HAL
open science

Le dictionnaire de Hugo a-t-il un bonnet rouge ?

Étienne Brunet

To cite this version:

Étienne Brunet. Le dictionnaire de Hugo a-t-il un bonnet rouge ? . Cahiers de Lexicologie, 1987, 50, pp.35-53. hal-01574300

HAL Id: hal-01574300

https://hal.science/hal-01574300v1

Submitted on 13 Aug 2017

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.

Le dictionnaire de Hugo a-t-il un bonnet rouge ?[1]

Etienne Brunet

« Ni talons rouges, ni bonnets rouges », recommande Hugo dans la *Préface d'Hernani*. Tout en clamant haut et fort le principe du « libéralisme en littérature » et tout en se félicitant « qu'à une littérature de cour succède une littérature de peuple », il n'en maintient pas moins la *nécessité de l'ordre* : « Dans les lettres, comme dans la société, point d'étiquette, point d'anarchie : des lois. »

Vingt-cinq ans plus tard, dans *Réponse à un acte d'accusation*, le Hugo de l'exil abandonne ces prudences et brandit l'étendard révolutionnaire :

J'ai mis un bonnet rouge au vieux dictionnaire.[2]

Il n'est pas difficile d'expliquer cette surenchère. Hugo vient de traverser la Révolution de 1848 et il a visité successivement l'endroit et l'envers des barricades. Devenu républicain, il tient à montrer que ses convictions n'ont pas tellement varié depuis 1830 et que son combat d'alors pour la liberté de l'art était aussi une lutte pour l'affranchissement de la pensée, puisque délivrer le mot, c'est délivrer la pensée.

Notre propos n'est pas de suivre Hugo dans les implications politiques de son plaidoyer mais d'examiner, à travers les exagérations plaisantes, la part réelle de Hugo, sinon dans le renouveau des formes littéraires, du moins dans la libération et l'enrichissement du vocabulaire. S'en tenir aux considérations lexicales, c'est certes restreindre la portée du discours hugolien qui s'en prend aux figures, aux tropes, à la rhétorique aussi violemment qu'à l'étroitesse du goût classique en

---

1. Article paru dans les *Cahiers de lexicologie*, n°50, 1987-1, p. 35-53.

2. En revenant sur la bataille *d'Hernani*, il rétablit une expression audacieuse qu'il avait dû atténuer pour la scène :

On entendit un roi dire : "Quelle heure est-il ?"

La version édulcorée propose une transcription moins prosaïque : "*Est-il minuit* ?"

matière de lexique[3]. Mais les méthodes quantitatives que nous prétendons utiliser ne permettent guère que la prise en compte d'unités simples. Et les mots sont des objets de cette sorte (du moins quand on en donne une définition graphique), alors que les figures de rhétorique, pour être comptables et accessibles à la machine, exigent un codage préalable, nécessairement humain, et donc sujet à l'arbitraire, à l'approximation, voire à l'incohérence. Mais s'en tenir au mot, c'est, pour Hugo, retenir l'essentiel : toute la place qui est disputée à la rhétorique, Hugo la donne au mot, dont le champ s'étend sans limite et se confond avec la pensée. Car il y a chez Hugo une théorie du mot, largement développée dans le poème qui fait suite à la *Réponse* et que Hugo intitule cavalièrement *Suite.* Certains vers en sont si célèbres que les commentaires deviennent superflus :

Car le mot, qu'on le sache, est un être vivant...
Les mots sont les passants mystérieux de l'âme...
Car le mot, c'est le Verbe, et le Verbe, c'est Dieu.

Mais on sait moins qu'il y a aussi chez Hugo un goût des chiffres autant que des lettres, ce dont témoignent précisément certains passages de la *Réponse à un acte d'accusation :*

J'y fis entrer le chiffre ; ô terreur ! Mithridate
Du siège de Cyzique eût pu citer la date.

Et Hugo, donnant l'exemple, de citer dans ce même texte *quatre-vingt-neuf* et *Louis-Quatorze*[4]. C'est assez pour qu'on se croie autorisé à faire entrer le chiffre dans l'oeuvre de Hugo, heureux de ne pas trouver chez lui cette ironie déclarée que Zola ou Giraudoux opposaient par avance à la statistique et qui eût dû décourager nos efforts[5].

On ne s'appesantira ni sur l'exposé des méthodes, ni sur la constitution du corpus hugolien. La méthodologie a été mise au point, pour l'essentiel, par Charles Muller, dont les manuels fourniront les explications à ceux des lecteurs à qui les formules de la lexicométrie sont

3. Il n'y a pas lieu de parler de syntaxe, puisque Hugo déclare accepter sans réserve les règles grammaticales : « Guerre à la rhétorique et paix à la syntaxe ».

4. Le mot *chiffre* qui compte 121 emplois dans notre corpus hugolien manifeste un excédent significatif quand la comparaison est faite avec les autres écrivains français.

5. Proust, que nous avons soumis au même traitement, ne cite que deux fois la statistique et les statisticiens, sans leur accorder un bien grand crédit Mais il admet que le cumul d'erreurs particulières peut produire une sorte de vérité générale, comme les couleurs du spectre produisent la lumière blanche. Voir A *la recherche du temps perdu,* Bibliothèque de la Pléiade, tome 3, p. 298 et 969.

peu familières[6]. Les données sont celles de l'Institut national de la langue française et ont été fournies sous forme de bande magnétique[7]. Elles contiennent quatre recueils de correspondance, quatre pièces de théâtre, trois romans (dont l'intégralité des *Misérables*), et huit recueils de poésie (dont *Les Contemplations* et *La Légende des siècles*)[8]. Il y a là plus de deux millions de mots, dont chacun trouve sa place dans un *Index* synoptique en cours de publication[9]. Certes tout Hugo n'est pas là et certains textes manquent fâcheusement (par exemple *Les Châtiments*). Mais l'étendue du corpus reste considérable et l'échantillon acceptable.

La première question que la lexicométrie se pose quand elle aborde un écrivain peut paraître lassante ou oiseuse : se trouve-t-on devant un vocabulaire riche ou pauvre ? Mais s'agissant de Hugo, l'affaire est d'importance puisque Hugo prétend ouvrir les portes du dictionnaire à tous les mots exclus par un goût trop étroit et accorder la citoyenneté lexicale aux représentants du peuple. Un tel libéralisme devrait multiplier les variétés. Notre relevé en recense 20.602. Est-ce peu ? Est-ce beaucoup ? La réponse est dans la comparaison. Zola, en compte moins (19.337 vocables ou « mots différents »), alors que pourtant l'étendue du corpus zolien est nettement plus grande (2.874.755 occurrences). Le vocabulaire de Hugo l'emporte donc en variété sur celui de Zola. La comparaison est moins facile avec Chateaubriand, Proust et Giraudoux, dont les données nous sont connues. Car si l'effectif des vocables est plus grand chez Hugo, celui des occurrences l'est aussi, et la conclusion reste en suspens à cause de la corrélation évidente qui lie *V* et *N* (les vocables et les occurrences). Il faut donc utiliser une formule de pondération qui

6. *Initiation aux méthodes de la statistique linguistique*, Hachette, 1973, et *Principes et méthodes de statistique lexicale*, Hachette, 1977. On n'oubliera pas les travaux de Pierre Guiraud, auxquels le nom de Wagner est attaché (R.-L. Wagner et P. Guiraud. La méthode statistique en lexicologie, *Revue de l'enseignement supérieur*, 1959).

