



**HAL**  
open science

## Les catégories grammaticales chez Hugo

Étienne Brunet

► **To cite this version:**

Étienne Brunet. Les catégories grammaticales chez Hugo. A. Zampolli. Computational lexicology and lexicography. Special issue dedicated to Bernard Quemada, VI, Giardini, pp.91-115, 1991, *Linguistica Computazionale*. halshs-01367636

**HAL Id: halshs-01367636**

**<https://shs.hal.science/halshs-01367636>**

Submitted on 16 Sep 2016

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

**Chapitre 11 (numérique) de : Étienne BRUNET, Tous comptes faits.**  
Écrits choisis, tome III. Questions linguistiques, *Bénédicte PINCEMIN (éd.)*,  
Paris : Éditions Champion, sous presse (publication prévue en 2016).  
Publié en ligne par la revue *Texte ! Textes & Cultures*, <http://www.revue-texto.net>  
Volume XXI – n°2 (2016). Coordonné par Audrey MOUTAT.  
Mis à disposition sous licence CC BY-NC-ND 3.0 France  
<http://creativecommons.org/licenses/by-nc-nd/3.0/fr>

## Les catégories grammaticales chez Hugo<sup>1</sup>

Il n'y a pas de CATÉGORIE GRAMMATICALE chez Hugo. Entendons par là que cette expression ne se rencontre pas sous sa plume<sup>2</sup>. Il y a certes quatre<sup>3</sup> CATÉGORIES, mais elles n'appartiennent pas à la grammaire<sup>4</sup>. On ne trouve

---

1. NDÉ : Article paru dans Zampolli, A., Cignoni, L., Peters, C. (éd.), *Computational lexicology and lexicography. Special issue dedicated to Bernard Quemada*, vol. I, collection *Linguistica Computazionale* vol. VI, Pisa : Giardini, 1991, p. 91-115 (1991a). On doit préciser qu'à cette époque les logiciels de lemmatisation comme *Cordial* ou *TreeTagger* n'existaient pas.

2. Du moins dans le corpus Hugo qui a été constitué avec les données de l'Institut national de la langue française et qui comprend deux millions d'occurrences. Voici les textes qu'il regroupe, avec les symboles utilisés dans les graphiques :

<b>Let</b> <i>Lettres à la fiancée</i>	<b>Rhi</b> <i>Le Rhin</i>	<b>Rue</b> <i>Chansons des rues...</i>
<b>Her</b> <i>Hernani</i>	<b>Co1</b> <i>Correspondance 1848</i>	<b>Co2</b> <i>Correspondance 1866</i>
<b>Aut</b> <i>Feuilles d'automne</i>	<b>Con</b> <i>Les contemplations</i>	<b>Mer</b> <i>Les travailleurs de la mer</i>
<b>Dam</b> <i>Notre-Dame de Paris</i>	<b>LS1</b> <i>Légende des siècles 1</i>	<b>Co3</b> <i>Correspondance 1873</i>
<b>Bor</b> <i>Lucrece Borgia</i>	<b>Mi1</b> <i>Les misérables 1</i>	<b>LS2</b> <i>Légende des siècles 2</i>
<b>Tud</b> <i>Marie Tudor</i>	<b>Mi2</b> <i>Les misérables 2</i>	<b>LS3</b> <i>Légende des siècles 3</i>
<b>Ruy</b> <i>Ruy Blas</i>	<b>Mi3</b> <i>Les misérables 3</i>	
<b>Ray</b> <i>Les rayons et les ombres</i>	<b>Fin</b> <i>La fin de Satan</i>	

Le roman des *Misérables* a été découpé en trois parties, vu son ampleur.

3. Au moment où apparaît, en toutes lettres, notre premier chiffre, nous devons prévenir le lecteur qu'il y en aura d'autres au cours de cette étude, qui anticipe sur un ouvrage à paraître aux éditions Slatkine : *Le vocabulaire de Victor Hugo* (1988a). Mais il y aura peu de détail sur la technique de calcul, et nous renvoyons le lecteur à notre ouvrage, ou mieux aux manuels classiques de Charles Muller. Bornons-nous à dire ici que les résultats se fondent le plus souvent sur un écart réduit, c'est-à-dire en dernier ressort sur la loi normale.

4. L'adjectif GRAMMATICAL est ignoré de notre corpus. Quant au mot GRAMMAIRE, qui a la fréquence 7, il est presque toujours associé aux souvenirs scolaires, parfois avec nostalgie :

*Oh ! que de soirs d'hiver radieux et charmants*

*Passés à raisonner langue, histoire et grammaire (Contemplations, livre 4)*

parfois avec humour, à propos des puissants : « *Tout obéit au succès, même la grammaire* », ou de Thénardier « *ruffian lettré à la grammaire près* », ou de l'éducation des couvents, qui enseignent « *l'histoire, la chose qu'on appelle ainsi au couvent, la*

pas davantage de PARTIES DU DISCOURS, même si l'on compte beaucoup de PARTIES et beaucoup de DISCOURS. Hugo s'intéresse plus au lexique ou à la versification qu'à la syntaxe. Il utilise volontiers les mots VOCABULAIRE et VOCABLE, sans parler des mots VERS OU RIME<sup>5</sup>, et met en œuvre avec délices le vocabulaire maritime, « la vieille langue de mer »<sup>6</sup>, ou l'argot du baigneur et des bas-fonds<sup>7</sup>. Même les mots étrangers ne le rebutent pas, et s'il s'agit de déshabiller un écossais, Hugo détaille avec complaisance toutes les pièces, nommément désignées, de l'accoutrement<sup>8</sup>. Même un produit aussi banal que la bière peut provoquer l'intempérance lexicale de Hugo et déclencher une logorrhée comique : « ce breuvage que les norvégiens appellent *wel*, les goths *buska*, les carinthiens *vo*, les esclavons *oll*, les dalmates *bieu*, les hongrois *ser*, les bohèmes *piva*, les polonais *pwo*, et que nous nommons la *bière* »<sup>9</sup>.

Par contre Hugo ne manifeste à l'égard de la syntaxe et de la grammaire qu'un sentiment révérencieux et silencieux. Et là-dessus, dans *Réponse à un acte d'accusation*, il a dit une fois pour toutes sa doctrine : *Guerre à la rhétorique et paix à la syntaxe*. Il respecte la syntaxe, d'une part en appliquant ses règles, d'autre part en les acceptant sans mot dire, sans les discuter, sans même en faire mention. Presque tous les mots qui servent à désigner quelque objet de la grammaire sont exclus de son vocabulaire et l'on cherchera vainement dans le corpus, et même dans sa

---

*géographie, la grammaire, les participes, les rois de France, un peu de musique, à faire un nez... »*. Tendresse et ironie se trouvent parfois mêlées, à l'endroit d'un homme de l'Ancien Régime, qu'il a connu jadis :

*Vous me faisiez parfois réciter ma grammaire. (Contemplations, Écrit en 1846)*  
ou à l'endroit d'un grammairien des temps antiques, le « *grammairien Philétas, lequel était si petit et si menu qu'il était obligé de plomber ses souliers pour n'être pas emporté par le vent.* », *Les Misérables, Marius*.

5. La RIME (verbe ou substantif) est omniprésente sous la plume de Hugo (37 emplois), non seulement dans la correspondance, mais aussi dans la poésie, et particulièrement dans les *Chansons*. Quant au mot VERS, la difficulté a été de séparer les prépositions et les substantifs, et, parmi ces derniers, les VERS LUISANTS, et ceux dont l'éclat n'est que poétique (32 occurrences contre près de 400).

6. Plusieurs pages des *Travailleurs de la mer* sont consacrées à des considérations linguistiques sur cette langue, p. 654-657. Et l'éditeur de la Pléiade croit utile de livrer un glossaire des mots techniques employés dans ce roman, p. 1736.

7. Tout un chapitre est consacré à l'argot dans les *Misérables*. Voir en particulier p. 996 et 1002.

8. Il lui arrive de se tromper et d'écrire BUG-PIPE pour BAG-PIPE (*Les Travailleurs de la mer*, p. 632 et note p. 1377).

9. *Le Rhin*, p. 216.

correspondance, une PRÉPOSITION, un DÉMONSTRATIF, un POSSESSIF, une COORDINATION, un INDICATIF, un IMPÉRATIF, un CONDITIONNEL ou un INFINITIF. Quand les mots techniques sont cités, c'est souvent une seule fois, généralement dans sa *Réponse à un acte d'accusation* : GRAMMAIRIEN, ADVERBE, PRONOM, LINGUISTE, NÉOLOGISME, VOCATIF, SUBJONCTIF<sup>10</sup>. Pas un mot de l'ADJECTIF, même pas pour répondre au Musset de *Dupuis et Cotonet*. Deux mentions seulement du SUBSTANTIF et deux de la SYNTAXE<sup>11</sup>. Si le mot VERBE est fort sollicité (fréquence 48), c'est dans le sens latin, comme équivalent du langage ou de la parole, et trois fois seulement dans une acception grammaticale<sup>12</sup>. C'est aussi le sens commun qui prévaut dans l'emploi du NOM (1274 occurrences) ou de l'ARTICLE (279 occurrences).

