	58	1
MARIE	239	13	TROUCHE	113	1	DIDE	79	3	ARTAUD	57	2
CHRIST	233	10	ZACHARIE	112	2	LENGAIGNE	79	1	MECHAIN	57	2
CHARLES	230	16	ROBINEAU	112	1	PHILOMENE	78	2	BOURDEU	57	1
PIERRE	227	15	GRANDE	112	1	PAUL	77	11	GAGA	57	1
SEVERINE	224	2	CHARBONNEL	112	1	DELANGRE	77	1	PRIMAT	57	1
ROSALIE	219	4	VICTOR	111	10	CABUCHE	77	1	ORIENT	56	9

ETC.

Il n'y a donc que 5 noms géographiques parmi les 100 premiers de la liste. Encore certains d'entre eux accèdent-ils au rang de symbole et représentent moins un lieu qu'une réalité sociale inscrite dans ce lieu. Dans beaucoup de leurs emplois, PARIS, PLASSANS ou la FRANCE désignent la

population de Paris, de Plassans ou de la France. Et l'on pourrait soutenir que Paris, élevé à la dignité de mythe, est devenu un personnage à part entière.

Parmi les noms de personne, ROUGON occupe la première place, comme c'était à prévoir, suivi de MOURET mais non de MACQUART qu'on trouve beaucoup plus loin dans la liste (au 52<sup>e</sup> rang derrière LANTIER, 47<sup>e</sup>). En revanche c'est la branche des MACQUART qui s'assure le premier rang parmi les noms individuels : NANA (1033), JEAN (956), PAULINE (831), GERVAISE (791), JACQUES (674), CLAUDE (639), ÉTIENNE (575). Seul le Rougon SACCARD, héros de la *Curée* et de l'*Argent*, s'intercale dans la tête de liste (875 emplois, rang 6). Quant à la répartition selon le sexe, il semble qu'il y ait équilibre. En écartant les patronymes quand plusieurs héros les partagent, et en s'en tenant aux cinquante premiers de la liste, on relève 21 hommes mêlés à 21 femmes<sup>11</sup>.

Appliqué aux noms propres, dont la spécialisation lexicale est si accusée, le calcul de l'écart réduit n'a pas grande signification probabiliste. Les noms propres sont étroitement attachés à un thème. Quand on passe d'un roman à un autre, même s'il s'agit d'un cycle, le personnel est renouvelé comme l'administration américaine lors d'un changement de président. Ce calcul permet toutefois de mettre en relief les héros de chaque roman et de tenir compte de l'étendue de chacun. Aussi l'avons-nous appliqué à tous les noms propres qui ont plus de 200 occurrences dans le corpus (il y en a 74). En voici la liste alphabétique, avec l'indication du code utilisé pour chacun dans la courbe du graphique 7. Il s'agit là d'une carte où chaque nom a pour abscisse la valeur de l'écart réduit (quand elle est positive et significative).

---

11. On a aussi observé le même équilibre dans la *Recherche du temps perdu*, malgré la domination d'Albertine. Même équilibre aussi chez Zola parmi les noms communs de personne, du moins parmi ceux auxquels la lemmatisation laisse la marque du sexe :

femme	5421	15,2	homme	5141	-11,8
sœur	649	3,4	frère	1049	9,2
mère	2080	10,7	père	1881	-3,3
tante	437	11,6	oncle	610	16,4
filles	2831	12,4	fil	892	-1,4
madame	808	2,0	monsieur	1400	-2,8
dame	304	12,4	garçon	1048	14,3
mademoiselle	228	1,4	mari	254	1,8

On prendra garde toutefois que cette égalité des sexes représente en réalité une promotion féminine, eu égard à l'époque, et par rapport aux écrivains de son temps Zola accorde à la femme une attention particulière (tous les écarts réduits étant positifs dans le clan des femmes).

ALBINE	a	327	FRANCE	Z	276	MAXIME	Y	383
ANGELIQUE	b	242	FRANCOISE	A	340	MIETTE	Z	240
AUGUSTE	c	301	GERVAISE	B	791	MIGNON	1	212
BERTHE	d	386	GUILLAUME	C	553	MOURET	2	1091
BOCHE	e	219	HELENE	D	666	MUFFAT	3	249
BUTEAU	f	496	HENRI	E	211	NANA	4	1033
CAMILLE	g	282	HENRIETTE	F	305	OCTAVE	5	539
CAROLINE	h	263	JACQUES	G	674	PARIS	6	1507
CATHERINE	i	268	JEAN	H	956	PASCAL	7	369
CHANTEAU	J	378	JEANNE	I	453	PAULINE	8	831
CHARLES	k	230	JESUS	J	300	PIERRE	9	227
CHRIST	l	233	JOSSERAND	K	351	PLASSANS	0	451
CLAUDE	m	639	LANTIER	L	292	QUENU	*	251
CLOTINDE	n	281	LAURENT	M	477	RENEE	(	366
COUPEAU	o	439	LAZARE	N	539	ROSALIE	-	219
DENISE	p	656	LISA	O	403	ROSE	+	483
DIEU	q	529	LISE	P	282	ROUBAUD	=	251
DUVEYRIER	r	547	LORILLEUX	Q	257	ROUGON	&	1299
ETIENNE	s	241	LOUISE	R	319	SACCARD		875
FAUCHERY	t	575	MACQUART	S	276	SANDOZ	,	246
FAUJAS	u	249	MADELEINE	T	563	SERGE	:	319
FELICITE	v	459	MAHEU	U	268	SEVERINE	/	224
FLORENT	w	381	MARIE	V	239	SILVERE	£	313
FOUAN	x	461	MARTHE	W	418	THERESE	]	470
	y	273	MAURICE	X	513			