7. L'exploitation a donc été à notre charge et elle comporte une grande part de programmation spécifique. Mais ces contraintes techniques tombent si l'on envisage des monographies de moindre étendue ou si l'on ne s'intéresse qu'à certains mots, même dans de très larges corpus. Par la voie de la télématique les données de l'INaLF sont en effet accessibles à tout chercheur, grâce à un protocole très simple d'interrogation qui ne nécessite qu'un rapide apprentissage.

8. Il faut ajouter *Le Rhin*, dont le classement fait difficulté.

9. *Le vocabulaire de Victor Hugo*, 1988, éditions Slatkine-Champion. Vol.1, 484 p., vol.2, 637 p., vol.3, 556 p., +27 microfiches normalisées contenant l'index synoptique des oeuvres de Hugo (6878 p.).

neutralise les différences de taille. L'indice *w*, que nous avons proposé à cet effet[10], donne les résultats suivants :

	occurrences	vocables	indice *w*
Hugo	2 074 286	20602	10,13
Chateaubriand	1 398 984	19606	9,71
Zola	2 874 755	19337	10,97
Proust	1 267 069	18322	9,84
Giraudoux	0 671 364	15771	9,43

Comme cet indice, dont la valeur est habituellement proche de 10, évolue en raison inverse de la richesse lexicale, on serait amené à admettre que la variété lexicale de Hugo n'atteint pas celle de Chateaubriand, ni celle de Proust ou de Giraudoux. Mais deux ou trois faits sont à prendre en considération : d'une part, de 1789 à nos jours, un phénomène d'inflation lexicale s'est donné libre cours et il n'est peut-être pas légitime de mesurer à la même toise des écrivains que plus d'un siècle sépare. D'autre part, la notion de *richesse lexicale* est assez ambiguë, selon qu'elle repose ou non sur les facilités de la suffixation, de la préfixation et de l'invention lexicale. On a pu montrer que Zola, méfiant à l'égard de l'abstraction, a peu souvent recours aux infixes et aux ressources de la langue qui permettent de faire du neuf avec du vieux. On observe la même réserve chez Hugo à l'égard des suffixes abstraits et cela représente un handicap numérique par rapport aux écrivains qui n'ont pas ces scrupules. Enfin et surtout le genre littéraire perturbe les résultats. Il se trouve que le corpus de Hugo, à la différence des autres corpus précités, incorpore des textes en vers et des lettres. Or ni la poésie, ni la correspondance, ne se prêtent aux curiosités lexicales et à l'enrichissement de la langue. En limitant la confrontation à l'intérieur d'un genre commun, par exemple la prose romanesque, Hugo reprend l'avantage, sinon sur Giraudoux, du moins sur Chateaubriand et Proust[11] :

	occurrences	vocables	Indice *w*
Hugo (roman)	844.915	17.212	9,45

10. En voici la formule : $w = {}_{N}V^{-a}$ pour a=0,185.

11. La comparaison directe, sans calcul, peut être faite entre les *Travailleurs de la mer* (*N* = 134.135, *V* = 8.824), et le *Temps retrouvé* (respectivement 144.752 et 8.240). Le roman de Hugo, qui est plus court, a pourtant plus de vocables.

Cette valeur de 9,45 n'est d'ailleurs atteinte à aucun moment par la prose du XIXe siècle, les huit tranches qu'on y distingue se distribuant comme suit :

1800	**1825**	**1835**	**1845**	**1855**	**1865**	**1875**	**1885**
11,35	10,84	10,67	10,11	10,23	9,70	9,72	9,84

Mais si l'on isole le genre poétique, l'originalité du vers hugolien se dissout dans la moyenne du siècle : 10,29 dans la poésie de Hugo contre respectivement 10,29, 10,45, 10,42, 9,45, 9,74, 9,08, 9,48, 8,97 dans les huit tranches chronologiques du corpus en vers au XIXe siècle[12]. La conclusion doit donc être nuancée : certes le lexique virtuel de Hugo, largement alimenté d'ailleurs par la lecture systématique des dictionnaires, a une dimension considérable et il se déploie largement dans son discours lorsque la situation le permet. Mais précisément, quoi qu'il prétende, la liberté n'est pas entière dans le genre poétique, où Hugo atteint assez vite les limites du genre, et, s'il se montre accueillant aux nouveautés, aux curiosités, et aux trivialités du vocabulaire, il les entasse au grenier – c'est-à-dire dans la prose, et dans un genre, le roman, où les règles existent à peine – mais non à l'étage noble de la poésie. A ce point de vue, la séparation des genres n'est nullement abolie, malgré la *Préface* de *Cromwell.* Observons en effet le tableau 1 et sa traduction graphique 1. A l'étage le plus bas – à la cave – Hugo entrepose les objets utilitaires, les mots de tous les jours qui se répètent sans cesse et qui ont cours dans la correspondance et le théâtre en prose. L'étage intermédiaire est celui du théâtre en vers et de la poésie et l'indice y montre une valeur moyenne. Enfin les romans – *Le Rhin* se joint à eux – occupent le plan supérieur, où la richesse lexicale est la plus forte. Si on néglige la distinction des genres, on peut mesurer l'évolution chronologique de Hugo[13] en considérant la pente de la courbe, qui est ascendante. Contrairement à ce qu'on a observé chez d'autres écrivains, le lexique de Hugo n'a donc pas tendance à se replier ou à se recroqueviller avec l'âge[14]. Le coefficient de corrélation chronologique (r = 0,47) atteint le

12. En réalité l'indice de la poésie hugolienne ne tient pas compte du théâtre. Mais cet indice, calculé à part pour le théâtre en vers, montre une valeur très voisine : 10,21.

13. On ne se cache pas que la date de publication et la date avouée de composition n'ont qu'une relation lâche avec la date réelle de composition. Tout ce qu'on peut dire d'un recueil de Hugo, c'est qu'aucune pièce n'en a été rédigée après la date de parution du recueil. Cette règle triviale ne suffit pas à assurer la chronologie, et dans au moins un cas – celui de *La fin de Satan* – nous avons préféré l'époque de la rédaction à celle de la publication, trop longtemps différée.