### 1. Comparaison externe

Mais même si l'artisan ne travaille pas en devanture, en expliquant ce qu'il fait, il n'en résulte pas qu'il n'a pas conscience de son art, encore moins qu'il travaille sans outils. Les parties du discours sont ces outils dont les écrivains se servent aussi naturellement que Monsieur Jourdain faisant de la prose. Mais l'usage et le dosage qu'ils en font diffèrent de l'un à l'autre. On a pu observer que la plume de Chateaubriand était favorable à la catégorie nominale (adjectifs et substantifs), que celle de Zola cultivait plutôt les participes et les verbes, que celle de Proust accumulait les mots de relation, et qu'enfin celle de Giraudoux fuyait les adjectifs, en recherchant les noms et les verbes. D'aucuns ont pu penser que Hugo souscrirait au choix de Chateaubriand et nous l'avons cru avant d'entreprendre cette recherche<sup>13</sup>. En réalité le dosage des catégories chez Hugo évoquerait plutôt les caractéristiques de Giraudoux : au moins le goût pour le substantif et le verbe, sinon l'aversion pour l'adjectif. On trouvera dans les tableaux 1 et 2a (et dans le graphique 2b qui accompagne le tableau 2a) le profil grammatical de Hugo.

10. « Je voudrais que dans ces pages splendides et charmantes, vous *pourchassiez* (tyrannie de l'imparfait du subjonctif) quatre mots. », *Correspondance (1849-1866)*, p. 429.

11. Il est inutile de chercher sous la plume de Hugo des termes plus savants encore, ou plus récents, comme LINGUISTIQUE, STYLISTIQUE, SÉMANTIQUE, MORPHOLOGIE, PHILOGOLOGIE, MÉTRIQUE.

12. Un exemple est très connu : « Tout conjugue le verbe *aimer*. » (*Contemplations*). Les autres sont dans *Réponse à un acte d'accusation*.

13. Voici l'aveu de notre imprudence : « On imagine que le schéma adopté par Chateaubriand est celui qui convient le mieux à l'écriture hugolienne, mais gageons qu'un tel pari est risqué ». *Le Vocabulaire de Zola*, tome 1, p. 151 (1985a).

Tableau 1. La distribution des parties du discours. Effectifs

	substant	adj-part	adj.subs	homogr	verbes	-ment	adverbes	prépos	gramm	total
Letres	13274	7054	493	25	15542	559	4880	10920	41143	93890
	1276	1158	100	9	796	148	64	31	149	3731
Hernani	4755	1804	220	9	3537	49	878	2619	8962	22833
	1001	635	90	5	492	23	61	27	142	2476
Feuilles	5978	2305	200	15	3036	33	931	3427	8941	24866
	1269	808	96	6	582	15	63	29	133	3001
Notre-Dame	37535	16420	2055	60	23384	1006	6572	26097	63386	176515
	4104	2680	353	21	1443	275	82	34	176	9168
Borgia	3715	1494	154	3	3317	78	887	2626	8445	20719
	874	561	71	3	441	56	55	28	141	2230
Tudor	4345	1740	197	6	4058	76	1076	2703	10746	24947
	850	549	68	2	439	46	55	26	144	2179
Ruy Blas	6122	2781	196	14	4374	143	1114	3893	11231	29868
	1365	911	91	8	610	74	69	29	149	3306
Rayons	7021	3111	278	16	3287	67	960	3889	9968	28597
	1494	945	115	10	566	42	59	31	122	3384
Rhin	43754	21169	2653	107	22746	1116	6664	32592	68998	199799
	4596	2891	414	20	1449	299	78	34	168	9949
Corresp 1	19698	10553	750	28	17388	684	5552	15968	46781	117402
	2347	1602	190	12	925	185	73	33	160	5527
Contempl	20710	8919	760	55	11823	105	2349	10912	32194	87827
	2511	1512	205	17	940	43	68	31	148	5475
Légende 1	17517	8005	755	38	9980	117	2082	10107	26219	74820
	2782	1556	234	13	996	58	69	30	154	5892
Misérabl 1	36245	15835	1777	117	26905	1054	6834	25361	66544	180672
	4203	2560	340	22	1421	268	79	31	167	9091
Misérabl 2	38212	16989	1911	78	27443	1021	6939	26492	68639	187724
	4745	2748	381	19	1549	275	79	31	178	10005
Misérabl 3	33418	15062	1501	66	24747	859	6301	23331	60584	165869
	4238	2605	329	21	1439	254	80	32	185	9183
Fin de S.	11248	4755	454	15	6948	56	1261	6229	17271	48237
	1874	1119	143	7	780	20	59	31	141	4174
Chansons	7745	3067	308	17	4419	37	786	4243	10958	31580
	2155	920	145	10	724	28	62	28	141	4213
Corresp 2	32370	16411	1566	74	27887	949	7335	25409	74021	186022
	3069	1926	274	15	1188	210	76	31	172	6961
Trav. Mer	29693	11985	1413	56	18908	626	5000	20136	46318	134135
	4261	2408	301	15	1357	202	78	31	171	8824
Corresp 3	19308	10164	833	27	16497	551	4040	14927	42567	108914
	2401	1500	208	11	926	168	73	30	166	5483
Légende 2	21445	9150	958	53	12943	136	2606	11871	33601	92763
	2879	1563	243	11	1049	53	70	31	159	6058
Légende 3	8320	3468	332	18	5127	58	1000	4571	13393	36287
	2018	1012	133	9	745	37	68	31	136	4189
poésie	99984	42780	4045	227	57563	609	11975	55249	152545	424977
	5089	2961	469	28	1681	123	78	32	178	10639
théâtre vers	10877	4585	416	23	7911	192	1992	6512	20193	52701
	1741	1183	138	9	776	85	74	29	164	4199
théâtre prose	8060	3234	351	9	7375	154	1963	5329	19191	45666
	1282	850	105	4	634	81	60	28	155	3199
romans	175103	76291	8657	377	121387	4566	31646	121417	305471	844915
	8251	4894	727	36	2457	512	99	34	206	17216
correspond	128404	65351	6295	261	100060	3859	28471	99816	273510	706027
	6213	4003	595	24	1954	453	85	35	183	13545
total	422428	192244	19764	897	294296	9380	76047	288323	770907	2074286
	9824	5993	950	41	2840	601	105	36	212	20602

Précisons d'abord qu'il n'est pas très aisé d'établir les effectifs, et surtout de maintenir la correspondance des normes pour les corpus qu'on

met en regard. En dehors d'une véritable désambiguïsation, il faut décider que la forme *CHARME* est un verbe ou un substantif, ou que *VERS* est nom ou préposition. On reprend ici la complainte habituelle aux lemmatiseurs, mais sur un ton plus angoissé. Car les catégories grammaticales sont concurrentes et tout choix qui décide en faveur de l'une est fait au détriment d'une autre. Quand un mot était ambigu, on n'a pas cherché à établir un partage, à partir de pourcentages douteux<sup>14</sup>. On a versé toutes les occurrences du côté de la catégorie présumée majoritaire.

On ne se cache pas que cette procédure est expéditive, grossière et arbitraire. Mais la statistique suscite et pardonne ces fautes vénielles. Le seul péché mortel est l'inconstance et l'incohérence. La statistique étant essentiellement une méthode comparative exige impérieusement que les termes de la comparaison soient traités sur le même pied. C'est pourquoi le tri des catégories doit être exécuté parallèlement sur les corpus que l'on confronte<sup>15</sup>. Ce découpage auquel nous avons procédé plusieurs fois dans le passé, notamment à propos de Proust et de Zola, a été remis sur le chantier, avec le souci de réduire encore l'incertitude, en isolant les catégories faciles à reconnaître (comme celle des adverbes en *-MENT* et celle des prépositions) et en limitant au maximum les cases fourre-tout où avaient d'abord été rangés les homographes. L'une de ces cases a disparu, celle des homographes profonds, qui ont été répartis dans la population normale (colonne 4 du tableau 1). L'autre cas est moins grave, parce qu'il est interne à la catégorie nominale ; il concerne les adjectifs qui peuvent être substantivés. Ici (colonne 3 du même tableau) les retouches ont été rares<sup>16</sup>.

Le tableau 2 indique le choix de Hugo parmi les parties du discours. Comme on redoutait l'influence de l'époque et du genre, la comparaison a été établie à quatre niveaux, d'abord avec le corpus total du TLF, puis avec le sous-ensemble de l'époque de Hugo, de 1815 à 1885, tous genres réunis, puis en distinguant la prose et les vers dans le corpus de Hugo et dans celui du temps. Or les quatre comparaisons amènent aux mêmes conclusions.

---

14. Les pourcentages de ce type ont été établis par le TLF, à partir de l'examen de plus de 300 000 occurrences. À supposer que l'échantillon est fidèle à la population dont il est issu, il resterait à valider l'application de tels pourcentages à un autre corpus.

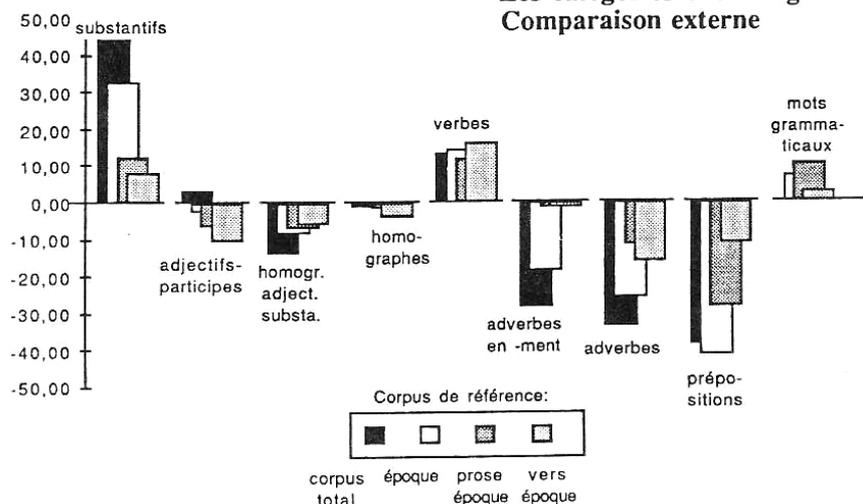
15. Quand un corpus est compris dans un autre, le parallélisme ne doit pas pousser à traiter seulement les mots du plus petit, que le plus vaste partage nécessairement. Il faut traiter aussi les mots extérieurs, afin de respecter le dosage réel des catégories.