En règle générale la spécificité des noms propres est telle qu'un seul des écarts pour chaque nom est positif (très largement), les 21 autres étant négatifs (faiblement). Cette carte désigne à l'évidence les héros principaux de chaque texte, et les distribue souvent sous la forme du trio classique : une femme et deux hommes (rarement l'inverse). On a ainsi les triangles Thérèse-Camille-Laurent, Madeleine-Guillaume-Jacques, Renée-Maxime-Saccard, Lisa-Quenu-Florent, Marthe-Mouret-Faujas, Albine-Serge-Dieu, Gervaise-Coupeau-Lantier, Hélène-Henri-Jeanne, Nana-Muffat-Fauchery, Pauline-Lazare-Chanteau, Françoise-Buteau-Jean, Séverine-Roubaud-Jacques, Henriette-Maurice-Jean, Pascal-Clotilde-Félicité. La carte permet aussi de distinguer les romans collectifs des romans individuels. Dans les premiers la carte est remplie au centre et c'est le cas de la *Fortune des Rougon*, de *Pot-Bouille* ou de la *Terre*. Dans les seconds, deux ou trois noms se portent à l'extrémité faisant le vide autour d'eux et c'est le schéma qu'on observe dans *Thérèse Raquin*, *Madeleine Férat*, la *Faute de l'abbé Mouret* ou une *Page d'amour*. On remarquera enfin la timidité des noms de lieu qui ne s'aventurent jamais très loin sur la carte. PLASSANS fait quelques pas dans la *Fortune*, la *Conquête*, *Pot-Bouille* et le *Docteur Pascal* et PARIS s'avance un peu dans la *Curée*, le *Ventre*, *Son Excellence*, une *Page d'amour*, *Nana*, *Au Bonheur des dames*, *L'Œuvre*, la *Bête humaine*, *l'Argent* et la *Débâcle*. Mais PARIS dispersant ses occurrences dans un grand nombre de textes, les écarts qui le concernent sont d'autant plus limités.