14. Chateaubriand résiste pareillement à la fatigue lexicale et à l'usure des ans.

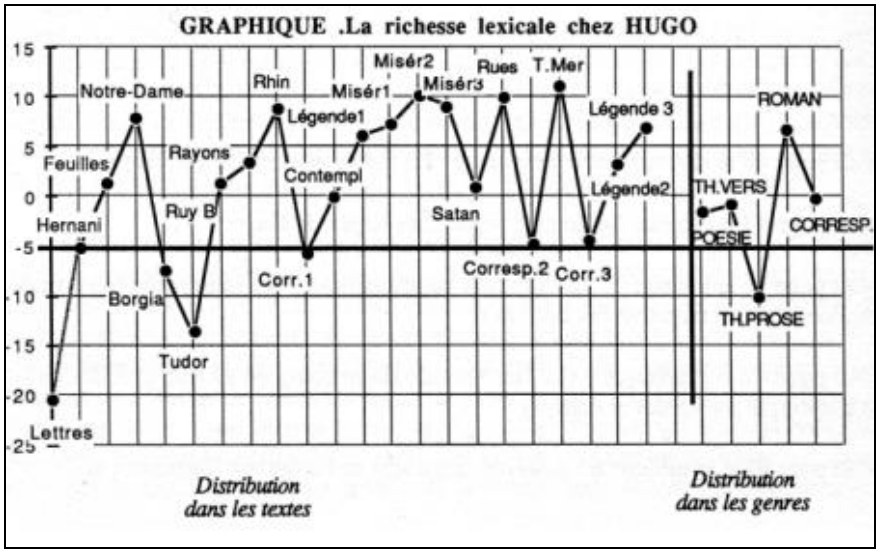
seuil significatif de 3 %[15]. Et le progrès est sensible à l'intérieur des genres, et notamment dans la série des textes en vers : 10,64, 9,99, 9,99, 9,79, 10,13, 9,51, 10,04, 9,14, 9,81, 9,43. La variété lexicale culmine à la fin de la série, dans un recueil plus libre que les autres : *Les Chansons des rues et des bois* et dans le foisonnement épique de *La Légende des siècles*[16].

	Chrono	Occurrences	Vocables	Indice *W*
Lettres à la fiancée	1	93890	3731	12,19
Hernani	2	22833	2476	10,64
Les Feuilles d'automne	3	24866	3001	9,99
Notre-Dame de Paris	4	176515	9168	9,34
Lucrèce Borgia	5	20719	2230	10,88
Marie Tudor	6	24947	2179	11,50
Ruy Blas	7	29868	3306	9,99
Les Rayons et les Ombres	8	28597	3384	9,79
Le Rhin	9	199799	9949	9,24
Correspondance 1	10	117402	5527	10,70
Les Contemplations	11	87827	5475	10,13
La légende des siècles 1	12	74820	5892	9,51
Les Misérables 1	13	180672	9091	9,41
Les Misérables 2	14	187724	10005	9,11
Les Misérables 3	15	165869	9183	9,22
La Fin de Satan	16	48237	4174	10,04
Les Chansons des Rues	17	31580	4213	9,14
Correspondance2	18	186022	6961	10,60
Les Travailleurs de la Mer	19	134135	8824	9,01
Correspondance3	20	108914	5483	10,58
La Légende des siècles 2	21	92763	6058	9,81
La Légende des siècles 3	22	36287	4189	9,43
Total		2074286	20602	10,13
Poésie		424977	10639	10,29
Théâtre en vers		52701	4199	10,21
Théâtre en prose		45666	3199	11,14
Roman		844915	17216	9,45
Correspondance		706027	13545	10 ,18
Total		2074286	20602	10,13

Tableau 1 . La richesse lexicale chez Hugo

15. Rappelons que cette notion de seuil signifie qu'on a 3 chances sur 100 de se tromper lorsqu'on affirme la tendance de Hugo à enrichir son vocabulaire.

16. On peut toutefois hésiter à attribuer au temps ce qui n'est peut-être que l'effet du choix des thèmes. Le lyrisme plus intime qui a la faveur de Hugo dans *Les Feuilles d'automne*, *Les Rayons et les ombres* et *Les Contemplations* n'appelle pas les mêmes ressources lexicales que la poésie ornementale, satirique ou épique. Si notre corpus avait enregistré les *Orientales* et *Les Châtiments*, l'évolution ne serait probablement plus si visible.



Graphique 1. La richesse lexicale chez Hugo

La question des hapax (nous entendons par là les mots rencontrés une seule fois dans le corpus XIX-XXe du T.L.F.) est liée de plus près au problème qui nous occupe. Nous sommes ici aux frontières du dictionnaire, là où convergent les mots étrangers, les noms propres, les régionalismes, l'argot, les mots techniques... et aussi les mots estropiés qui souffrent d'une lettre tordue ou perdue. Comme Hugo est assez friand de couleur locale, d'argot, de termes techniques, de patois, ces vocables ont un effectif non négligeable, et il a fallu les examiner l'un après l'autre. Nous en avons retenu près de 500, dont on trouvera un extrait dans le tableau 2, limité aux premières lettres de l'alphabet[17]. Là encore l'audace lexicale de Hugo se déploie là où les risques sont les moindres, c'est-à-dire dans les genres qui imposent le moins de règles : roman et récit de voyage. La plupart des hapax appartiennent au *Rhin* (couleur locale), à *Notre-Dame de Paris* (vieux mots), aux *Misérables* (argot, termes militaires, mots techniques), et surtout aux *Travailleurs de la mer* (régionalismes, termes marins). Ni la correspondance, ni le théâtre, ni la poésie n'ont de part appréciable dans l'effectif des hapax. *Hernani* n'en contient aucun, non plus que *Les Rayons et les ombres.* On n'en rencontre que deux au théâtre et 42 parmi les vers, dont *35* dans la seule *Légende*

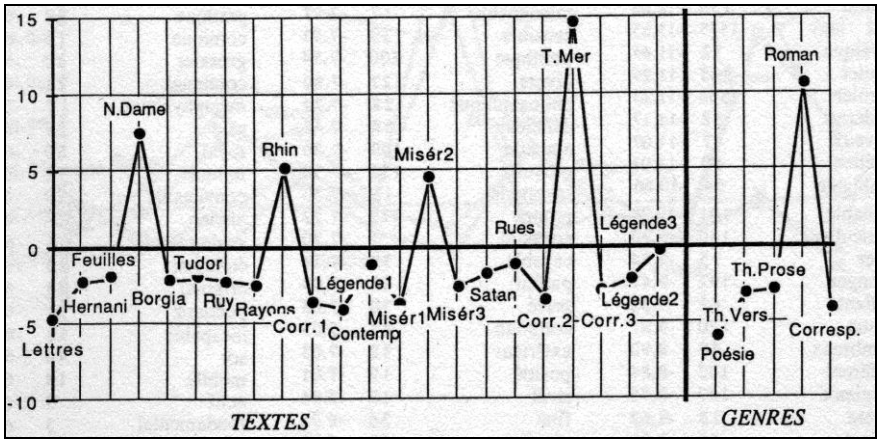
17. Les numéros renvoient aux pages de l'édition de la Pléiade, sauf pour *Le Rhin* et la Correspondance (édition Ollendorf). Les codes alphabétiques (de *a* à *g*) précisent la zone dans la page.