16. Cette catégorie a été intégrée aux adjectifs dans un traitement ultérieur.

Tableau 2a. Les parties du discours. Comparaison externe

effectifs observés		substant	adj-part	adj.subs	homogr	verbes	-nient	adverbes	prépos	gramm
corpus entier		13378715	6466802	736800	31467	9726934	425007	2908468	10482713	26113503
époque		6143065	2953342	320425	14053	4380330	172801	1273195	4745946	11680829
corpus prose		5294311	2555602	280760	11925	3877439	151954	1128569	4165998	10247804
corpus vers		415311	188861	18431	1209	236971	3222	60239	245690	657275
prose Hugo		311567	144876	15303	647	228822	8579	62080	226562	598172
vers Hugo		110861	47365	4461	250	65474	801	13967	61761	172738
effectifs théoriques Hugo										
corpus entier		394921	190891	21749	929	287126	12546	85854	309435	770835
époque		402174	193349	20978	920	286771	11313	83353	310707	764720
prose époq		305002	147227	16174	687	223377	8754	65016	240001	590370
vers époq		108573	49373	4818	316	61950	842	15748	64229	171828
écarts réduits										
corpus entier		44,43	3,14	-13,67	-1,06	13,58	-28,69	-33,98	-38,53	0,08
époque		33,04	-2,60	-8,67	-0,79	14,54	-18,80	-26,18	-41,54	7,32
prose époq		12,25	-6,31	-7,06	-1,57	11,87	-1,93	-11,86	-28,26	10,46
vers époq		8,08	-10,52	-5,99	-4,32	16,47	-1,66	-16,51	-11,33	2,55

### Les catégories chez Hugo. Comparaison externe



Graphique 2b. Les parties du discours. Comparaison externe

Les écarts sont toujours de même signe pour la même catégorie, sauf dans un cas seulement : les adjectifs (ou participes) sont en excédent si l'on envisage le corpus entier, alors que l'usage qu'en fait Hugo est déficitaire si l'on s'en tient au XIX<sup>e</sup> siècle, où précisément l'adjectif était fort en faveur. Hugo apprécie plus que tout le substantif (là aussi l'usage s'étant modifié au cours de deux siècles, l'excédent est plus considérable si l'on considère le corpus entier). Il fait aussi bon accueil au verbe, sans doute parce que le théâtre et dans une moindre mesure le roman

s'appuient sur la catégorie du verbe. Mais ce n'est pas une explication suffisante, puisque l'excédent persiste dans le corpus en vers, d'où le roman est exclu et où le théâtre a peu de part. Hugo accepte enfin le dosage normal des mots grammaticaux. Dans la zone négative on trouve essentiellement les adverbes de toute espèce et les prépositions. Hugo a peu de goût pour les adverbes en *-MENT*, qu'il appelle des adverbes qualificatifs<sup>17</sup> et dont la lourdeur est embarrassante dans le cadre du vers. Mais cet encombrement est ressenti par les autres poètes et quand le vers de Hugo est confronté à celui de l'époque, l'aversion de Hugo n'a plus rien de spécifique (écart réduit non significatif de -1,66). Il en est ainsi de la prose, où la faveur de l'adverbe en *-MENT* est variable selon qu'il s'agit de l'époque contemporaine ou du temps de Hugo, et surtout selon que la prose est littéraire ou « technique ». Si la comparaison porte sur l'époque de Hugo, qui est moins favorable à cette catégorie, et sur les seuls textes littéraires, à l'exclusion des essais, l'écart observé chez Hugo atteint à peine le seuil significatif ( $z = -1,93$ ). En ce qui concerne les autres adverbes, le déficit est constant, quels que soient les ensembles comparés. Quant aux prépositions, le déficit qui les frappe s'oppose violemment à l'excédent observé chez Chateaubriand.

Restent les deux catégories où intervient l'adjectif et qui sont déficitaires. Mais elles sont ambiguës et l'on ne sait si le déficit est imputable aux adjectifs purs ou aux éléments qui s'y trouvent mêlés : substantifs dans un cas, participes dans l'autre. On a certes des présomptions dans la situation des adjectifs substantivés : comme les substantifs purs sont largement excédentaires, il est difficile d'attribuer à la substantivation la désaffection qui frappe les adjectifs substantivés (ou en état de l'être). Mais les adjectifs qui peuvent changer de statut sont peut-être d'une espèce particulière qui les fait aptes à désigner une personne (c'est le cas de beaucoup de suffixes en *-IEN* ou en *-ISTE*) ou un concept (certains composés en *-IQUE*). Quant à la confusion des adjectifs et des participes, elle laisse le jugement en suspens. Et pour y voir clair, un examen plus approfondi de ce cas d'espèce s'impose, et nous l'entreprendrons plus loin. Mais dès maintenant on peut douter que l'essentiel du romantisme – à tout le moins celui de Hugo – se réduise à l'emploi de l'adjectif, comme le suggérait l'impertinence de Musset. Le romantisme hugolien paraît préférer les catégories pleines, celles où la substance est plus solide, c'est-à-dire le substantif et, à un moindre degré,

---

17. « J'ai peur qu'on ne rattrape maintenant à mes dépens tout le temps qu'on a si ... (mettez un adverbe qualificatif perdu », *Correspondance (1849-1866)*, p. 234.

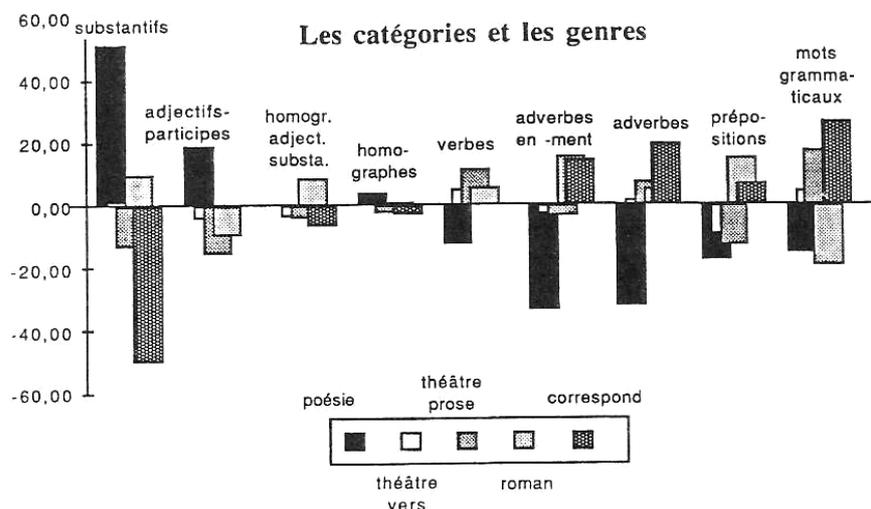
le verbe. Il est difficile de mesurer les effets produits par un tel dosage des parties du discours. Peut-être pourrait-on attribuer la plénitude, la nervosité et l'efficacité de l'écriture hugolienne au choix des ingrédients les plus consistants. En tous cas cette tendance du tempérament propre de l'écrivain n'est guère perturbée par les influences du temps et de l'époque. La neutralisation successive des facteurs suspectés laisse intacts les mêmes traits, superposés dans le graphique 2b.

## 2. Étude interne

Les difficultés d'harmonisation disparaissent quand on s'enferme dans le même corpus. Celui de Hugo est assez vaste et assez varié pour qu'on ne se sente pas à l'étroit et que des jeux d'opposition ou d'alliance puissent commander la distribution des parties du discours entre les textes et entre les genres. Les effectifs du tableau 1 ont été convertis en écarts réduits (tableau 3a), et ceux-ci en représentation graphique dans le graphique 3b.

Tableau 3a. La distribution interne des parties du discours (écarts réduits)