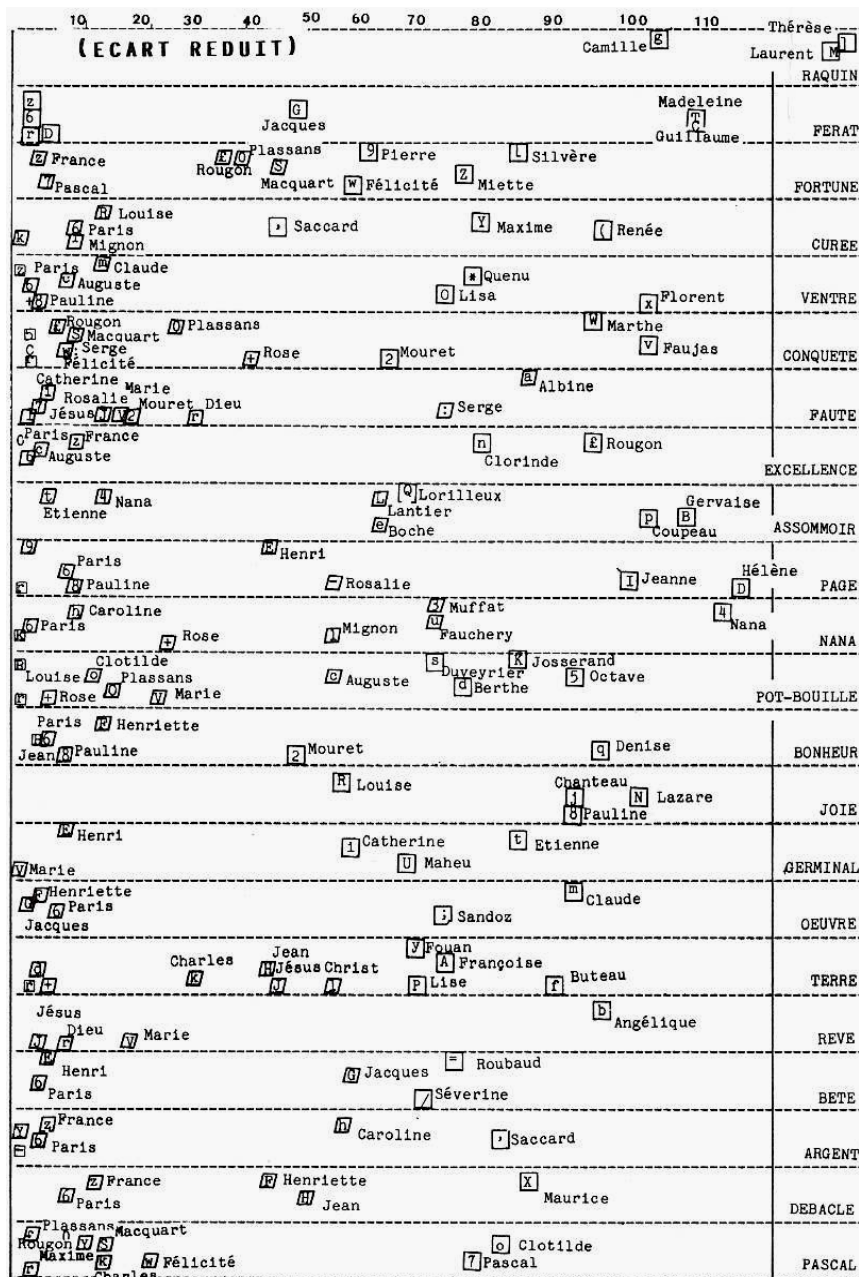


Figure 7. Noms propres spécifiques de chaque texte

Cette constatation nous amène à étudier la spécificité des noms de lieu. Ils ont beau être nettement moins fréquents et moins nombreux que les noms de personne, leur répartition est beaucoup plus régulière. Comparons en effet les tableaux 8 et 9. Le premier relève les noms propres les mieux répartis (classes de répartition 22 à 5), le second recueille au contraire ceux qui concentrent leurs emplois dans 1, 2 ou 3 textes<sup>12</sup>. Dans ce second lot on ne rencontre guère que des noms de personnages, alors que les noms de lieux foisonnent là où la répartition est la mieux assurée. Ainsi relève-t-on dans le tableau 8 le nom de PARIS ( $R = 22$ ), celui de la FRANCE ( $R = 20$ ), de PLASSANS 16, de MARSEILLE 15, de L'ITALIE 13, de LYON 13, L'ANGLETERRE 12, L'AMERIQUE 11, L'EUROPE 11, BATIGNOLLES 11, MONTMARTRE 11, la PROVENCE 11, VERSAILLES 11, BOULOGNE 10, la CHINE 10, L'ESPAGNE 10, LONDRES 10, ORLEANS 10, ROME 10. Sur les 43 noms propres présents dans au moins 10 textes, la moitié sont des noms géographiques<sup>13</sup>. On s'explique assez facilement la régularité relative des noms de lieu, qui sont généralement des noms réels, même si Zola les déguise parfois sous un nom emprunté<sup>14</sup>. Et comme Zola est peu soucieux de dépaysement et ne franchit pas les frontières, le même cadre géographique se répète d'un roman à l'autre, surtout le cadre parisien où Zola situe le plus souvent son intrigue. Au contraire l'invention règne sans partage parmi les noms de personnages, Zola déguisant même les personnages réels du monde politique, Napoléon III mis à part<sup>15</sup>. Le choix dès lors s'étend à l'infini et la répartition se réduit dans les mêmes proportions.

Les noms de lieu permettent en outre de construire l'espace culturel d'un écrivain. Nous avons eu l'occasion d'étudier le système référentiel du temps et de l'espace dans l'œuvre de Chateaubriand et d'observer chez cet écrivain la profondeur et l'étendue de sa culture historique et de sa connaissance du globe<sup>16</sup>. Combien limité en comparaison apparaît le cadre spatio-temporel de Zola !

---

12. On s'est contenté des noms les plus fréquents ( $f > 50$ ) car les listes de répartition 1, 2 et 3 contiennent des centaines d'unités.

13. On prendra garde que certains noms de saints sont en réalité des noms de rue, de monument, de quartier ou de localité, par exemple (Saint) GERMAIN, (l'île Saint) LOUIS, (la) MADELEINE, (Saint) HONORE.

14. Comme Aix sous le nom de PLASSANS.

15. C'est dans la *Débâcle* qu'il se rapproche le plus de la réalité historique et qu'il cite le plus grand nombre de personnages réels de THIERS à MAC-MAHON.

16. « Les noms propres dans l'œuvre de Chateaubriand », *Annales de la Faculté des Lettres de Nice* n°28, les Belles Lettres, 1979, p. 83-94 (1979b).