des siècles. Hugo a donc longtemps hésité à introduire des intrus dans la citadelle du dictionnaire. Encore ces hapax ont-ils un statut qui les assimile plutôt à des étrangers qu'aux gens du peuple. En les mêlant au discours, Hugo produit plus d'étrangeté que de familiarité. Et la poésie accepte plus volontiers le dépaysement que le déclassement. Le « mot propre » n'est donc pas nécessairement un « rustre », mais plus volontiers l'ambassadeur d'une contrée lointaine. Ainsi se justifient – nous reprenons l'extrait ci-dessus – ces termes bizarres dont Hugo parsème *La Légende des siècles : alète, andryade, aragonal, argiraspide...,* et qu'on trouverait difficilement dans les dictionnaires, même encyclopédiques[18].

mot	texte	page	zone	mot	texte	page	zone
abaca	Mer	720	e	argyronète	Mer	935	g
abax	Cor2	180	g	arimaspe	Rhin	203	a
abodrite	Rhin	115	c	arpenteuse	Mer	934	f
abreuveur	Mer	630	a	arquebusé	Misl	318	f
acarien	Rhin	161	g	artien	Dame	158	f
accompagneresse	Dame	236	c	asphatilte	F.Sat	801	e
accortement	Rhin	376	f	assolant	Cor3	192	d
aciculiforme	Rhin	162	b	assorteur	Rhin	419	b
affieffeur	Mis2	1004	a	astrologien	Dame	145	a
afflation	Mer	903	c	atigeur	Mis2	1013	f
aiguilleter	Mer	842	c	augmentable	Cor2	157	b
alète	LS2	392	e	banquetier	Mer	744	g
aise	Rhin	212	c	barant	Mis2	1011	e
alunière	Rhin	57	g	baranton	Mis2	1011	e
andryade	LS1	412	a	baraqué	Misl	123	b
antenale	Rhin	202	e	barbeyer	Mer	656	g
anthylle	Rhin	211	d	barbière	Mis2	613	d
antipapisme	Mer	687	b	baronnage	LS1	145	b
antonomase	Misl	282	a	barroter	Mer	778	f
apitage	Mer	980	a	basbourdis	Mer	656	e
apostillant	Cor3	257	f	baudroyeur	Dame	403	a
aquilège	Mer	701	e	baverolle	Dame	211	d
aragonal	LS1	154.	b	bayardant	Rhin	208	d
arcasien	Mis2	996	f	becqueteur	LS3	478	a
arcasse	Mer	670	c	becquillard	Mis2	1013	f
archegaye	Dame	399	c	belgiquois	Cor3	121	b
archichambellan	Rhin	423	c	bénéventin	Rhin	123	a
archidiaconat	Misl	54	d	bichot	Mis2	1011	b
argiraspide	LS2	198	a	bisaiguë	Mis3	1270	f
argotière	Dame	95	d	bisqué	Cor2	310	a
				etc...			

Tableau 2. Les hapax de Hugo (extrait)

18. Le plus utile pour des contrôles de ce genre est le *Grand Larousse encyclopédique du XIXe siècle.*



Graphique 2. Les hapax de Hugo dans le corpus du T.L.F.

Du mot propre, on passe naturellement au nom propre, car la frontière est parfois indécise entre les deux, comme le montre par exemple le mot *aragonal* qu'on vient de citer. Et ici on touche à un trait spécifique de l'écriture hugolienne : l'invasion généralisée des noms propres. On en dénombre 82.951 au total (il s'agit d'occurrences, non de vocables), alors que, vu la taille du corpus, on en attendrait 65.955 si la proportion était la même qu'ailleurs. L'écart réduit qui mesure cette différence atteint une valeur monstrueuse : + 67. Si on récuse le langage probabiliste, on peut exprimer le même fait par un simple rapport : un mot sur 25 est un nom propre chez Hugo, alors que la proportion est de 1 sur 31 dans l'ensemble du corpus. Chez Proust le rapport s'établit à 1 contre 30 et chez Zola à 1 contre 35. Seul Chateaubriand, avec une valeur de 1/23, l'emporte sur Hugo. Mais si l'on raisonne sur les vocables – ce qui vaut mieux présentement – Chateaubriand doit s'incliner.

	Vocabulaire	*noms*	*proportion*
Hugo	20.602	8.881	0,43
Chateaubriand	19.606	6.797	0,35
Zola	19.337	2.940	0,15
Proust	18.322	2.976	0,16

L'effectif des noms propres atteint donc presque la moitié de l'effectif des vocables communs, quand la moyenne s'établit aux alentours de 15 %. Reste à savoir si cette multiplication des individualités a quelque chose à voir avec la libération des mots, que revendique Hugo. Il est bien exact que dans l'ordre social la Révolution et le Romantisme ont conjointement oeuvré à l'exaltation de l'individu, et le large accueil

fait aux noms propres relève de cette tendance, qui explique aussi le goût de la couleur locale et des voyages dans l'espace ou dans le temps. Mais ces individualités que Hugo aime à répandre, dans sa prose comme dans ses vers, n'ont généralement rien du *vulgum pecus*, même s'il s'agit de criminels ou d'inconnus. Le seul fait de la citation vaut une promotion et l'anoblissement. Là encore l'ambiguïté demeure : cette armée de noms propres qui envahit l'oeuvre de Hugo ne brandit nullement le drapeau rouge. Le goût du nom propre est plutôt un apanage aristocratique, l'héritage d'un temps où le nom qu'on porte s'identifie à la terre qu'on a et où le nom est un titre au double sens du terme : titre de noblesse et titre de propriété. Ainsi s'explique la faveur du nom propre chez Chateaubriand, et, à un moindre degré, chez Proust. Un authentique républicain comme Zola s'en méfie au contraire et s'en éloigne. La génération romantique, née sous l'Empire, loin d'adopter l'égalitarisme démocratique, a le culte élitiste du héros et rêve de se faire un nom, un nom propre.

Les critères quantitatifs que nous venons d'utiliser laissent donc la question en suspens. Certes avec Hugo une foule de mots pénètre dans la littérature mais ils n'ont pas accès partout et, le temple littéraire restant compartimenté, on les admet difficilement dans la poésie. Et d'autre part ceux auxquels Hugo délivre un passeport ne sont peut-être pas des

drôles patibulaires
Habitant les patois ; quelques-uns aux galères
Dans l'argot ; dévoués à tous les genres bas ;
Déchirés en haillons dans les halles ; sans bas,
Sans perruque ; créés pour la prose et la farce.