écarts réduits									
	substant	adj-part	adj.subs	homogr	verbes	-ment	adverbes	prépos	gramm
Letres	-43,27	-18,08	-13,74	-2,51	19,69	6,68	25,08	-19,09	34,24
Hernani	1,55	-6,82	0,17	-0,28	5,26	-5,37	1,42	-9,90	5,20
Feuilles	12,92	0,01	-2,41	1,30	-8,33	-7,54	0,65	-0,50	-3,14
Notre-Dame	8,76	0,50	9,51	-1,95	-10,96	7,69	1,31	10,42	-9,04
Borgia	-7,80	-9,78	-3,11	-2,00	7,00	-1,63	4,65	-4,76	8,53
Tudor	-10,38	-11,97	-2,66	-1,47	8,77	-3,49	5,37	-13,06	15,41
Ruy Blas	0,51	0,25	-5,29	0,30	2,11	0,69	0,58	-4,04	1,25
Rayons	15,80	9,01	0,34	1,04	-12,18	-5,52	-2,75	-1,37	-6,45
Rhin	15,98	20,50	18,07	2,33	-35,00	7,44	-8,12	30,43	-20,29
Corresp 1	-28,04	-3,24	-11,35	-3,29	5,83	6,84	19,58	-2,83	15,52
Contempl	21,58	8,83	-2,71	2,82	-5,84	-14,98	-15,68	-11,98	-2,53
Légende 1	18,81	13,10	1,61	1,01	-6,28	-12,26	-12,86	-2,93	-9,70
Misérabl 1	-2,99	-7,36	1,40	4,60	8,31	8,68	2,70	1,64	-2,43
Misérabl 2	-0,10	-3,25	3,03	-0,37	5,20	6,19	0,72	2,59	-4,48
Misérabl 3	-2,05	-2,61	-2,08	-0,71	8,25	4,15	2,94	1,89	-4,46
Fin de S.	14,54	4,30	-0,26	-1,30	1,27	-11,11	-12,21	-5,88	-4,96
Chansons	16,51	2,61	0,41	0,91	-0,93	-8,92	-11,01	-2,23	-7,24
Corresp 2	-29,69	-6,62	-5,14	-0,75	9,64	3,90	6,54	-2,92	19,48
Trav. Mer	14,87	-4,14	3,90	-0,27	-0,92	0,82	1,21	11,29	-16,36
Corresp 3	-19,81	0,71	-6,53	-3,01	8,63	2,71	0,76	-1,77	10,67
Légende 2	19,01	6,10	2,55	2,08	-1,94	-14,16	-13,95	-9,22	-4,82
Légende 3	10,92	1,83	-0,75	0,59	-0,30	-8,36	-9,14	-6,72	-0,81
	substant	adj-part	adj.subs	homogr	verbes	-ment	adverbes	prépos	gramm
poésie	51,22	19,17	-0,07	3,58	-12,48	-33,58	-32,39	-17,64	-15,23
théâtre vers	1,41	-4,34	-3,89	0,04	5,08	-3,04	1,38	-9,63	4,39
théâtre prose	-13,00	-15,52	-4,08	-2,45	11,26	-3,69	7,14	-12,93	17,23
romans	9,51	-9,36	8,78	0,79	5,67	15,66	4,94	15,07	-19,80
correspond	-49,94	-0,40	-6,49	-3,12	-0,43	14,52	19,80	6,60	26,72

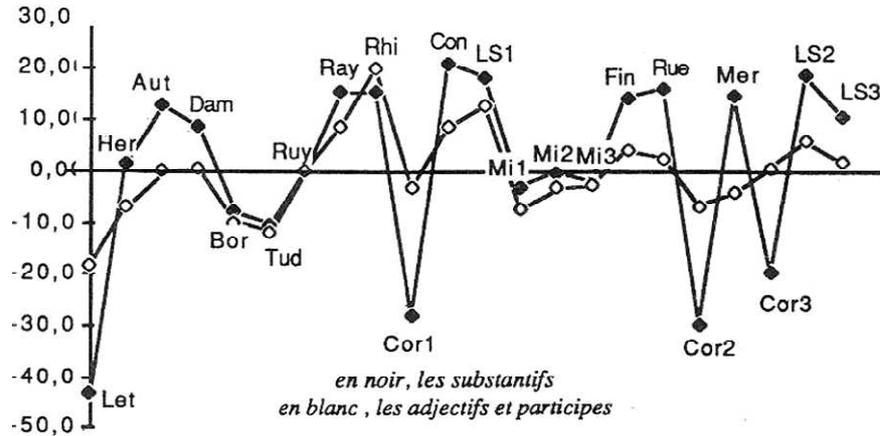


**Graphique 3b. La distribution interne des parties du discours (écarts réduits)**

On s'intéressera d'abord aux genres dans la moitié inférieure de ce tableau. La poésie (en noir sur le graphique) accapare les catégories nominales et délaisse toutes les autres espèces. La valeur élevée de l'écart réduit montre que les choix poétiques sont extrêmement nets : substantifs +51, adjectifs + 19, verbes -12, adverbes en -MENT -33, autre adverbes -32, prépositions -17, mots-outils -15. Les options de la correspondance (en gris foncé sur le graphique) sont tout à l'opposé : déficit de substantifs (-49) et excédent des adverbes et des mots grammaticaux. Le théâtre se prononce pour le verbe et les mots-outils. Quant au roman, ses goûts sont éclectiques puisque les substantifs lui plaisent, souvent introduits par des prépositions, mais tout aussi bien les verbes, accompagnés des adverbes. Dans l'opposition forte qui dresse la poésie contre le genre épistolaire, théâtre et roman offrent deux compromis différents qui pour s'accorder auraient besoin d'un autre compromis.

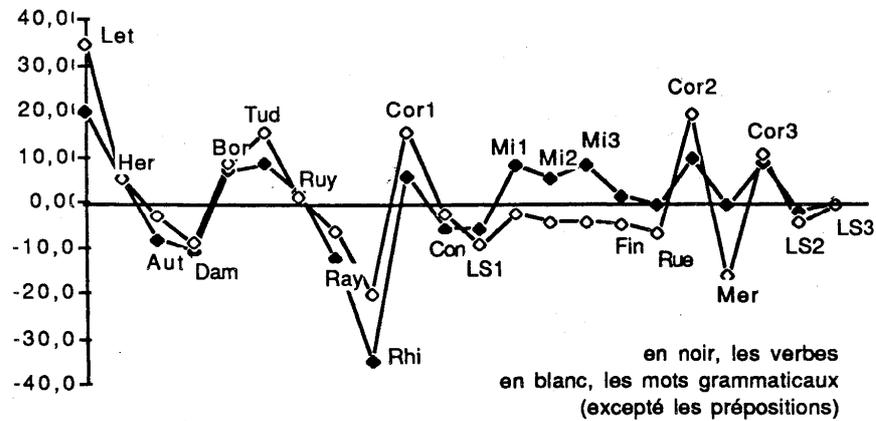
C'est l'analyse factorielle qui mettra tout le monde d'accord, Mais avant d'y parvenir, observons le détail des textes, en feignant d'ignorer à quels genres ils appartiennent. Le graphique 4 nous rappelle immédiatement l'existence de ces genres. Les catégories nominales, auxquelles il est consacré et dont le parallélisme éclate aux yeux, rejoignent en effet les textes poétiques au haut du graphique et s'éloignent des textes de correspondance et du théâtre en prose (rejetés en bas), Si les substantifs n'écartent pas le roman et le théâtre en vers, par contre l'adjectif n'admet que le voisinage des textes poétiques. L'adjectif

serait donc l'apanage exclusif de la poésie, s'il n'y avait l'exception remarquable du *Rhin*.



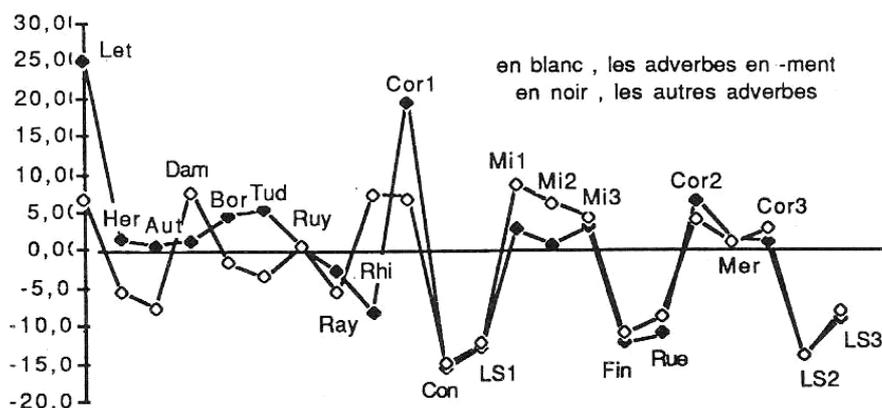
Graphique 4. Les catégories nominales

Un parallélisme semblable marque aussi la distribution des verbes et des mots de relation dans le graphique 5. Les deux courbes s'accordent pour repousser les recueils poétiques et faire bon accueil à la correspondance et au théâtre en prose.



Graphique 5. Verbes et mots de relation

L'incompatibilité des adverbes et de la poésie est encore plus forte dans le graphique 6, Les deux séries d'adverbes suivent des chemins qui divergent peu. On note seulement que les romans préfèrent les adverbes en -MENT, la correspondance et le théâtre les autres adverbes.



Graphique 6. Les adverbes

Les trois graphiques qu'on vient de parcourir à grands pas soulignent tous l'excentricité du *Rhin*. C'est ce texte qui manifeste l'aversion la plus forte pour le verbe et les mots de relation. C'est le *Rhin* qui attire le plus d'adjectifs et d'adverbes en *-MENT* (qui sont en réalité des adjectifs du verbe). C'est là enfin qu'on trouve la concentration la plus intense de ces adjectifs substantivés qui désignent les gens. Ces choix reflètent assez la situation particulière de ce récit de voyage, qui est moins dramatique que descriptif, et qui a moins d'action et de dialogue que d'évocations et de réflexions. Les catégories qui y sont sollicitées sont celles auxquelles a recours Chateaubriand dans les *Mémoires d'outre-tombe*.

Au reste l'excentricité du *Rhin* est soulignée par l'analyse factorielle de la figure 7. Ce récit de voyage a beau prendre la forme épistolaire, en prenant position à l'extrême droite du graphique, il se situe aux antipodes des recueils de correspondance qui occupent la moitié gauche. Et s'il s'accorde avec la poésie dans le premier facteur, il s'oppose à elle dans le second facteur, en se fixant tout au nord, quand les *Contemplations* et les autres recueils poétiques se placent au sud. Au total, le *Rhin* se range du côté des romans, mais sans conviction. Dès le troisième facteur, il reprend sa liberté, en occupant là encore une position extrême.