Tableau 9. Les noms propres à forte concentration

1			2			3					
JOSSERAND	351	1	DUBUCHE	83	1	COUPEAU	656	2	NANA	1033	3
LORILLEUX	257	1	PROSPER	83	1	DENISE	529	2	FLORENT	461	3
ROUBAUD	251	1	RAMOND	83	1	MAURICE	513	2	MARTHE	418	3
FAUCHERY	249	1	LOUBET	82	1	BUTEAU	496	2	RENEE	366	3
SANDOZ	246	1	LAPOULLE	82	1	FAUJAS	459	2	SERGE	319	3
DUVEYRIER	241	1	GAGNIERE	82	1	CHANTEAU	378	2	LISE	282	3
KAHN	186	1	GUEULIN	80	1	ALBINE	327	2	QUENU	251	3
MARTINE	186	1	NEGREL	80	1	SILVERE	313	2	MIGNON	212	3
DELAHERCHE	181	1	LENGAIGNE	79	1	CAMILLE	282	2	GENEVIEVE	162	3
RAQUIN	175	1	DELANGRE	77	1	CLORINDE	281	2	SEDAN	155	3
DEBERLE	173	1	CABUCHE	77	1	FOUAN	273	2	AURELIE	149	3
TRUBLOT	169	1	BEJUI	77	1	MAHEU	268	2	BAUDU	147	3
TEUSE	165	1	PALOCQUE	76	1	MUFFAT	249	2	CLEMENCE	131	3
RASTOIL	162	1	SARRIETTE	76	1	ANGELIQUE	242	2	RAMBAUD	124	3
CAMPARDON	162	1	NOIRAUDE	75	1	MIETTE	240	2	CLARISSE	119	3
FAGEROLLES	160	1	MALIGNON	74	1	SEVERINE	224	2	FELICISSE	118	3
DELHOMME	158	1	BONNEMORT	74	1	BOCHE	219	2	FANNY	93	3
BORDENAVE	157	1	CADINE	73	1	VERONIQUE	185	2	LUCIEN	82	3
CHAVAL	156	1	ARCHANGIAS	72	1	MAHEUDE	180	2	LUCIE	79	3
GAVARD	155	1	ROCHAS	72	1	ZOE	164	2	DIDE	81	3
CONDAMIN	152	1	JANTROU	71	1	MONTSOU	160	2	ESTELLE	73	3
LEVAQUE	152	1	JUZEUR	70	1	LERAT	149	2	MATHEU	72	3
HENNEBEAU	148	1	MICHAUD	70	1	VANDEUVRES	146	2	CECILE	71	3
DELESTANG	148	1	BOURRETTE	70	1	VIRGINIE	145	2	BEAUMONT	61	3
GOUJET	144	1	ROUDIER	70	1	HUTIN	136	2	SUZANNE	59	3
POIZAT	139	1	DELOCHE	69	1	JULIETTE	135	2	VICTOIRE	55	3
SATIN	129	1	DAIGREMONT	67	1	FALOISE	130	2	BISMARCK	51	3
BOURDONCLE	127	1	BOURRAS	67	1	STEINER	122	2			
JEANLIN	124	1	LECOEUR	67	1	ROGNES	120	2			
BOUCHARD	122	1	LEBIGRE	66	1	ZACHARIE	112	2			
BACHELARD	122	1	HURET	66	1	THEOPHILE	98	2	BERTHE	386	4
CORREUR	122	1	BONGRAND	66	1	CHOUTEAU	98	2	ROSALIE	219	4
HUBERTINE	120	1	CAZENOVE	65	1	PARADOU	95	2	CHRISTINE	185	4
NORMANDE	120	1	NENESSE	65	1	BONNEVILLE	87	2	SATURNIN	135	4
HUBERT	117	1	PORQUIER	65	1	MEUSE	86	2	ARISTIDE	124	4
LABORDETTE	117	1	PEPE	64	1	HORTENSE	86	2	FLORE	122	4
FONTAN	117	1	MES BOTTES	64	1	VABRE	82	2	OLYMPE	108	4
TROUCHE	113	1	GRIVET	63	1	HAMELIN	80	2	LAURE	93	4
ROBINEAU	112	1	GILBERTE	63	1	CLARA	80	2	SIDONIE	89	4
GRANDE	112	1	CLOYES	62	1	PHILOMENE	78	2	SICARDOT	61	4
CHARBONNEL	112	1	DENIZET	62	1	JOUVE	75	2	CAEN	51	4
GILQUIN	111	1	ESCORAILLES	62	1	AUGUSTINE	75	2			
JORY	110	1	PIERRON	62	1	SIMONNE	70	2			
MAHOUDEAU	110	1	VUILLET	62	1	HEDOUIN	69	2			
SAGET	110	1	PEQUEUR	61	1	LARSONNEAU	68	2			
RASSENEUR	109	1	MEHUDIN	61	1	BEAUCE	68	2			
VALERIE	109	1	HUGON	61	1	ADELAIDE	61	2			
BEU	109	1	LISON	61	1	ARTAUD	57	2			
HOURDEQUIN	108	1	MUCHE	60	1	MECHAIN	57	2			
SILVINE	106	1	GASPARINE	60	1	ZEPHYRIN	56	2			
WEISS	105	1	VALLAGNOSC	60	1	GRADELLE	55	2			
DESFORGES	104	1	TROUILLE	60	1	JACQUELINE	52	2			
GRANOUX	100	1	MOUQUETTE	59	1						
VOREUX	100	1	MAFFRE	58	1						
GUNDERMANN	97	1	BOUTHEMONT	58	1						
MACQUERON	97	1	REMILLY	58	1						
MARJOLIN	94	1	BOURDEU	57	1						
MARSY	93	1	PRIMAT	57	1						
BUSCH	93	1	GAGA	57	1						
MARTY	92	1	FENIL	56	1						
MAZAUD	92	1	LYDIE	56	1						
LUCY	91	1	REQUILLART	55	1						
POISSON	90	1	SOUVARINE	55	1						
PECQUEUX	90	1	PACHE	54	1						
GREGOIRE	90	1	FETU	54	1						
FAVIER	89	1	GRANDMORIN	54	1						
DAGUENET	89	1	COLOMBAN	54	1						
BAZELLES	89	1	BEAUDOIN	53	1						
ROUQUETTE	89	1	PLOUGUERN	53	1						
FOUCHARD	88	1	SOULEIADE	52	1						
MISARD	87	1	VETEUILL	51	1						
GOURD	86	1	SAULAIES	51	1						
BOVES	85	1	JEANBERNAT	51	1						