Pour en décider, il ne suffit pas de compter des effectifs et d'opérer sur des classes, dans l'anonymat. Encore faut-il s'attacher à l'identité, à l'individualité des mots.

Mais la tâche est rude, s'il faut examiner un à un les 20.000 mots de Hugo et les comparer aux 70.000 unités du corpus général de l'INaLF. C'est pourtant ce que nous avons fait avec l'aide de l'ordinateur. Notre Index, pour chaque vocable, compare la fréquence observée chez Hugo à celle qu'on relève dans le corpus XIXe-XXe siècles et livre un écart réduit, positif ou négatif selon qu'on constate un excédent ou un déficit. Un second test prend pour référence les oeuvres littéraires contemporaines de Hugo, de 1815 à 1885. Et pour faire bonne mesure, on a comparé la prose de Hugo à celle de son temps, et de la même façon la poésie de Hugo à celle qui l'entoure. La place nous manque ici pour exposer le détail de cette recherche et nous renvoyons le lecteur à notre

ouvrage. Nous n'en donnerons qu'un extrait (tableau 3), relatif aux adjectifs, dont nous avons fait deux lots : ceux qui sont exceptionnellement fréquents chez Hugo, eu égard à l'époque, et ceux qui au contraire ont une fréquence anormalement basse (tableau 3-bis).

mot	**fréq.**	**écart**	**mot**	**Fréq.**	**écart**	**mot**	**fréq.**	**écart**
sombre	1198	46.50	Géant	289	23.50	bourru	51	15.14
hideux	360	37.49	Vaillant	159	21.82	splendide	164	15.05
obscur	573	33.05	Haut	1710	20.86	chère	677	14.76
charmant	1031	31.41	Hagard	109	20.27	bon	2110	14.74
farouche	319	30.56	redoutable	211	20.01	sépulcral	55	14.70
difforme	131	30.08	excellent	499	19.63	vil	167	14.61
lugubre	272	29.65	serein	191	19.51	admirable	368	14.49
effrayant	354	28.97	doux	1210	19.04	nain	85	14.47
noir	1701	28.81	mystérieux	403	18.87	vieux	2054	14.47
sinistre	331	28.24	insondable	59	18.27	affreux	399	14:24
formidable	284	27.73	altier	91	17.74	effaré	149	14.23
cher	1379	27.16	âpre	175	17.20	vivant	588	14.19
ténébreux	194	27.01	immense	688	17.17	plein	1280	14.17
pensif	251	26.96	misérable	425	16.66	énorme	427	14.16
profond	1108	26.75	morne	231	16.39	nu	352	14.11
cordial	163	26.07	joyeux	318	15.84	horrible	332	13.54
monstrueux	250	24.01	fier	392	15.41	chauve	86	13.43
auguste	196	24.12	blême	138	15.40	saint	1364	13.42

Tableau 3. Les adjectifs caractéristiques de Hugo : zone positive (des excédents)

Dans les deux cas le tri a été fait à partir de la valeur de l'écart réduit. Ainsi prenons pour exemple la tête de liste, le mot *sombre.* On en relève 1.198 emplois chez Hugo, et, compte tenu de la part du corpus Hugo dans celui de l'époque (p = 2.074.286 / 31.681.697 = 0,0655, d'où q = 1-p = 0,9345), on est en droit d'en attendre : 5.394 x 0,0655 = 353. La différence absolue est donc de : 1.198 - 353 = 845, et l'écart réduit : z = 845 / racine carré de (353 x 0,9345) = 46,50, cette valeur étant la plus forte qu'on ait enregistrée parmi les adjectifs. Aucun lecteur de Hugo ne sera surpris de cette première place non plus que du premier rang obtenu par *ombre* parmi les substantifs. Le couple *ombre-sombre* forme également la rime la plus fréquente chez Hugo. Que les chiffres s'accordent ainsi avec le sentiment du lecteur, donne quelque assurance sur leur fiabilité. Le reste de la liste a le même caractère d'évidence : les adjectifs *hideux, obscur, charmant, farouche, difforme, lugubre, effrayant, noir, sinistre, formidable,* etc., sont bien ceux qu'on a conscience d'avoir rencontrés mille fois sous la plume de Hugo. Or on ne trouvera pas un *gueux* parmi eux. Certes la tonalité d'ensemble est effrayante mais il s'agit d'un style relevé, tout à fait digne du théâtre classique. Pas un adjectif ne figure ici qui ne puisse être employé honorablement dans la tragédie de Racine, ... ou de Campistron. Sur les

54 mots relevés en tête de notre extrait, nous en avons retrouvé effectivement 44 dans le théâtre de Corneille[19]. Voici ceux qui manquent : *difforme, cordial, hagard, serein, insondable, joyeux, splendide, bourru, sépulcral, effaré.* Mais à part *effaré, insondable* et *cordial* tous ont été utilisés au théâtre avant 1800[20]. Les adjectifs préférés de Hugo n'ont donc rien de révolutionnaire. Leur allure est noble et digne et ils ne chantent pas la *Carmagnole.*

mot	fréq.	écart	mot	fréq.	écart	mot	fréq.	écart
nouveau	853	-17.22	intellectuel	16	-8.22	fort	1248	-6.87
moral	105	-15.53	ennemi	186	-8.22	personnel	70	-6.87
général	361	-14.45	scientifique	. 6	-8.04	spirituel	56	-6.85
jeune	1238	44.39	supérieur	112	-8.04	intéressant	23	-6.78
naturel	138	-12.65	raisonnable	17	-7.91	pratique	29	-6.77
seul	1579	41.55	sensible	39	-7.91	comique	12	-6.74
physique	32	-11.44	politique	300	-7.84	grossier	30	-6.72
dernier	845	-11.29	divers	127	-7.80	continuel	21	-6.66
premier	1566	-11.23	philosophique	28	-7.74	métaphysique	7	-6.58
moderne	38	-11.17	délicieux	38	-7.48	sale	28	-6.55
nerveux	7	-11.07	aimable	109	-7.36	royal	80	-6.54
chrétien	59	-11.04	moindre	152	-7.33	intime	77	-6.45
semblable	64	40.66	favorable	15	-7.30	convenable	20	-6.44
véritable	101	-10.24	propre	322	-7.22	ancien	504	-6.39
particulier	130	-9.90	constant	7	-7.20	analogue	5	-6.29
léger	75	-9.84	capable	58	-7.11	étonnant	20	-6.28
étranger	192	-9.47	parfait	83	-7.09	original	23	-6.20
brillant	42	-9.30	privé	30	-7.08	poétique	31	-6.17
connu	170	-9.27	.commun	83	-7.03	incapable	18	-6.15
nombreux	49	-8.90	extérieur	82	-7.03	sot	40	-6.15
différent	102	-8.84	positif	19	-7.01	mobile	14	-6.14
religieux	167	-8.75	rival	13	-6.96	actif	3	-6.07
légère	15	-8.62	fine	36	-6.95	fondamental	3	-6.06
ridicule	55	-8.29	régulier	20	-6.93	pieux	37	-6.01

