



HAL
open science

L'exploitation des grands corpus: Le bestiaire de la littérature française

Étienne Brunet

► **To cite this version:**

Étienne Brunet. L'exploitation des grands corpus: Le bestiaire de la littérature française. *Literary and Linguistic Computing*, 1989, 4 (1), pp.121-134. 10.1093/lc/4.2.121 . hal-01574221

HAL Id: hal-01574221

<https://hal.science/hal-01574221>

Submitted on 12 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.

L'exploitation des grands corpus : le bestiaire de la littérature française¹

Etienne Brunet

L'application des méthodes quantitatives à la littérature a suscité des espoirs, parfois déçus, des prétentions souvent trompeuses, et des renoncements, sans doute prématurés. Les raisons qui expliquent cet échec relatif tiennent au déséquilibre plusieurs fois constaté entre des ambitions trop vastes et des données trop étroites. Certains ont voulu aborder, non sans imprudence, des problèmes complexes de datation ou d'attribution de textes, en cherchant à s'appuyer sur des garanties scientifiques, l'objectif étant d'asseoir l'intime conviction du critique sur des preuves matérielles, comme celle des juges et des jurés l'est au tribunal. Mais où trouver les empreintes digitales d'un écrivain ? A quels témoignages sûrs peut-on reconnaître la griffe de Voltaire ou la plume de Flaubert ? Et la recherche de s'égarer à la poursuite des indices, sans obtenir jamais la trace des empreintes digitales. Car ces indices varient d'une recherche à l'autre, sans atteindre le niveau de généralité et de constance qu'on attend d'un instrument de mesure. Et parfois même, par une pétition de principe assez contestable, les indices ont été retenus comme probants dans la mesure où ils étaient discriminants et donnaient lieu à des écarts importants. Mais même lorsqu'il n'est pas partiel, le choix d'un indice reste un critère trop partiel, s'il ne procède pas d'un dépouillement exhaustif et d'une enquête élargie. Car le témoignage de la statistique n'est solide que dans les grands nombres. Sa valeur est collective, comme celle des portraits-robots, et non individuelle, comme celle des empreintes digitales. Or faute de moyens, les enquêtes réalisées jusqu'ici ont rarement disposé de données suffisamment amples et

1. Article publié dans *Literary and linguistic computing*, vol.4, n° 2, 1989, Oxford, p. 121-134. Seule la mise en page a été revue et corrigée.

homogènes pour permettre l'accumulation et le recouvrement des témoignages et faire apparaître dans l'entrelacs brouillé des lignes et des traits une figure reconnaissable, un portrait-robot.

Mais le temps est venu où l'ampleur des données répond à l'ambition des objectifs, au moins dans le domaine français. On sait que l'entreprise du *Trésor de la langue française* a procédé au dépouillement des principaux textes de notre littérature nationale. Or ces données amassées depuis vingt ans sont désormais constituées en une véritable base qui peut être consultée à toute heure par la voie télématique et qui est disponible à tout chercheur universitaire pourvu d'un terminal. Nul besoin de demande écrite, et plus de ces retards et de ces malentendus propres aux prestations de service. Le chercheur, au bout de sa ligne téléphonique, peut pêcher sans attendre n'importe lequel des 150 millions de mots du grand corpus, s'intéresser à une expression, à un écrivain, à un genre littéraire, à une époque, ou à tel ensemble de textes qu'il précise à sa guise. Et la réponse lui est donnée immédiatement, sous la forme choisie, qui peut être une fréquence, un index de références ou même un contexte d'une ou plusieurs phrases entourant le mot pivot. Muni de tels matériaux le chercheur est libre d'employer les méthodes traditionnelles, il a seulement gagné du temps en constituant un fichier complet (quoique impur²) de références. Nous emprunterons une autre voie qui fait confiance non seulement aux ressources de l'informatique mais aussi aux méthodes de la **statistique**. Ces méthodes, qui sont toujours comparatives, permettent de comparer les uns aux autres des mots, des textes et plus précisément l'usage différencié des mots dans les textes. Les données de base prennent alors la forme d'un tableau à deux dimensions, où la cellule (i,j) contient la fréquence du mot i dans le texte j .