La géographie de Zola s'aventure rarement au-delà de l'Europe. Si on limite l'enquête aux seules unités qui ont au moins 10 occurrences voici comment apparaissent les continents au lecteur des *Rougon-Macquart* : l'Afrique passe presque inaperçue avec 37 mentions de l'AFRIQUE elle-même et 14 de l'ALGERIE<sup>17</sup>. L'AMERIQUE est à peine mieux lotie avec 25 occurrences à quoi s'ajoutent 23 citations de CAYENNE et 17 du MEXIQUE<sup>18</sup>. L'Asie est plus souvent citée, sinon l'Asie lointaine (ASIE 15, CHINE 20, INDE 12, JAPON 8), du moins le Proche-Orient dont rêvent les héros de l'*Argent* : ORIENT 56, PALMYRE 39, BEYROUTH 20, JERUSALEM 25, CONSTANTINOPLE 15, SMYRNE 12, SUEZ 10<sup>19</sup>. Notons que l'actualité semble inspirer les rares escapades de Zola en dehors de l'Europe. Les références au Mexique, à l'Algérie, au Proche Orient paraissent suggérées par l'expédition de Maximilien, la colonisation africaine et le percement du canal de Suez.

L'Europe même est tronquée et orientée dans la vision de Zola, qui se souvient de son ascendance italienne. C'est l'ITALIE qui est en effet le plus souvent citée (81 emplois), après la FRANCE 276 et devant l'ALLEMAGNE 50, l'ANGLETERRE 34, la BELGIQUE 34, l'ESPAGNE 23, le LUXEMBOURG 17, l'AUTRICHE 16, la POLOGNE 15 et la RUSSIE<sup>20</sup>. Et de même la capitale italienne l'emporte sur les autres, ROME 75, BERLIN 39, LONDRES 33, BRUXELLES 13. Si l'on excepte certains lieux plus historiques que géographiques comme AUSTERLITZ 14, SADOWA 13 et SOLFERINO 16, l'Italie s'impose encore par TURIN 10, VENISE 11, FLORENCE 7, MILAN 3, alors que beaucoup de capitales sont ignorées comme Lisbonne, Varsovie, Prague<sup>21</sup>.

Reste la France, où l'équilibre là non plus n'est pas respecté, d'une part parce que Paris et ses quartiers l'emportent sur le reste du pays, d'autre part parce que les quatre points cardinaux ne pèsent pas le même poids. Si l'Est avec 600 emplois, toutes villes et régions confondues, fait pendant à l'Ouest (500 occurrences), en revanche le Nord (300 mentions, surtout dans *Germinal*) est loin de faire contrepoids au Midi où la seule ville de PLASSANS compte 451 occurrences, devant MARSEILLE 71, LYON 42, la

---

17. On relève aussi 9 occurrences du MAROC, 8 de l'ÉGYPTE et 1 de la TUNISIE.

18. Une seule mention du CANADA, une seule des ÉTATS-UNIS, et deux du BRÉSIL. Aucune de l'Argentine, du Chili, des Antilles et pas une ville citée, ni New York, ni Montréal, ni Chicago, ni Rio, ni Buenos Aires.