Plus largement la synthèse représentée dans le graphique 7b met en opposition le clan du verbe (avec ses alliés : adverbes et mots de relation), et le clan du nom (avec son acolyte l'adjectif). Le statut de la préposition ne lui permet pas de s'éloigner du substantif. On la trouve donc à droite, dans la zone d'influence du nom<sup>18</sup>. Ce premier facteur

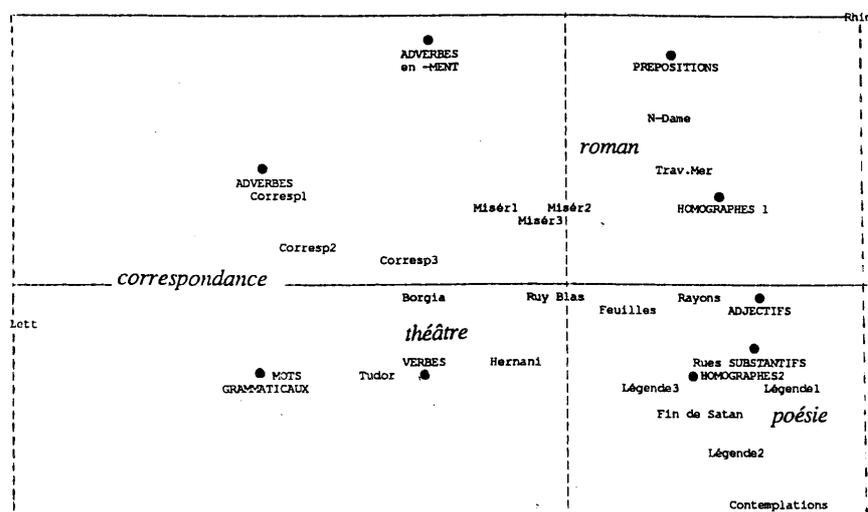
18. Mais la préposition affirme sa spécificité dans le deuxième facteur.

parcourt les textes sur l'axe horizontal et les distribue par genres : la poésie et la correspondance se situent aux extrêmes, tandis que le roman et le théâtre occupent l'entredeux. Si l'on prête attention aux indications numériques reproduites au haut du tableau 7a, on remarque que les textes et les catégories acceptent largement ce premier critère de classement (qui rend compte de 68 % de la variance). Ce qu'on appelle la « qualité d'explication » et qu'on lit dans la première colonne donne entière satisfaction, puisque presque partout la valeur 900 (sur 1000) est dépassée. Une seule catégorie échappe à cet assentiment général : celle des homographes profonds (coefficient 461) qui se trouvent empêchés par leur division intérieure de prendre parti. Parmi les textes un seul refuse de souscrire au référendum : *Ruy Blas*. Son coefficient 307 (sur 1000) manifeste le désarroi d'un texte qui appartient au théâtre et à la poésie en même temps et que l'axe horizontal écartèle<sup>19</sup>.

Tableau 7a. Analyse factorielle des catégories grammaticales

	I1	QLT	POID	INR	1#F	COR	CTR	2#F	COR	CTR	3#F	COR	CTR	4#F	COR	CTR
1 Lett	996	43	241	-1117	988	348	-65	3	4	-50	2	11	-61	3	32	
2 Hern	849	44	10	-121	274	4	-148	408	20	88	143	33	-35	23	11	
3 Feui	881	44	11	144	385	6	-45	37	2	5	0	0	-157	458	222	
4 Dame	897	50	32	176	212	10	311	662	103	50	17	13	-27	5	8	
5 Borg	906	44	17	-274	849	21	-27	8	1	65	47	18	-12	2	1	
6 Tudo	916	43	39	-400	765	44	-168	135	26	57	16	14	-16	1	2	
7 Ruy	307	45	3	-63	273	1	-18	22	0	-10	7	0	-8	4	1	
8 Rayo	970	46	20	275	751	22	-26	7	1	-87	75	34	-118	138	128	
9 Rhin	986	54	167	624	556	135	519	385	310	-176	45	167	14	0	2	
10 Cor1	971	46	83	-585	832	101	162	64	26	-144	51	95	-99	24	9	
11 Cont	976	41	63	394	447	41	-420	507	154	-89	23	32	-2	0	0	
12 L g1	979	44	49	451	804	57	-188	139	33	-96	36	40	7	0	0	
13 Mis1	652	50	21	-81	71	2	146	229	23	163	282	130	81	70	66	
14 Mis2	939	49	9	-11	3	0	154	569	25	107	273	55	63	95	39	
15 Mis3	710	48	8	-67	115	1	118	353	14	92	218	40	31	24	9	
16 Fin	911	42	28	275	511	21	-238	384	51	3	0	0	47	15	19	
17 Rues	984	44	23	304	766	26	-152	191	21	49	20	10	30	7	8	
18 Cor2	975	45	57	-502	870	73	60	13	3	-105	38	49	125	54	14	
19 Mer	935	49	31	220	334	15	214	315	48	195	260	183	-61	26	37	
20 Cor3	909	44	27	-297	638	25	38	10	1	-132	126	76	137	135	167	
21 Leg2	983	43	46	370	551	37	-325	427	96	-5	0	0	30	4	8	
22 Leg3	990	43	14	171	406	8	-203	572	38	-1	0	0	30	12	8	
1 SUBS	997	247	178	370	836	217	-141	122	105	57	20	81	-57	20	160	
2 ADJ	975	99	94	392	711	98	-48	10	5	-234	253	534	13	1	4	
3 HOM1	836	72	60	364	700	62	143	108	31	56	17	22	46	11	30	
4 HOM2	461	18	28	374	391	16	-141	56	8	62	11	7	33	3	4	
5 VERB	966	196	92	-257	617	83	-156	228	102	88	73	151	71	48	201	
6 MENT	968	73	91	-303	325	43	419	622	275	36	4	9	67	16	66	
7 ADV	998	84	160	-607	848	198	185	79	61	12	0	1	-175	70	516	
8 PREP	969	95	102	271	303	45	400	660	324	-25	3	6	25	3	12	
9 GRAM	979	116	194	-565	840	238	-191	96	90	-128	43	189	19	1	8	

19. La situation de *Hernani* est la même et sa position sur le graphique est également hésitante, à mi-chemin entre le théâtre et la poésie, entre le substantif et le verbe.



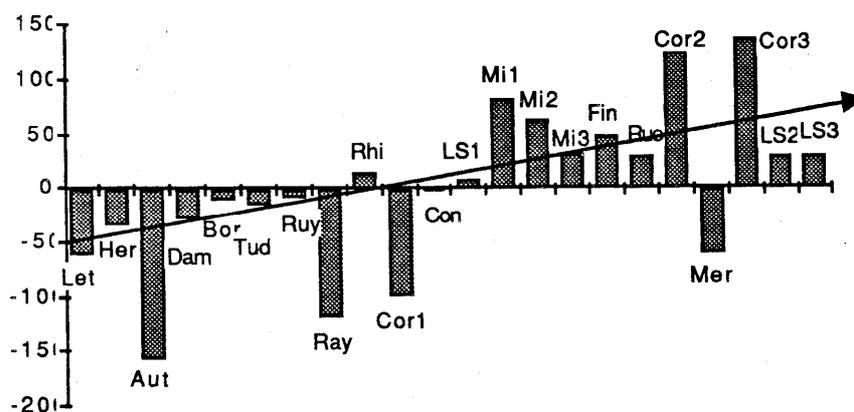
Graphique 7b. Analyse factorielle des catégories grammaticales

Mais l'expression des différences et des préférences ne se réduit pas au seul premier facteur. Le second n'est pas négligeable puisqu'il rend compte de 21 % de la variance. C'est lui qui exprime la spécificité du roman au haut du graphique et l'emploi romanesque des catégories aptes à nuancer l'action (les adverbes en *-MENT*) ou à enrichir les circonstances (les prépositions). La poésie et le théâtre se dispensent de ces équipements un peu sophistiqués et campent de l'autre côté, en bas du graphique, dans une région plus dépouillée et plus pure, où règne la simplicité du verbe et du nom.

Mais où est donc la chronologie ? Les deux premiers facteurs que nous venons d'examiner recouvrent les genres. Et c'est le genre encore qui s'impose dans le troisième, pour opposer les romans à la correspondance, et les verbes aux adjectifs<sup>20</sup>, ce qui amène le *Rhin*, qui est plus riche d'adjectifs que de verbes, à abandonner ses amis romanesques et à rejoindre les recueils épistolaires. La chronologie n'apparaît que tardivement et timidement, avec le quatrième facteur. Observons dans le tableau 7a la colonne qui porte le symbole 4#F et qui contient les coordonnées de ce facteur. Les 8 premières sont négatives, et

20. L'opposition entre verbes et adjectifs est souvent plus radicale que celle qui se manifeste entre le verbe et le substantif. Mais la catégorie des adjectifs n'est pas pure ici, puisqu'elle englobe les participes. Et comme les participes – au moins les participes passés – ont partie liée avec le verbe, l'opposition est émuée.

les 11 dernières positives, à une exception près. Il y a donc une tendance, faiblement esquissée, qui intéresse surtout les adverbes et les verbes, les premiers diminuant, les seconds augmentant avec le temps. Cette tendance a été matérialisée dans le graphique 8, à partir des coordonnées du facteur 4. Mais la netteté du mouvement ne doit pas faire illusion, le facteur 4 ne représente que 2 % de l'inertie totale.