Tableau 3-bis. Les adjectifs caractéristiques de Hugo : zone négative (des déficits)

Inversement les adjectifs dont l'emploi répugne à Hugo – et dont la liste figure au bas du même tableau – ont un caractère nettement prosaïque ou utilitaire, et appartiennent au langage de l'information, de la science, de la critique, de la politique, de la conversation. On les trouve dans les journaux, non dans un drame ou dans un recueil poétique. Si le mot *poétique s'y* trouve, c'est précisément parce que ce n'est pas un mot poétique mais un outil de la communication, comme d'autres mots en *ique* qu'on trouve pareillement dans la liste : *scientifique, politique, philosophique, métaphysique, comique, physique, pratique.*

19. En consultant l'index établi par Charles Muller dans *Le Vocabulaire du Théâtre de Pierre Corneille,* éditions Slatkine, Genève.

20. Encore le mot *cordial* – qui n'a rien en soi de spécifiquement hugolien – ne doit de figurer ici qu'à l'importance de la correspondance dans le corpus de Hugo.

On a pu faire les mêmes observations à propos des substantifs, et constater dans les listes, positive et négative, des spécificités que les préférences de Hugo en matière lexicale vont au langage soutenu[21]. On se limitera aux quelques exemples que Hugo cite dans sa *Réponse à un acte d'accusation*, et qui sont moins probants, dans les faits, que ce qu'il prétend.

	fréquence		Ecarts réduits			
	total	poésie	corpus	époque	époque prose	époque vers
albâtre	11	8	1.2	-0,3	-1,5	-2,2
neige	160	76	1,1	0,3	-2,5	-4,3
jais	14	3	4,3	2,7	3,7	-0,4
ivoire	28	13	-0,3	-1,7	-2,1	-4,4
narine	31	14	0,0	0,5	1,2	-1,9
nez	141	12	-4,7	-4,5	-3,6	-2,6
perruque	26	7	1,8	0,1	-0,7	2,2
crinière	29	18	4,5	2,3	-0,1	0,1
poire	15	1	0,6	0,6	1,1	-0,3
âne	81	41	4,3	3,3	-0,7	6,2
cochon	10	2	-5,5	-5,1	-4,8	1,1
vache	42	8	-2,5	-1,2	-1,2	-0,6
génisse	6	3	0,8	0,8	0,6	-1,5

S'il est vrai que Hugo répand moins souvent que les versificateurs de son temps l'*albâtre*, la *neige* et *l'ivoire*, les écarts cessent d'être significatifs – et même cessent souvent d'être négatifs – quand la comparaison est faite plus largement avec l'époque, tous genres confondus, ou avec le corpus entier. Il n'y a pas trace d'une proscription systématique. Et cela est vrai pour la *perruque*, la *narine*, le *jais* ou la *génisse*, dont les exemples ne sont pas rares sous la plume de Hugo. Inversement les mots auxquels Hugo pense avoir donné droit de cité, à savoir le *nez*, le *cochon* et la *vache* se rangent du côté des déficits[22]. Ainsi donc, quoi qu'il en dise, Hugo est relativement plus accueillant à la *narine* qu'au *nez*, à la *génisse* qu'à la *vache*. Et quand il prétend avoir nommé le cochon par son nom, cela n'est vrai qu'au moment où il l'écrit,

21. On a même neutralisé le genre littéraire en limitant la comparaison au seul corpus en vers. Le vers hugolien ne se détache nullement par quelque audace plébéienne du vocabulaire mais par une surenchère, un renforcement de la spécificité, et presque de l'exterritorialité, poétique.

22. Le mot *crinière* que Hugo oppose peu légitimement à *perruque* est en réalité un mot noble et poétique, quand il est appliqué au lion ou au cheval.

car le mot *cochon* n'apparaît pas dans notre corpus avant ce texte, sinon dans le *Rhin*[23] Quant au mot *poire*, il n'y a pas d'autre exemple « poétique » qu'on puisse citer de ce fruit sous la plume de Hugo[24].

Mais, dira-t-on, est-ce bien légitime de comparer Hugo à son temps, alors que lui-même fait référence au passé ? Il n'y a de révolution que s'il y a un ancien régime, et dès lors c'est cet état antérieur qu'il faut considérer. Malheureusement il est difficile de remonter au-delà de 1789, car les données de l'INaLF n'ont pas été homogénéisées et lemmatisées en ce qui concerne la littérature antérieure à la Révolution. Nous avons pourtant tenté l'expérience, au moins pour la liste précédente, en procédant aux réajustements nécessaires, c'est-à-dire en regroupant les formes du singulier et du pluriel. Et pour empêcher l'interférence du genre, nous avons borné notre enquête aux seuls textes poétiques (chez Hugo aussi, naturellement). On a recensé ainsi 119 textes en vers, dont 31 d'ailleurs ne contenaient aucun mot de la liste. Ce corpus est très comparable à celui de Hugo en étendue et ne le dépasse que de 13 %. Aussi par simplification, nous considérerons les deux ensembles comme égaux, pour éviter les calculs et les décimales. Voici les résultats :

mot	corpus	Hugo	mot	corpus	Hugo	mot	corpus	Hugo
albâtre	35	8	*âne*	73	41	*cochon*	10	2
crinière	30	18	*génisse*	28	3	*ivoire*	33	13
jais	1	3	*narine*	10	14	*neige*	141	76
nez	303	12	*perruque*	23	7	*poire*	9	1
vache	63	8						

Pour 4 éléments, les affirmations de Hugo sont confirmées : il y a dans ses vers moins *d'albâtre*[25], moins *d'ivoire*, moins de *perruques* et

23. Journet et Robert avaient vainement cherché ce *cochon* dans l'oeuvre poétique de Hugo sans le trouver. Pour être exact, le cochon réapparaît une fois dans un vers de Hugo, quelque part dans *La Légende des siècles*, mais non dans la pièce *Sultan Mourad*, dont le héros, un cochon, n'est désigné que par les termes *pourceau, porc fétide, bête impure* ou *lépreuse, être difforme, abject* ou *immonde.*

24. Même absence des *toutous* et des *papas* que Hugo n'introduit dans ce texte que par pure provocation et sans donner lieu à récidive.