Il suffit donc de choisir les mots (les lignes du tableau) et les textes ou ensembles de textes (les colonnes). Le logiciel STELLA, créé par Jacques Dendien, permet très aisément de dresser la liste des premiers (par la commande *listemots*) et des seconds (par les commandes *biblio*, *listerefs* et *defcorpus*), avant que la commande *frequence* ne soit lancée qui met en rapport les deux listes et établit le tableau à exploiter.

2. Nous disons *impur* parce que les problèmes de lemmatisation n'ont été réglés que de façon grossière et expéditive. Mais l'index permet précisément au chercheur de faire le tri parmi les références d'un mot homographe. Quant au traitement statistique l'approximation n'est pas trop gênante, pourvu que les normes restent constantes.

Nous choisirons à titre de démonstration un cas facile et même trivial, où le doute n'est guère permis. Et nous nous proposerons de vérifier si Colette aime les chats et plus largement les bêtes, ou du moins si son vocabulaire, par quelque intempérance lexicale, porte témoignage de cet amour des bêtes. Il est peu probable que Colette ait eu beaucoup de goût pour les mesures et pour les comptes. Passe encore de compter les sous. Mais compter les mots, au lieu de les peser! Compter les bêtes, au lieu de les caresser! Ce projet insensé n'a sans doute jamais effleuré l'esprit de Colette, même dans les soirées agitées où le décompte des moutons aide, paraît-il, à trouver le sommeil. Mais pour rassurer Colette et épargner à son ombre le cauchemar des nombres, la cuisine des chiffres ne sera pas montrée³. Il suffit de savoir que les courbes qu'on va produire s'obtiennent avec des règles de trois, des racines carrées et bien d'autres ingrédients dont le nom seul peut dissuader l'appétit et que l'ordinateur digère sans broncher.

a – Au reste, dans le cas de Colette, les calculs n'ont rien de monstrueux car ils ont porté sur des données relativement étroites. Car dans cette banque de données littéraires qui est la plus importante du monde et qui compte plus de 150 millions de mots et plus de 3000 textes, la part de Colette est assez mince puisqu'on n'a retenu d'elle que **cinq titres** : *Claudine à l'école*, *Dialogues de Bêtes*, *La Maison de Claudine*, *La Naissance du jour*, et *Sido*.⁴ On admettra volontiers qu'un tel corpus conviendrait assez peu pour étudier par exemple le mal d'aimer, mais son exploitation se justifie ici puisque le thème que nous abordons s'accorde avec la figure centrale du cycle. Car il s'agit bien là d'un cycle, qui est ordonné autour de Sido et de la nature, et dont le lien est plus organique que la ficelle trop commerciale qui attache la série des *Claudine*. Sido est tout à la fois la source, le modèle, le refuge. Et Colette en vient à s'identifier à elle, et à cultiver comme un héritage tous les traits moraux qu'elle a reçus de sa mère, et tout particulièrement l'amour des plantes et des bêtes :

3. Ceux qui voudraient forcer la porte de la cuisine trouveront une excellente initiation dans les ouvrages de Charles Muller : *Initiation aux méthodes de la statistique linguistique*, Hachette Université, 1973, 188 p., et *Principes et méthodes de statistique lexicale*, Hachette Université, 1977, 206 p. Voir aussi l'ensemble de la collection des *Travaux de linguistique quantitative*, publiée par les Editions Slatkine, Genève.

4. Quelques-uns de ces textes ont été écrits à Saint Tropez. Pourquoi Saint Tropez inspire-t-il si profondément l'amour des animaux ou bien pourquoi les défenseurs des bêtes choisissent-ils ce coin, quelque temps conservé, du paradis perdu ? A défaut de Colette, la question pourrait être posée à Brigitte Bardot.