19. On visite aussi la SYRIE 9, la PERSE 5, la TURQUIE 2, la PALESTINE 4, le LIBAN 6.

20. Au dessous de la fréquence 10 on rencontre l'ÉCOSSE 8, la SUISSE 6, la SUEDE 6, le DANEMARK 4, la HOLLANDE 3, la NORVEGE 3, la BAVIERE 3, la GRECE 2, la HONGRIE 2.

21. On relève 6 mentions de MOSCOU, 4 de SAINT PETERSBOURG, 5 de MADRID, 5 de VIENNE, 2 de GENEVE et 1 d'ATHENES.



PROVENCE 28, NICE 21 et TOULON 13<sup>22</sup>. Quant à PARIS, c'est le nom propre le plus cité, avant même le nom des ROUGON. À ses 1507 occurrences il faut encore ajouter les mentions de la SEINE (143), des rues, des quartiers, des faubourgs (MONTMARTRE 91, BOULOGNE 40, BATIGNOLLES 47, PASSY 39, MONCEAU 40, BELLEVILLE 13, SOUFFLOT 17, SEVRES 21, RIVOLI 29, SAINT-OUEN 18, NEUILLY 12, etc.). Mais tout lecteur de Zola est sensible à la présence obsessionnelle de Paris et les chiffres sont ici assez inutiles et plus encore les commentaires<sup>23</sup>.

Les chiffres fournissent par contre des informations moins attendues sur la place des noms propres dans la phrase. On vient de voir que les noms propres n'ont pas la même répartition dans le cycle des *Rougon-Macquart*, suivant qu'il s'agit de personnes ou de lieux. Les premiers passent quand les seconds demeurent, les personnages défilant sur une toile de fond (les noms de lieu) qui ne bouge guère. Les deux catégories diffèrent également quant à leur fonction syntaxique – ce qui a des incidences sur leur position dans la phrase. Et c'est par le biais de la place des mots qu'on s'approchera de la syntaxe, puisqu'aussi bien les programmes de reconnaissance syntaxique directe ne sont guère applicables à un si grand corpus. Il a fallu d'abord comptabiliser toutes les phrases du corpus en s'appuyant sur les ponctuations fortes – ce qui a supposé la désambiguïsation préalable du point (dont les points de suspensions et les points d'abréviation ont été détachés). On a relevé ainsi 178 858 phrases, c'est-à-dire autant d'intervalles précisément mesurés, à l'intérieur desquels on a guetté l'apparition d'un nom propre. En réalité on s'est d'abord intéressé à un groupe de noms de personne choisis parmi les plus fréquents. Puis un deuxième lot constitué d'autant de noms de lieu fréquents a été soumis au programme, avant qu'un échantillon aléatoire du vocabulaire commun ne donne la contre-épreuve. Le procédé

---

22. Précisons que nos chiffres ne prennent en compte que les termes qui ont plus de 10 occurrences. Si *Germinal* campe dans le Nord, la *Débâcle* parcourt l'Est et la *Joie de vivre* s'installe à l'Ouest face à l'Atlantique. On sent que Zola choisit ses sujets de telle façon qu'il puisse faire le tour de l'hexagone. Mais les deux pôles dominants restent Paris et le Midi.

23. PARIS occupe aussi une place prépondérante dans le roman contemporain étudié par Gunnel Engwall à partir d'un corpus de 25 titres et un demi-million d'occurrences. Mais en un siècle la géographie littéraire s'est modifiée et la place faite à l'Amérique et à l'Afrique y est beaucoup plus forte. Voici le classement des lieux les plus cités par les romanciers modernes entre 1962 et 1968 : PARIS 164, FRANCE 57, ALGERIE 51, VERDUN 51, ALGER 26, AMERIQUE 22, ESPAGNE 18, ITALIE 17, ANGLETERRE 16, STRASBOURG 15, AVIGNON 14, NEW YORK 13 etc. Évidemment un tel classement est soumis aux contingences du choix des textes et à l'influence de l'actualité et principalement de la guerre d'Algérie.

de mesure consiste à faire défiler toutes les phrases et à y chercher l'un des mots retenus. En cas de succès on mesure la longueur de la phrase et la position du mot trouvé à l'intérieur de celle-ci. On convertit alors cette valeur absolue en valeur relative (on divise le nombre position par le nombre longueur). Et on procède à une sommation de cette valeur afin d'obtenir en fin de traitement une moyenne qui exprime la position d'ensemble des mots retenus dans la phrase. Si la catégorie envisagée n'a pas de place préférentielle et si le nombre d'observations est suffisant pour étouffer les variations aléatoires, on s'attend à obtenir une valeur moyenne de 0,50. Or tel n'est pas le cas des 29 noms de personnes, choisis parmi les plus fréquents, dont la liste est reproduite en haut du tableau 10 et qui se situent toujours en avant du centre de gravité de la phrase.