25. Voici les huit contextes poétiques – et très classiques – où resplendit *l'albâtre :*

- Et sur son cou *d'albâtre /*Les nue avait moins d'escarboucles luisant/Que ces fauves rochers n'ont de flaques de sang (*La Légende des siècles. Le petit roi de Galice*).

moins de *génisses.* Mais pour les 9 autres, elles ne sont pas vérifiées, soit que les dénominations nobles soient en surplus chez Hugo (*jais, narine*), soit que les mots familiers ou déclassés obtiennent un classement plus mauvais encore dans le vers hugolien (*âne, cochon, nez, poire, vache*). Dans le domaine lexical, Hugo se croit donc – ou se veut – plus iconoclaste qu'il n'est en fait.

Mais on peut n'avoir pas tort, même en fournissant deux ou trois mauvais exemples. En citant malencontreusement le cochon et la vache, Hugo veut sans doute suggérer – et peut-être à bon droit – que la gent animale est particulièrement victime des exclusives classiques et que pour être complète la Révolution doit libérer le peuple des bêtes (au moins au niveau des mots). Il serait donc judicieux d'élargir l'enquête au bestiaire dans son ensemble, ou du moins dans ses espèces les plus communes (nous en avons choisi 50). Le romantisme, sensible aux thèses de Rousseau, a célébré l'innocence des bêtes et la vertu des plantes et des simples. Et Hugo, en défendant le crapaud contre les garnements et le cochon contre les poètes trop délicats, participe à ce retour à la nature. Mais, même s'il est capable de braver les interdits et d'approcher les intouchables, un crapaud, ou une araignée, sa plume aime à caresser les animaux les plus nobles que la création (ou l'imagination) ait enfantés : l'*aigle,* le *lion,* le *tigre,* le *loup,* l'*ours,* le *dragon,* le *cygne.* Les seigneurs prennent le pas dans les graphiques 3 sur les animaux domestiques[26].

- Le granit est la brute informe de la nuit, */L'albâtre* ne sait pas que l'aube existe et luit, /Le porphyre est aveugle et le marbre est stupide ; (*La Légende des siècles. Les 7 merveil. Temple Ephèse*).

- Le marbre obscur s'émeut dans la nuit infinie / Sous la parenté sombre et sainte du génie, / Et l'albâtre enfoui ne veut plus être noir (*La Légende des siècles. Les 7 merveil. Temple Ephèse*).

- Rien ne peut te ternir, ô pur *albâtre* (*La Légende des siècles. L'amour, VI. En Grèce !*).

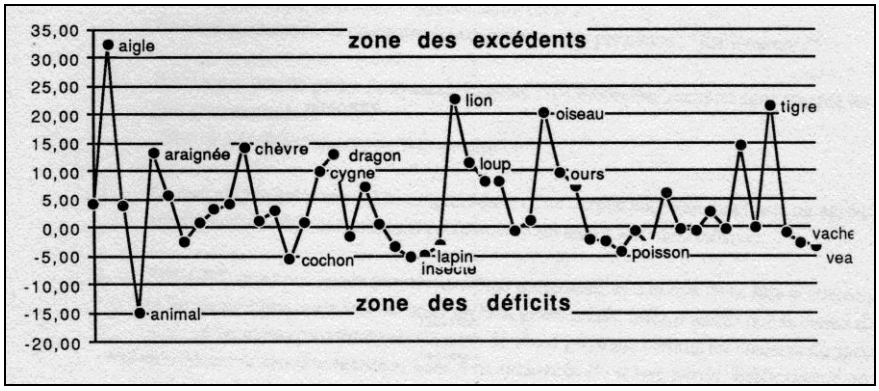
- Notre *albâtre* vaut le charbon (*Les Chansons des rues. Silhouettes 2 écrit en 1827*)

- Je massacrai *l'albâtre,* et la neige, et l'ivoire (*Les Contemplations. Réponse à Accusation*).

- Du haut du temple ou du théâtre, /Colosse de bronze ou *d'albâtre,* /Salué d'un peuple idolâtre, /Je surgirais sur la cité (*Les Feuilles d'automne. A* M. *David, statuaire*).

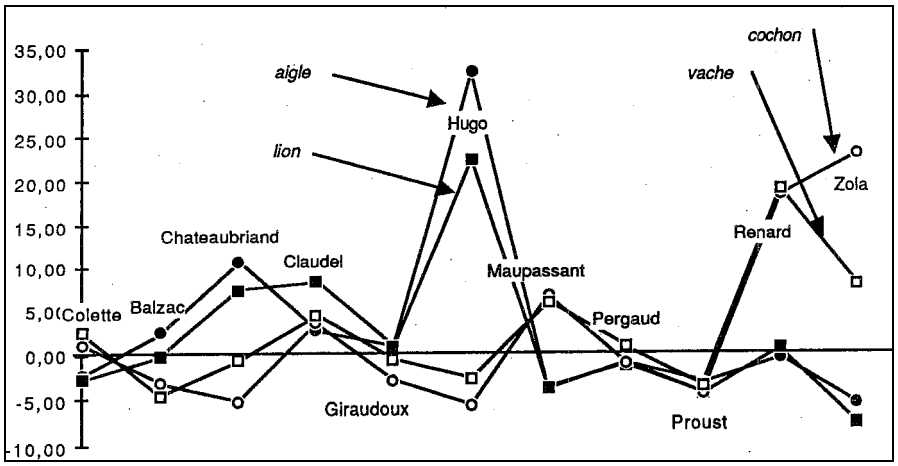
- Il faut que, sous le ciel, de soleil inondée,... /La figure de pierre, ou de cuivre, *ou d'albâtre,* /Porte divinement sur son front calme et fier /La beauté, ce rayon, la gloire, cet éclair ! (*Les Rayons et les ombres. Au statuaire David*).

26. Le graphique reposant sur des écarts réduits a été obtenu en comparant le bestiaire de onze écrivains, dont certains sont des animaliers patentés : Balzac, Chateaubriand, Claudel, Colette, Hugo, Maupassant, Pergaud, Proust, Renard et Zola, soit un ensemble de 16 millions de mots.



Graphique 3. Le Bestiaire de Hugo comparé au corpus XIXe-XXe

La société animale paraît donc régie par des clivages anthropomorphiques qui sont le reflet des structures humaines. En particulier il est des animaux nobles, comme il est des manants. Et il semble bien que la révolution chez les bêtes ait suivi d'assez loin celle de 1789 chez les hommes. Le graphique 4 met en relief la persistance dans la première partie du XIXe siècle de la ségrégation animale.