Graphique 8. L'influence du temps sur les parties du discours

### 3. Adjectifs et participes

Les catégories grammaticales n'ont pas de délimitations catégoriques. Elles forment un continuum et, du verbe au substantif, des mots pleins aux mots-outils, on peut se promener en pente douce sans jamais rencontrer de rupture. Soit des termes comme *PENDANT*, *DURANT* ou *SUIVANT* ; selon le contexte, on peut avoir affaire à un verbe, un adjectif, un nom, une préposition. Même si le dictionnaire voit là des homographes et leur attribue des entrées distincte<sup>21</sup>, l'écrivain et son lecteur reconnaissent l'étymologie où les divergences puisent leur source commune – ce qui permet des ambiguïtés, des jeux, des glissements sémantiques et des déguisements grammaticaux. Or l'adjectif est la catégorie qui se prête le plus volontiers à ces changements de costume. On a déjà eu affaire aux adjectifs substantivés, sans qu'il y ait trop d'embarras ; comme l'adjectif et le substantif entrent dans la classe nominale, la catégorie hybride y entrait aussi à double raison. Mais la confusion des adjectifs et des participes, et donc des classes nominale et verbale, pose un problème si

21. Le *Petit Robert* fait la distinction pour *SUIVANT* et *PENDANT*, mais non pour *DURANT*.

épineux qu'aucun lexicométricien n'a su le résoudre. Même Gunnell Engwall, qui n'a pas reculé devant la désambiguïsation de 32 708 articles LE, a renoncé à séparer, dans les participes, ceux qui appartiennent au verbe et ceux qui sont des adjectifs<sup>22</sup>. En les rejetant tous du côté du verbe, elle a certes épuré les adjectifs, mais l'affluent boueux des participes a peut-être pollué les verbes.

Nous avons tenté la décantation sans jeter le trouble ailleurs. En s'appuyant sur la finale, il était possible en effet de distinguer les formes en -É, en -I, en -S, en -T, en -U, en -ANT et celles qui n'avaient aucune de ces marques et qui ne pouvaient donc pas être des participes. On obtenait ainsi par soustraction la classe des adjectifs purs. Il y avait certes de l'arbitraire à décréter que les formes en -É ou en -I étaient des participes passés et les formes en -ANT des participes présents, mais on avait tout de même progressé en isolant les vrais adjectifs. Quant aux formes en -U, en -S et en -T, divers critères de tri – et dans les cas ultimes une liste d'exceptions – ont permis de retenir ceux qui pouvaient se prévaloir d'un verbe existant et jouer le rôle d'un participe. Le détail du relevé figure dans le tableau 9.

Contentons-nous d'abord du total obtenu pour chaque variété, en le comparant à celui qu'on relève dans le corpus du TLF<sup>23</sup>. Voici les différences constatées :

	é	i	s	u	t	ant	adj	total
corpus	1349711	249479	116339	520980	415050	659360	3947579	7258498
réel Hugo	37624	7353	3494	13954	11106	21463	117014	212008
théoriq. Hugo	39840	7364	3434	15378	12251	19463	116522	214251
écart réduit	-11,27	-0,13	1,04	-11,66	-10,50	14,56	1,46	-4,92

22. « Dans le cas du participe passé, plusieurs solutions ont été proposées en vue de faire le départ entre les valeurs verbale et adjectivale : relevé de la présence ou non-présence de l'auxiliaire ou établissement d'une liste de participes à considérer comme adjectifs. Comme nous Lyne (1973 : 88) insiste sur l'impossibilité de distinguer les deux valeurs. Et il ramène aussi au verbe toute occurrence du participe passé ». Gunnell Engwall, *Vocabulaire du roman français (1962-1968)*.

23. En réalité on n'a pas renouvelé pour cette recherche l'exploration du grand corpus XIX-XX<sup>e</sup>, puisque nous l'avions déjà réalisée, à propos de Zola. On a donc repris les effectifs du TLF, tels qu'ils figurent dans notre *Vocabulaire de Zola* (1985a), tome 1, p. 211 (même s'ils ne s'accordent pas tout-à-fait avec ceux du tableau 2). La catégorie inclut ici les adjectifs substantivés.

Tableau 9. Participes et adjectifs. Effectifs

	vocables	occurr.	en é	en i	en s	en u	en t	ant	adj	total
Lettres	3731	93890	1775	292	75	582	328	505	3990	7547
Hernani	2476	22833	294	64	38	129	176	269	1054	2024
Feuilles d'a.	3001	24866	397	104	26	126	121	330	1401	2505
Notre-Dame	9168	176515	3487	532	378	1330	878	1771	10100	18476
Borgia	2230	20719	316	59	20	115	145	167	826	1648
Tudor	2179	24947	377	127	38	139	132	135	989	1937
Ruy Blas	3306	29868	480	115	40	191	185	486	1480	2977
Rayons	3384	28597	487	126	38	141	137	458	2002	3389
Rhin	9949	199799	4833	783	423	1423	1089	1892	13379	23822
Corresp.1	5527	117402	2282	375	196	1088	635	763	5964	11303
Contempl	5475	87827	1032	416	127	457	454	1425	5768	9679
Légende 1	5892	74820	992	302	110	389	378	1450	5139	8760
Misér. 1	9091	180672	3697	582	349	1188	744	1657	9395	17612
Misér. 2	10005	187724	3656	664	298	1252	906	1729	10396	18901
Misér. 3	9183	165869	3663	573	296	1102	837	1637	8455	16563
Fin de S.	4174	48237	665	240	69	206	259	716	3054	5209
Chansons...	4213	31580	347	163	50	172	132	388	2123	3375
Corresp. 2	6961	186022	2931	558	295	1473	1509	1458	9754	17978
Trav. Mer	8824	134135	2666	378	205	893	605	1356	7295	13398
Corresp. 3	5483	108914	1665	363	210	948	808	916	6087	10997
Légende 2	6058	92763	1136	379	128	466	444	1442	6113	10108
Légende 3	4189	36287	446	158	85	144	204	513	2250	3800
total1	20602	2074286	37624	7353	3494	13954	11106	21463	117014	212008
poésie		424977	5502	1888	633	2101	2129	6722	27850	46825
th.vers		52701	774	179	78	320	361	755	2534	5001
th.prose		45666	693	186	58	254	277	302	1815	3585
roman		844915	17169	2729	1526	5765	3970	8150	45640	84949
corresp		706027	13486	2371	1199	5514	4369	5534	39175	71648

Si l'écart est négatif (-4,92) pour l'ensemble, on sait à quoi attribuer ce déficit : aux participes passés ou plutôt aux formes en -é, en -u, ou en -t, qui ont un écart réduit respectivement de -11, -12 et -10. Les participes présents – les formes en -ant – ont au contraire un très large excédent ( $z = 14$ ). Quant aux vrais adjectifs, leur nombre tombe à la moitié de l'ensemble (117 014 sur 212 008). Et la proportion qu'ils représentent dans le discours hugolien est à peine plus forte que celle qu'on observe dans l'ensemble des écrivains du corpus. L'écart réduit est positif mais il n'atteint pas le seuil significatif ( $z = 1,46$ ). Plus d'un lecteur de Hugo sera surpris par ce résultat. On a rencontré si souvent sous la plume de Hugo des êtres SOMBRES, FAROUCHES, MONSTRUEUX, PENSIFS, et des choses OBSCURES, SINISTRES, PROFONDES et FORMIDABLES qu'on en conclut un peu vite – comme fait Musset – à une manie hugolienne de l'adjectif. Mais la lecture est toujours plus sensible aux excédents qu'aux déficits. Dans une foule on remarque plus les grands que les petits, parce que leur tête dépasse. Et l'on a peu conscience du sous-emploi de certains adjectifs qui ne plaisaient guère à Hugo ou qui sont devenus à la mode après lui, comme NOUVEAU, MORAL, GÉNÉRAL, NATUREL, PHYSIQUE, SEUL, PREMIER, DERNIER, MODERNE,

VÉRITABLE, PARTICULIER, NOMBREUX, DIFFÉRENT, INTELLECTUEL, SCIENTIFIQUE, SUPÉRIEUR, POLITIQUE, etc. Comme la statistique radiographie la foule, elle voit les petits comme les grands, les déficits aussi bien que les excédents, et dans le cas des adjectifs chez Hugo elle estime que les premiers compensent les seconds<sup>24</sup>.