*Dans le coeur, dans les lettres de ma mère
étaient lisibles l'amour, le respect des créatures
vivantes. Je sais donc où situer la source de ma
vocation. (La Naissance du jour, Flammarion,
1984, p. 65).*

b – S'il s'agit bien d'une *vocation*, d'une préoccupation consciente et constante, on doit s'attendre à une invasion des animaux dans le vocabulaire de Colette. Jamais en effet on n'en a tant trouvés dans notre littérature depuis le *Roman de Renart* jusqu'aux romans de J. Renard. Et cela apparaît à la seule lecture des titres : *Dialogues de Bêtes*, *Le Coeur des Bêtes*, *La Paix chez les Bêtes*, *Ma Mère et les Bêtes* (Colette préfère les bêtes aux animaux), *La Chatte* (ce titre a servi deux fois : pour un roman et pour un ballet), *Splendeur des Papillons*, sans parler des noms propres *La Toutouque* ou *Prrou*, *Poucette et quelques autres*. Et Colette a dû mettre de l'ordre dans son *arche de Noé* (le mot est d'elle), en parquant la gent féline dans un recueil spécifique : *Chats* et la gent canine dans un autre : *Chiens*⁵. Nul besoin de l'ordinateur pour ce relevé, déjà fort significatif. Mais dans le texte même de Colette mille animaux se cachent ou se pavent, comme dans les tapisseries des mille fleurs ou dans ces devinettes d'enfant qui invitent à découvrir un animal tapi dans le lacs des branches ou des nuages. Il n'est pas impossible d'établir à la main le décompte et nous avons fait ainsi pour *Claudine à l'école*⁶. Mais

5. Chez Albin Michel, en 1949 et 1950 respectivement. Le même éditeur devait poursuivre ce ménage dans la ménagerie de Colette, en proposant en 1958 *Bêtes libres et prisonnières de Colette*.

6. Ce sondage partiel suffit à mettre en lumière un trait constant de Colette, qui est de prendre pour référence le monde animal, lors même qu'elle parle des hommes. La technique descriptive pour rendre compte d'un visage, d'une expression, d'un sentiment consiste à faire la transposition dans le monde des bêtes. La petite Aimée Lanthenay *aime le bien-être comme une chatte qui a froid* (Albin Michel, *Les « Claudine »*, 1982, p. 38), cette *jolie oiselle* (p. 40) *regarde de ses yeux de chat* (p. 45), la *chatte Lanthenay* (p. 54) a une *allure assurée et coquette de chatte gâtée* (p. 82), ses yeux *brasillent de joie comme ceux d'une chatte en malice* (p. 98), elle soupire *paresseuse comme une chatte qu'elle est*. Et sa soeur Luce partage avec elle la même félinité : *elle rit de ses yeux verts pareils, tout-à-fait, aux yeux de Fanchette, ma belle chatte* (p. 73), *chatte inférieure* (p. 102), *chatte énervée* (p. 105), cette *petite bête vicieuse, miaulante et griffante* (p. 108) est *toujours chatte et frôleuse* (p. 143) au moment où la narratrice prend congé d'elle : *adieu, féline petite Luce* (dernière phrase du roman). Et Colette ne se contente pas des chats : la même Luce est ailleurs un *petit singe* (p. 73), un *angora blanc* (p. 172) ; tel autre personnage *tressaute comme un cheval* (p. 133), a la naïveté d'un *gros hanneton* (p. 51), *lève ses yeux de veau* (p. 137), ou *ses yeux brillants d'oiseau* (p. 145), ou *ses yeux d'animal méchant* (p. 157), ou *ses yeux de biche ahurie* (p. 52). En réalité ce procédé n'est que l'inversion de

le recours au traitement automatique offre plus de garanties d'exhaustivité, de fiabilité et d'objectivité.