Ainsi les 1299 emplois de ROUGON ont pour position moyenne la valeur 0,39, c'est-à-dire la 4<sup>e</sup> place d'une phrase de 10 mots. De tous les noms étudiés c'est GERVAISE qui se porte le plus souvent en tête de phrase, avec une valeur moyenne de 0,32. Le nom qui se rapproche le plus de la distribution normale est celui de FAUJAS (0,48). On est tenté d'en tirer des conclusions sur la force, la présence de Gervaise qui s'impose comme le personnage le plus net, le plus franc et le mieux réussi de Zola, et qui s'oppose au masque un peu figé de l'abbé Faujas. On croit pouvoir faire la distinction entre les prénoms et les noms de famille, ces derniers se rapprochant du centre de la phrase quand ils désignent une entité collective<sup>24</sup>.

Plusieurs facteurs jouent certainement qui influent sur la place des noms propres. Le plus évident semble être la fonction grammaticale qui est attachée aux noms de personnes et qui statistiquement est plus souvent celle de sujet. Les emplois de vocatif renforcent la tendance des noms de personnes à s'installer en début de phrase mais dans les romans, où la part du dialogue est limitée, ce facteur semble secondaire comme on peut le déduire de l'exemple de DIEU. Ce mot intervient en effet souvent dans les invocations et les jurons. Il est pourtant le seul nom de la série à être refoulé au-delà du centre de gravité de la phrase (moyenne 0,59 pour 546 emplois).

---

24. Mais quand le nom de famille est accaparé par un personnage unique, comme BUTEAU, son statut est assimilé à celui d'un prénom. Il en est de même des surnoms ou noms empruntés, qui ne sont pas transmissibles, comme SACCARD. En de tels cas il n'y a jamais association d'un prénom et d'un nom (cette association étant hiérarchique pousse le prénom en avant et fait passer PIERRE ou EUGENE avant ROUGON).

Tableau 10. La place des noms propres dans la phrase

NOMS DE PERSONNE			MOYENNE			phrases courtes éliminées		
ROUGON	507	1299	0.3902	475	1225	0.3878		
MACQUART	120	276	0.4338	106	251	0.4219		
MOURET	420	1091	0.3845	391	1035	0.3778		
NANA	391	1033	0.3787	346	945	0.3666		
SACCARD	301	874	0.3442	286	845	0.3387		
GERVAISE	256	790	0.3236	238	750	0.3169		
OCTAVE	207	538	0.3856	190	498	0.3822		
BUTEAU	183	495	0.3689	171	474	0.3600		
LAZARE	214	537	0.3990	200	511	0.3916		
CLAUDE	210	638	0.3289	190	600	0.3166		
FAUJAS	221	458	0.4832	218	452	0.4819		
JEAN	349	955	0.3652	330	915	0.3609		
PASCAL	137	369	0.3700	130	353	0.3671		
DIEU	321	546	0.5884	308	523	0.5897		
SERGE	132	319	0.4137	116	288	0.4032		
ALBINE	134	327	0.4083	122	305	0.4006		
PIERRE	75	227	0.3320	74	223	0.3306		
COUPEAU	257	655	0.3929	242	628	0.3849		
DENISE	187	528	0.3533	174	503	0.3468		
MAURICE	170	513	0.3320	164	502	0.3275		
FLORENT	178	461	0.3865	171	444	0.3859		
PAULINE	321	831	0.3859	295	781	0.3771		
SILVRE	121	313	0.3871	115	299	0.3847		
MARTHE	143	418	0.3428	133	392	0.3402		
CLORINDE	112	281	0.3979	102	260	0.3909		
JACQUES	280	674	0.4149	266	646	0.4120		
CLOTILDE	168	439	0.3816	156	418	0.3729		
LANTIER	126	292	0.4314	122	284	0.4289		
EUGENE	45	121	0.3699	44	119	0.3728		
NOMS DE LIEU			fréquence total moyenne					
PARIS	863	1506	0.5731	856	1492	0.5736		
PLASSANS	250	451	0.5539	248	448	0.5526		
FRANCE	178	276	0.6442	176	273	0.6431		
MONTSOU	88	159	0.5560	88	158	0.5547		
SEDAN	91	154	0.5904	90	153	0.5899		
SEINE	83	143	0.5796	83	143	0.5796		
VOREUX	48	100	0.4791	48	100	0.4791		
ROUEN	56	98	0.5723	56	98	0.5723		
BAZEILLES	48	88	0.5416	47	86	0.5426		
MEUSE	53	86	0.6136	53	86	0.6136		
ITALIE	50	81	0.6145	50	81	0.6145		
ROME	43	75	0.5766	42	74	0.5742		
MARSEILLE	42	71	0.5881	41	70	0.5858		
BEAUCE	33	67	0.4939	33	67	0.4939		
VIORNE	35	62	0.5641	35	62	0.5641		
ALLEMAGNE	31	50	0.6126	30	49	0.6115		
BERLIN	29	39	0.7489	15	18	0.8362		
BOULOGNE	27	40	0.6764	27	40	0.6764		
BELGIQUE	22	34	0.6558	22	34	0.6558		
CAEN	30	51	0.5938	30	51	0.5938		
CHARTRES	27	42	0.6445	26	41	0.6408		
ESPAGNE	12	23	0.5237	11	22	0.5112		
EUROPE	30	52	0.5786	30	52	0.5786		
LILLE	23	41	0.5729	23	41	0.5729		
LYON	19	41	0.4606	19	41	0.4606		
LONDRES	21	33	0.6328	21	33	0.6328		
PALMYRE	14	39	0.3594	14	39	0.3594		
PASSY	25	39	0.6395	25	39	0.6395		
ORIENT	29	56	0.5146	29	56	0.5146		
REMILLY	29	57	0.5165	29	57	0.5165		
PROVENCE	19	28	0.6901	19	28	0.6901		
NOMS COMMUNS								
ABANDON	189	0.5836	AFIN	332	0.5967	ARGENT	1456	0.6015
ABATRE	61	0.6502	AGIT	57	0.4720	ARRIVE	136	0.5951
ABORD	1030	0.3508	AILLEURS	1371	0.2965	ART	139	0.6138
ABOMINABLE	202	0.5899	AIR	3647	0.6048	ARTICLE	112	0.5261
ABOMINATION	62	0.5933	ALLAIT	2010	0.4738	ASSEZ	902	0.5619
ACCIDENT	107	0.5183	ALORS	3696	0.1696	ASSIT	162	0.4021
ADIEU	146	0.5414	AMANT	317	0.6055	AVANT	1316	0.5153
ACTION	93	0.6469	AMOUR	722	0.5910	AVAIT	19852	0.4161
ADRESSE	139	0.6072	ANGOISSE	344	0.6143	AVOIR	3405	0.5333
AFFAIRE	754	0.5882	ANTIQUE	97	0.5651	AVAIS	590	0.4559
AFFREUX	146	0.5827	APPELAIT	244	0.4935	AUX	5431	0.5699