Graphique 4. *Nobles* et *rustres*

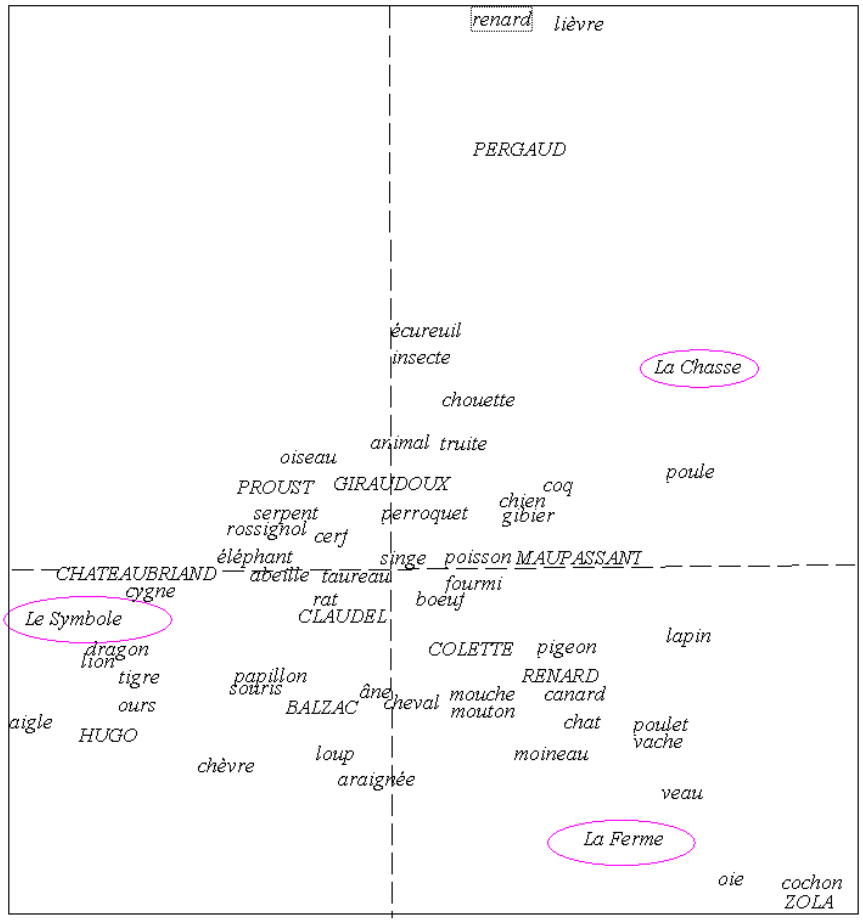
Chateaubriand, Balzac, et surtout Hugo accordent le premier rang au *lion* et à l'*aigle*, représentants de la noblesse et donnent rarement la parole à la *vache* et au *cochon*, délégués rustauds du tiers état. Ceux-ci prennent leur revanche à la fin du siècle, quand le réalisme et plus encore

le naturalisme renversent les barrières. Conduits par Maupassant, bovins et porcins envahissent alors la littérature, tandis que se retire le roi des animaux, suivi de sa cour. Au XXe siècle la cause est entendue et l'égalité sociale acceptée : ni Colette, ni Giraudoux, ni Claudel, ni Proust ne font la différence entre animaux nobles et rustres.

L'analyse factorielle représentée dans le graphique 5 rend compte de cette ségrégation parmi les bêtes et ceux qui en parlent. On y voit nettement une ligne de partage entre la droite et la gauche, c'est-à-dire entre une vision poétique et une description réaliste du règne animal. D'un côté Hugo, approuvé par Chateaubriand, et plus mollement par Balzac, Claudel et Proust, s'installe dans un lieu poétique où il accueille de préférence les animaux qui disposent de la puissance, de la grandeur, de la beauté, de la liberté ou qui, à tout le moins, sont porteurs d'une de ces vertus, comme l'*abeille*, symbole du travail, l'éléphant (la force), le *cygne* (la pureté), le *rossignol* (le chant de la nature), le *papillon* (la joie de vivre). Le symbole est ici si éloigné du réalisme qu'il admet des êtres irréels, comme le *dragon*, et que sa préférence va aux animaux les plus extraordinaires et les plus doués de personnalité, même si les traits sont accusés jusqu'à la sauvagerie. Les animaux féroces, comme le *lion*, le *tigre*, l'*ours*, le *loup*, l'aigle sont ici chez eux, en liberté. Car la poésie semble peu compatible avec la condition servile et elle repousse systématiquement les animaux domestiques. Seule la *chèvre* d'Esméralda s'aventure ici par humeur vagabonde (mais c'est une bête du cirque, non de la ferme). Le *taureau*, moins avachi que les autres, a également forcé le passage.

De l'autre côté, dans le camp des domestiques, Zola fait la loi, sur la marge droite du graphique, entouré d'une *oie* et d'un *cochon*. Jules Renard est dans la proximité, et plus loin Maupassant et Pergaud. C'est donc là que se sont installés les réalistes et les naturalistes. Et l'on voit bien qu'ils ont pris leurs modèles dans la réalité immédiate, et particulièrement à la ferme. Dans la zone d'influence de Zola et de Renard (quadrant inférieur droit), c'est toute la ferme en effet qui est reconstituée : avec l'étable (*veau, vache, boeuf*), l'écurie (*cheval*), la bergerie (*mouton*), la basse-cour (*poulet, canard, lapin*). Une inquiétude pourtant : où donc est passé le *chien ?* Le chien a déserté la ferme pour suivre les chasseurs, de l'autre côté de l'axe des x. C'est que le dernier quadrant, en haut à gauche, est celui des bois et de leurs habitants (*écureuil, chouette, renard*), du gibier (*gibier, lièvre, cerf*) et des chasseurs : Maupassant, Pergaud et Giraudoux... Ainsi les clivages sont

nets : le premier facteur sépare le symbole et la réalité, et dans la réalité le second facteur intervient pour distinguer la chasse et l'élevage.



Graphique 5. Analyse factorielle des animaux (facteurs 1 et 2)

Hugo se trouve donc en deçà de la Révolution chez les bêtes, comme Rousseau est en deçà de la Révolution chez les hommes. On pourrait poursuivre l'enquête dans d'autres champs sémantiques que traverse la même barrière entre le grand et le petit, par exemple dans la représentation que donnent les noms de métier et les étiquettes sociales. Là aussi Hugo est attiré par les grandes figures symboliques, qui peuvent être des bandits, des galériens ou des saints. Ce n'est pas encore le menu peuple que Hugo nous montre, mais un peuple transfiguré, comme l'épagneul qu'il déguise en lion :

Avec des épagneuls nous fîmes des lions.
(*Réponse à un acte d'accusation*)

Dans tout ce qu'il fait, Hugo agit *lionneusement*, pour reprendre l'expression de Marot. On a vu des lions comme Mirabeau et Danton dans les révolutions. Dans le domaine des mots, Hugo tient une position analogue : de grands gestes, des déclarations enflammées et des échos longuement prolongés. Mais en agitant – violemment et tardivement – le drapeau rouge, Hugo et le romantisme n'ont fait que préparer les réformes qui ouvraient la voie au réalisme, au naturalisme et aux mouvements plus radicaux encore.