Encore faut-il dresser la **liste des animaux** et la communiquer à l'ordinateur qui n'ayant aucun sens du sens des mots est tout-à-fait incapable de distinguer dans la création ce qui vit et se meut. Comme il nous est arrivé d'approcher le bestiaire de Giraudoux⁷, l'idée nous est venue d'emprunter la liste établie chez cet écrivain afin de justifier la comparaison des deux bestiaires. Giraudoux lui-même nous engageait dans cette voie, puisqu'il reconnaît en Colette le premier écrivain animalier de notre littérature : *Colette, qui marche pieds et âme nus, a eu la première près d'elle de vrais chiens et de vrais chats. (Littérature, La Bête et l'écrivain, Grasset, p. 190)*. Cette liste, quoique corrigée par celle que nous avons établie pour l'ensemble du corpus XIX^e-XX^e siècles du T.L.F., doit à Giraudoux quelques passe-droits, souvent ratifiés par Colette, comme la promotion du *perroquet* ou de l'*écureuil*, le repêchage de la *truite*, ou la lumière jetée sur la *chouette*. Avertissons d'emblée que nulle désambiguïsation n'a été faite et que l'emploi métaphorique n'a pas été distingué des autres emplois. Il peut ainsi se faire que la *chatte* désigne un objet qui n'est pas l'animal et qui, comme la fameuse fissure visitée par Dutertre, se rencontre parfois *dans les endroits les plus abrités du soleil*⁸ (même à Saint Tropez, s'il faut en croire les gendarmes !). De même les *dragons* non seulement sont de faux animaux mais il arrive que ce soit de vrais hommes, dans le sens militaire, et les *dragonnes* sont parfois leurs maîtresses, ou leurs *souris*. Et ces mêmes *souris* ont l'infortune supplémentaire de se confondre avec le verbe *sourire*, comme les *rates* avec le verbe *rater*. Il n'était pas difficile de procéder au tri des homographes en ce qui concerne Colette, mais fort malaisé d'étendre cette mesure au grand corpus du T.L.F. Aussi y avons-nous renoncé, afin d'assurer la compatibilité et la comparaison.

l'anthropomorphisme habituel aux fabulistes. Au lieu de montrer des animaux en livrée d'hommes, Colette nous fait voir des humains déguisés en bêtes, utilisant une technique assez courante chez les caricaturistes dont l'illustration la plus récente se trouve dans l'album *Ces animaux qui nous gouvernent*. Si Colette n'abandonne jamais cette caractéristique (il suffit de lire sa correspondance, où les mots de tendresse sont des noms de bêtes), elle va épurer les rapports entre l'homme et l'animal pour en faire une véritable relation affective et non cette corrélation logique et symbolique qui n'est qu'une comparaison.

⁷ Voir notre *Vocabulaire de Jean Giraudoux*, p. 529-535, Slatkine, Genève.

⁸ *Les Claudine*, Albin Michel, p. 70.

c – Comme les textes comparés n'ont pas la même étendue les fréquences absolues ne permettent pas une interprétation directe : les 11 occurrences d'*oiseau* dans *Claudine à l'école* ne font pas contrepoids aux 11 autres qu'on rencontre dans *Dialogues de Bêtes*, puisque ce dernier texte est quatre fois plus court.