Tableau 10. Participes et adjectifs. Écarts réduits

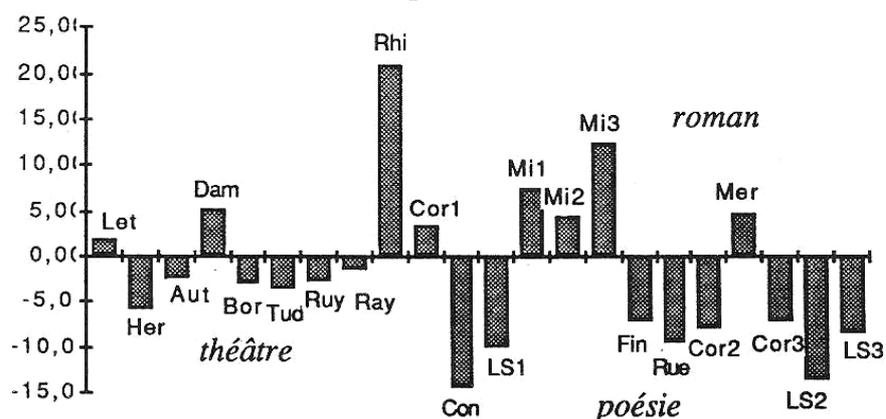
	é	i	s	u	t	ant	adj
<i>Lettres</i>	1,79	-2,29	-6,77	-2,02	-7,97	-15,32	-18,37
<i>Hernani</i>	-5,94	-1,89	-0,07	-2,00	4,89	2,14	-6,56
<i>Feuilles d'a.</i>	-2,56	1,70	-2,47	-3,21	-1,06	4,56	-0,05
<i>Notre-Dame</i>	5,27	-3,92	4,89	4,33	-2,28	-1,36	1,49
<i>Borgia</i>	-3,10	-1,69	-2,53	-2,08	3,25	-3,25	-10,08
<i>Tudor</i>	-3,57	4,13	-0,62	-2,24	-0,14	-7,71	-11,22
<i>Ruy Blas</i>	-2,67	0,89	-1,46	-0,71	2,00	10,14	-5,03
<i>Rayons</i>	-1,40	2,46	-1,48	-3,73	-1,31	9,49	9,75
<i>Rhin</i>	21,13	2,95	4,96	2,26	0,62	4,06	20,89
<i>Corresp.1</i>	3,40	-2,08	-0,13	10,93	0,26	-13,35	-8,34
<i>Contempl</i>	-14,36	6,06	-1,76	-5,63	-0,77	17,50	11,81
<i>Légende 1</i>	-10,09	2,30	-1,45	-5,19	-1,15	24,74	14,40
<i>Misér. 1</i>	7,68	-2,42	2,68	-0,82	-7,52	-5,14	-8,26
<i>Misér. 2</i>	4,51	-0,06	-1,07	-0,32	-3,28	-5,08	-1,98
<i>Misér. 3</i>	12,44	-0,64	1,04	-0,43	-1,79	-2,00	-9,72
<i>Fin de S.</i>	-7,18	5,34	-1,38	-6,66	0,05	9,82	6,46
<i>Chansons...</i>	-9,51	4,86	-0,44	-2,80	-2,87	3,41	8,15
<i>Corresp. 2</i>	-8,00	-4,14	-1,09	6,57	17,04	-11,15	-7,57
<i>Trav. Mer</i>	4,88	-4,62	-1,44	-0,32	-4,37	-0,89	-3,23
<i>Corresp. 3</i>	-7,18	-1,21	2,01	8,17	9,57	-6,46	-0,75
<i>Légende 2</i>	-13,63	2,83	-2,31	-6,47	-2,42	15,92	12,45
<i>Légende 3</i>	-8,34	2,61	3,08	-6,46	0,70	7,16	4,53
<i>poésie</i>	-28,18	11,02	-3,47	-15,90	-3,44	39,31	28,08
<i>théâtre vers</i>	-5,96	-0,58	-1,16	-1,86	4,75	9,10	-8,15
<i>théâtre prose</i>	-4,75	1,92	-2,18	-3,07	2,10	-7,93	-15,16
<i>roman</i>	19,35	-6,32	3,54	1,40	-10,70	-8,23	-12,04
<i>correspond.</i>	7,40	-3,24	0,35	13,66	11,79	-25,52	-4,03

L'hyperbole hugolienne ne tient donc pas à la structure du vocabulaire, à quelque déséquilibre des parties du discours, à un

24. Si l'on veut toucher du doigt ce phénomène de compensation, il suffit de feuilleter le dictionnaire des fréquences de Hugo (le second tome de notre ouvrage *Le vocabulaire de Victor Hugo*, 1988a). Si on examine les adjectifs les uns après les autres, les écarts se montrent aussi souvent dans la zone négative, et même plus souvent. Un simple sondage le laisse entendre, même s'il est limité aux adjectifs qui commencent par un A et dont la fréquence est supérieure à 100. On en compte 16 dont 9 sont négatifs.

développement cancérigène de l'adjectif ou de l'adverbe. Elle tient au choix de certains termes, adjectifs ou non. Elle tient à la sémantique, non à la syntaxe. Et nous rejoignons là le constat auquel nous avait conduit l'étude des *Rougon-Macquart* : contrairement à certains critiques, la machine ne songe pas à reprocher à Zola son abus de l'adjectif, puisque cet abus n'existe pas.

À l'endroit des participes, Hugo est plus sélectif que Zola. Comme Zola, et en réalité comme le XIX<sup>e</sup> siècle, il utilise très largement les formes en *-ANT* (10 % de plus que la moyenne). Mais contrairement à Zola il se tient à distance du participe passé. En réalité c'est le genre qui l'en détourne dans certaines situations. Car le participe passé trouve aisément à s'employer dans le récit, beaucoup moins au théâtre et encore moins en poésie. C'est ce qu'on voit à l'intérieur même du corpus Hugo. Suivons par exemple la trace des formes en *-É* dans le graphique 11. Elle mène tout droit au roman, et particulièrement aux *Misérables* et au *Rhin*, en se tenant à l'écart du théâtre et de la poésie<sup>25</sup>.



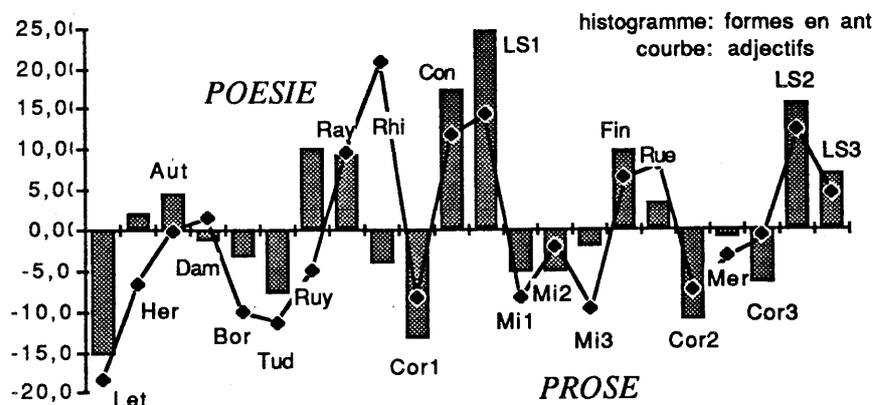
Graphique 11. Les formes en *-É*

Le graphique 12, qui superpose la distribution des formes en *-ANT* et des adjectifs<sup>26</sup>, donne au contraire la préférence à la poésie,

25. Les formes en *-É* n'ont pas une distribution homogène dans la correspondance. Excédentaires dans les deux premiers recueils, elles se font plus rares dans les deux derniers. Le motif en est accidentel : les effusions sentimentales ou familiales multiplient les occurrences de *ADORÉ* et *AIMÉ* dans les lettres de la première période (114 emplois de *ADORÉ* dans les *Lettres à la fiancée* sur 171 en tout, et 201 de *AIMÉ* sur 689). Ces mots sont évidemment moins fréquents quand Hugo s'adresse à ses éditeurs.

26. Le parallélisme n'est pas absolu. Il y a des moments où les adjectifs s'éloignent des participes présents. Ainsi ils prennent leur envol dans le *Rhin*, tandis que fléchit la

particulièrement à la *Légende des siècles*. Ce n'est pas là une surprise, mais une confirmation, Nous avons déjà noté dans notre *Vocabulaire français*<sup>27</sup> que les formes en *-ANT* et les adjectifs avaient la faveur du corpus en vers, et qu'inversement les participes en *-É*, mais aussi en *-U* et en *-T*, s'orientaient vers la prose littéraire. Ces traits se retrouvent chez Hugo, avec plus de force probante encore, puisque la diversité des auteurs est neutralisée et que le mélange des époques est réduit.



Graphique 12. Adjectifs et formes en *-ANT*

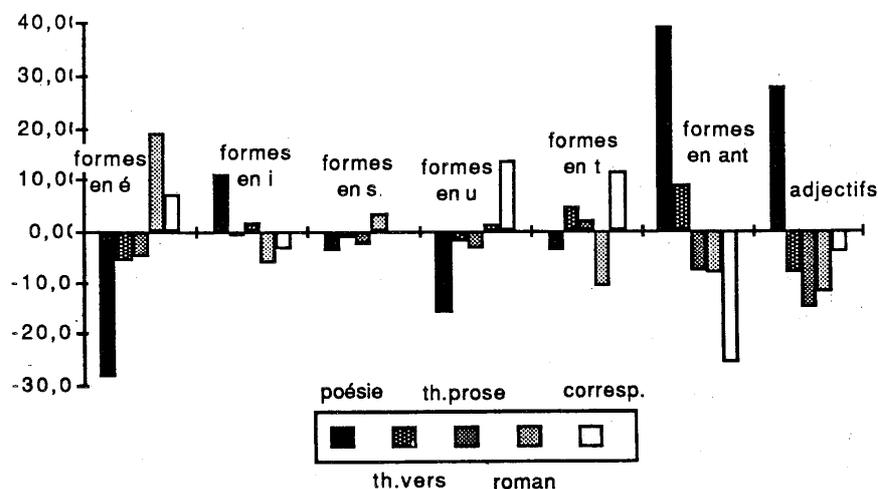
Les adjectifs, les formes en *-ANT* et les formes en *-É* ont des profils bien tranchés et contrastés. Il n'en est pas toujours ainsi des autres espèces, comme on le voit dans le graphique 13.

Mais l'ensemble des formes en *-I*, en *-S*, en *-U* et en *-T* ne représente qu'un effectif limité, en ce qui concerne les occurrences, et plus réduit encore si l'on envisage les vocables :

	<i>i</i>	<i>s</i>	<i>u</i>	<i>t</i>	total	catégorie	pourcentage
occurrences	7353	3494	13954	11106	35907	212008	17%
vocables	157	38	113	55	363	3431	10%

fréquence des formes en *-ANT*.

27. (1981a), voir p. 309-313 et 358-359.