Dieu ne semble donc pas un personnage à part entière chez Zola, mais plutôt un être abstrait, et presque un objet, même lorsque le thème est religieux, ce qui se produit dans la *Faute* et le *Rêve*. À de tels indices peut-être pourrait-on reconnaître l'athéisme de l'auteur, mais la contre-épreuve avec des écrivains chrétiens et hantés comme Bernanos serait bien utile<sup>25</sup>.

Les noms de lieu ont un comportement tout différent et on les rencontre plus volontiers dans la deuxième moitié de la phrase, là où la construction de la langue française place de préférence les circonstances. Tous ceux de tableau 10 (partie centrale) – du moins tous ceux qui ont une fréquence suffisante pour échapper aux variations aléatoires – ont une moyenne au-delà de la médiane, aux alentours de 0,6. Une seule exception mérite attention et concerne le *VOREUX* (0,4791 pour 100 emplois). Mais le nom de la mine de *Germinal* est-il seulement un nom de lieu ? N'est-ce pas plutôt un personnage monstrueux, une divinité souterraine qui dévore les victimes humaines comme le minotaure de Cnossos ? À ce titre, il aurait donc sa place en tête de phrase parmi les actants.

Quant au vocabulaire commun, un rapide sondage portant sur une trentaine d'unités nous le montre polarisé par les catégories grammaticales. Et c'est bien ce qu'on attendait de la phrase française où la place, la nature et la fonction des mots sont en relation directe. Comment s'étonner que les adverbes de phrase (*ALORS, D'ABORD, D'AILLEURS*) choisissent les premières places (0,17, 0,35 et 0,29) et que les auxiliaires et plus généralement les verbes précèdent la masse des substantifs ? L'éventail des fonctions (et le nombre de places) est en effet moins large devant le verbe français que derrière. Il en va de la phrase française comme au théâtre : on voit toujours les mêmes têtes au premier rang. Et là les noms propres ont leur fauteuil réservé, s'il s'agit de personnages.

---

25. On a eu un instant le soupçon que les phrases très courtes, souvent exclamatives ou elliptiques, pouvaient jeter la perturbation dans les relevés. C'est pourquoi on a renouvelé les calculs en excluant les phrases qui avaient moins de 4 mots. Cela ne change rien aux résultats, ni pour *DIEU* (moyenne 0,5897 contre 0,5884 pour 523 phrases au lieu de 546), ni pour aucun autre élément de la série. Voir colonne de droite du tableau 10.