	Claudine	Dialogues	Maison	Naissances	Sido	TOTAL
abeille	-0,54	-0,66	5,22	3,45	-0,66	3,18
âne	-0,68	0,71	-1,03	0,04	-0,71	-0,89
animal	-1,45	-0,88	-1,06	-0,06	-0,89	-1,99
araignée	0,97	-0,51	11,28	-0,71	1,45	5,82
boeuf	-0,46	-0,84	-0,42	-1,17	-0,84	-1,52
chat	10,13	53,01	42,09	26,69	16,00	59,44
cheval	-2,33	5,78	0,65	1,02	0,39	1,24
chèvre	-0,35	1,07	1,38	0,36	-0,60	0,72
chien	-0,41	14,09	15,04	2,55	8,85	15,11
chouette	-0,62	2,85	1,72	-0,44	2,85	2,06
cochon	0,89	0,80	1,00	0,12	-0,68	1,10
dragon	-1,08	1,32	0,45	-0,76	-0,55	-0,56
fourmi	0,20	1,72	-0,67	-0,64	-0,46	-0,07
gibier	-0,89	-0,45	-0,66	2,58	-0,45	-0,01
insecte	-1,68	-0,85	-0,44	0,52	1,51	-0,82
lapin	0,54	6,17	4,87	-0,82	-0,59	4,00
lièvre	0,07	3,59	2,09	0,80	3,59	3,65
loup	1,48	-0,81	1,31	-1,13	2,85	1,69
mouche	2,49	8,52	5,17	-0,10	-0,76	6,36
mouton	0,00	-0,71	4,67	-1,00	-0,71	1,28
oie	2,46	-0,45	2,35	-0,62	-0,45	2,09
oiseau	-0,07	4,81	1,54	1,90	0,08	3,05
ours	0,48	-0,60	-0,88	-0,84	-0,60	-0,86
papillon	2,39	-0,57	1,53	4,20	4,65	5,35
poisson	-1,97	2,04	-0,09	-0,65	0,01	-0,91
poule	1,07	5,52	2,21	3,58	-0,65	4,82
poulet	-0,95	16,20	0,73	5,37	1,61	7,75
ra	-1,20	4,32	3,61	0,34	-0,61	2,26
renard	-0,97	-0,49	2,05	0,80	5,63	2,33
rossignol	-0,96	-0,48	2,12	-0,67	-0,48	-0,22
serpent	0,52	0,59	1,61	0,87	1,89	2,24
singe	0,37	-0,62	1,25	-0,87	-0,62	0,04
souris	5,11	5,50	3,35	2,19	-0,50	7,30
tigre	-0,99	1,50	0,63	0,75	-0,50	0,33
vache	0,42	1,81	3,25	-1,08	1,81	2,44
veau	1,14	1,61	6,37	-0,67	1,61	4,38
total	1,04	24,51	20,58	8,28	7,50	23,98

Tableau 1. Le bestiaire de Colette (écarts réduits par rapport au corpus XIX^e-XX^e)

Les fréquences doivent donc être pondérées. Elles pourraient l'être par de simples pourcentages. Mais nous avons préféré une mesure plus fine, plus lisible et plus probante qui est celle de l'écart réduit⁹. Les écarts sont calculés en prenant pour « norme » le grand corpus du T.L.F. Pour apprécier la vraisemblance sinon l'exactitude du tableau 1, le lecteur familier de Colette portera son attention sur la ligne consacrée au *chat*, où la prévision est aisée. Le calcul confirme, avec des écarts énormes (respectivement 10, 53, 42, 26, 15), une prédilection qui n'a échappé à personne¹⁰. Plus globalement le tableau révèle que la majorité des écarts sont positifs, surtout sur la marge droite qui compare le corpus de Colette au corpus total et sur laquelle nous reviendrons. C'est dire que la faveur de Colette s'étend à l'ensemble du monde animal, même si le mot *animal* est lui-même déficitaire (on a dû écarter le mot *bête* comme étant trop ambigu, du fait de l'adjectif). Mais avant de comparer Colette aux autres écrivains, tâchons de la comparer à elle-même en observant les variations de l'écart pour le même mot, par exemple le premier, l'*abeille*, qui n'est en excédent que dans la *Naissance* et la *Maison*. La brusque invasion des *araignées* dans la *Maison*, et des *papillons* dans la *Naissance* et dans *Sido*, le grand intérêt porté aux *mouches*, au *lapin*, aux *oiseaux* et au *poulet* – c'est-à-dire à la nourriture – dans les *Dialogues* du chien et du chat, sont autant d'indications fournies par ce tableau. Mais la lecture linéaire n'en est pas facile et une synthèse en a été faite dans le