Graphique 13. Profils des catégories d'adjectifs et de participes

En de tels cas les écarts sont de plus faible amplitude et les accidents dus à des individualités lexicales s'y font sentir davantage. Ainsi les mots *INFINI* ou *JOLI* orientent sans doute les formes en *-I* du côté des adjectifs<sup>28</sup>. Néanmoins le profil des formes en *-U* et en *-T* a un air de famille qui évoque celui des formes en *-É*. Les formes en *-S*, trop peu représentées pour avoir un dessin bien net, se rapprochent aussi de ce modèle. La figure 13, qui juxtapose tous les profils, laisse entendre qu'il y a en réalité deux modèles à suivre. L'un est celui du participe passé et il est commun aux formes en *-É*, en *-S*, en *-U*, et en *-T*. L'autre, qui est celui de l'adjectif, rallie les formes en *-ANT* et en *-I*.

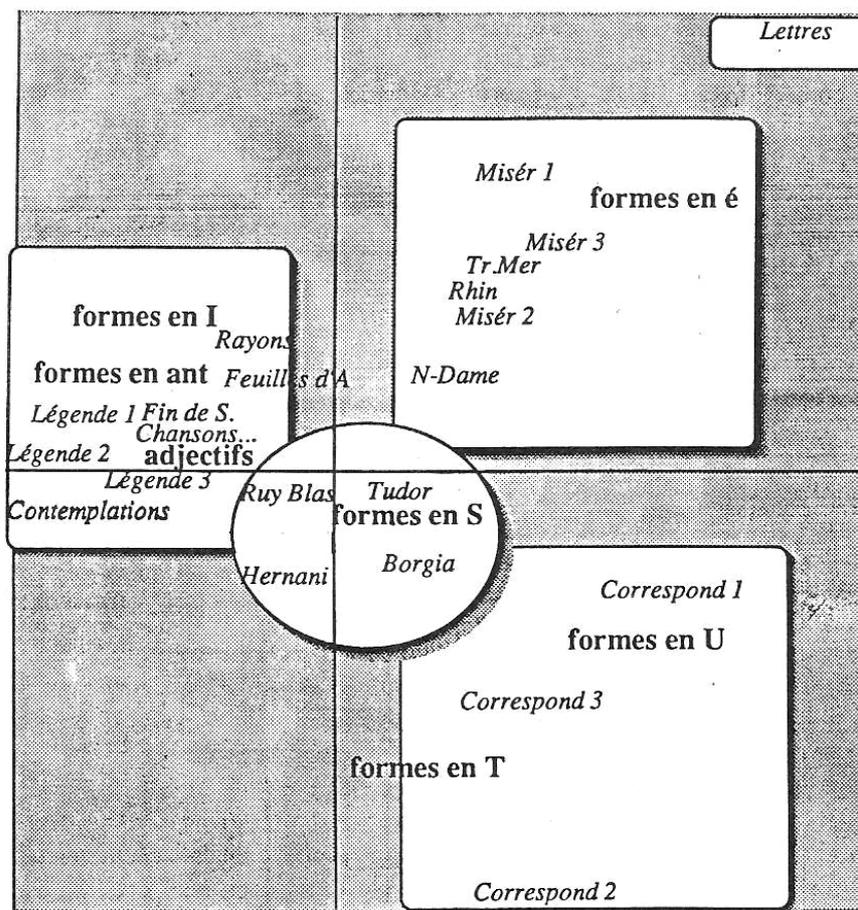
Ces deux modèles sont opposés nettement dans l'analyse factorielle réalisée à partir des données du tableau 10 (graphique 14). Adjectifs et formes en *-ANT* et en *-I* font bloc à gauche dans le cercle étroit que délimitent les huit recueils poétiques<sup>29</sup>. Les participes passés dispersent leurs variétés à droite dans le champ ouvert de la prose, les formes en *-É* se concentrant dans les romans<sup>30</sup> (quadrant supérieur droit) et les autres

28. Ce n'est sans doute pas la seule raison. Dans un corpus beaucoup plus large où 240 000 formes en *-I* avaient été recensées, on a observé aussi le rapprochement des adjectifs et des formes en *-I*.

29. Ce cercle a beau être étroit, il comprend cependant des zones d'influence, et les deux premiers recueils ne se mêlent pas aux autres.

30. Il faut noter l'intrusion des *Lettres à la fiancée* dans le camp romanesque au haut du graphique. Cela est dû, comme nous l'avons noté, à l'afflux de quelques mots tendres qui par un hasard lexical se trouvent en *-É* et que l'amoureux répète à satiété.

espèces préférant la correspondance (quadrant inférieur droit). Comme à l'accoutumée, le théâtre s'interroge à la croisée des chemins, au milieu du graphique. Les pièces en vers tirent à hue, du côté de la poésie, et les pièces en prose tirent à dia, de l'autre côté de l'axe vertical.



Graphique 14. Analyse factorielle des adjectifs et participes

C'est donc le genre qui polarise les lignes de force dans le champ magnétique des adjectifs et des participes. Or paradoxalement ces lignes demeurent assez semblables dans le corpus de Zola, où l'opposition des genres est moindre. Si l'on se reporte à l'analyse effectuée dans ce corpus romanesque à partir des mêmes catégories et avec les mêmes principes de

distinction<sup>31</sup>, on voit pareillement le regroupement des participes passés en -s, en -u et en -t sous la bannière des formes en -é, et l'alliance des adjectifs et des participes présents<sup>32</sup>. Où donc se fait la division ? Dans les catégories ou dans le genre ? Ou dans les deux à la fois ? L'analyse factorielle suppose que les oppositions se font jour semblablement entre les variables (ou les colonnes du tableau, ici les catégories) et aussi entre les individus (ou les lignes du tableau, ici les textes). Or même quand ces textes n'ont pas de marques distinctives, parce qu'ils appartiennent au même genre, ou à la même époque, ou au même auteur, ils finissent toujours par trouver des raisons de se combattre et de se disputer les mots. Assemblez des joueurs paisibles et indistincts, sur un terrain plat et neutre, avec un ballon rond et lisse, et revenez quelques instants plus tard. Vous trouverez des équipes qui se font la guerre.

### Bibliographie

- É. Brunet, *Le Vocabulaire français de 1789 à nos jours, d'après les données du Trésor de la langue française*, 3 vol., Travaux de linguistique quantitative 17, Genève-Paris : Slatkine-Champion, 1981 (1981a).
- É. Brunet, *Le Vocabulaire de Proust*, 3 vol., Travaux de linguistique quantitative 18, Genève-Paris : Slatkine-Champion, 1983 (1983a).
- É. Brunet, *Le Vocabulaire de Zola*, 3 vol., Travaux de linguistique quantitative 26, Genève-Paris : Slatkine-Champion, 1985 (1985a).
- J.-M. Cotteret et R. Moreau, *Recherches sur le vocabulaire du Général de Gaulle*, Travaux et recherches de science politique 3, Paris : Armand Colin, 1969.
- Dictionnaire des fréquences. Vocabulaire littéraire des XIX<sup>e</sup> et XX<sup>e</sup> siècles*, 4 vol., Didier, Paris, 1971.
- G. Engwall, *Fréquence et distribution du vocabulaire dans un choix de romans français*, Stockholm : Skriptor, 1974.
- G. Engwall, « Valeurs quantitatives et impression du lecteur », dans Th. Pettersson (éd.), *Papers from the Fifth Scandinavian Conference of Linguistics*, Acta Universitatis Lundensis, Sectio 1. Theologica

---

31. *Le Vocabulaire de Zola* (1985a), tome 1, p. 222.

32. Les formes en -t font cependant bande à part.

- Juridica Humaniora 30, Stockholm : Almqvist & Wiksell International, 1979, p. 93-104.
- G. Engwall, *Vocabulaire du roman français (1962-1968). Dictionnaire des fréquences*, Data linguistica 17, Stockholm : Almqvist & Wiksell International, 1984.
- A. Geffroy et P. Lafon, « L'insécurité dans les grands ensembles », *Mots*, 5, 1982, 129-141.
- G. Gougenheim *et al.*, *L'Élaboration du français fondamental (1er degré)*, nouv. éd., Paris : Didier, 1967.
- P. Guiraud, *Les Caractères statistiques du vocabulaire*, Paris : Presses Universitaires de France, 1954.
- P. Guiraud, *Problèmes et méthodes de la statistique linguistique*, Dordrecht : D. Reidel Publishing Company, 1959.
- B.-L. Gunnarsson, *Lagtexters begriplighet*, Stockholm : Liber, 1982.
- G. Henry, *Comment mesurer la lisibilité*, Éducation 2000, Bruxelles-Paris : Éditions Labor-Fernand Nathan, 1975.
- A. Juilland, D. Brodin et C. Davidovitch, *Frequency Dictionary of French Words*, The Romance Languages and Their Structures, First Series, FI, The Hague-Paris : Mouton, 1970.
- Méthodes quantitatives et informatiques dans l'étude des textes*, en hommage à Charles Muller, Genève-Paris : Slatkine-Champion, 1986 (1986c).
- Ch. Muller, *Étude de statistique lexicale. Le Vocabulaire du Théâtre de Pierre Corneille*, Paris : Larousse, 1967.
- Ch. Muller, *Initiation aux méthodes de la statistique linguistique*, Paris : Classiques Hachette, 1973.
- Ch. Muller, *Principes et méthodes de statistique lexicale*, Paris : Classiques Hachette, 1977.
- Ch. Muller, « De la lemmatisation », préface à P. Lafon, *Dépouillements et statistiques en lexicométrie*, Genève-Paris : Slatkine-Champion, 1984.
- P. Richaudeau, *Les Secrets de la communication efficace*, Collection Savoir Communiquer, Paris : La Bibliothèque du CEPL, 1975.