9 En voici la formule :

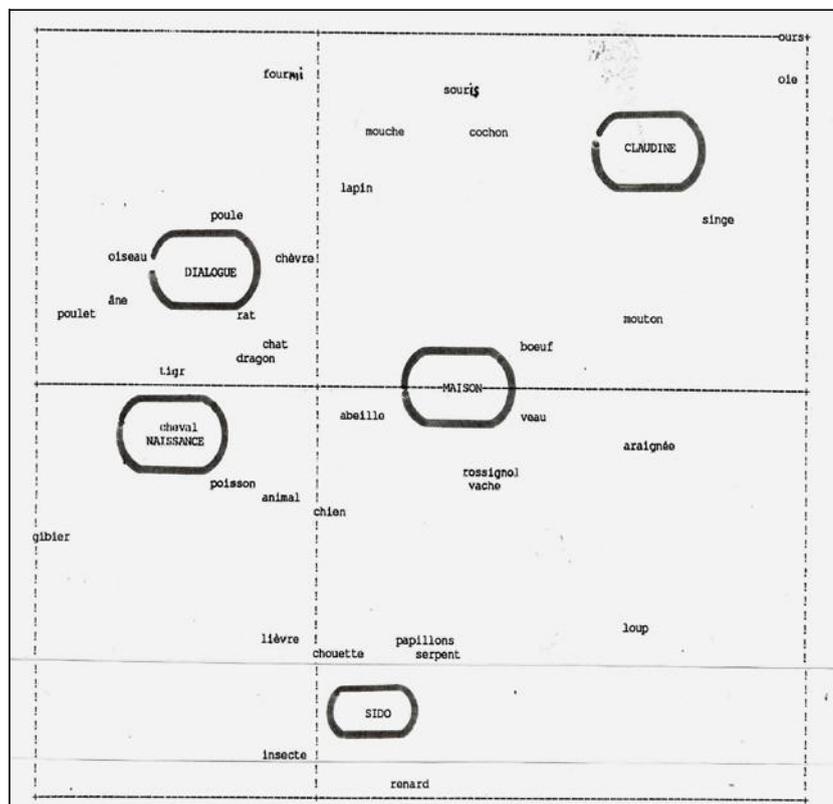
$$z = \frac{o - (f * p)}{\sqrt{(f * p * q)}}$$

pour p = proportion du texte dans le corpus, $q = 1 - p$, f = fréquence du mot dans le corpus, o = fréquence observée du mot dans le texte.

Étendue respective des cinq textes considérés : *Claudine à l'école* 96234 mots (ou signes), *Dialogues* 24.478, *La Maison* 52.910, *La Naissance* 47.380, *Sido* 24.601. La fréquence des noms d'animaux dans le corpus est donnée dans le tableau 2.

10. Il n'est peut-être pas opportun d'opposer les jugements qualitatifs et les observations quantitatives, puisque les premiers se fondent souvent sur les secondes et parfois s'y confondent. Quand nous disons d'un mot ou d'un thème qu'il est typique ou caractéristique d'un auteur, c'est une somme d'empreintes, et en dernier ressort une statistique implicite, qui nous autorise à le faire, avec cependant quelques chances d'erreurs. Et l'on pourrait dire du discours ce que Proust dit du temps, que nous mesurons à chaque instant, même approximativement : « Sans doute on se trompe souvent dans cette évaluation, mais qu'on ait cru pouvoir la faire signifie qu'on concevait l'âge comme mesurable. » (*Le temps retrouvé*, la Pléiade, p. 1046). Ainsi le discours est-il senti confusément comme mesurable, ce qui justifie notre approche.

graphique 1 qui reproduit le résultat de ce qu'on appelle une « analyse factorielle ».



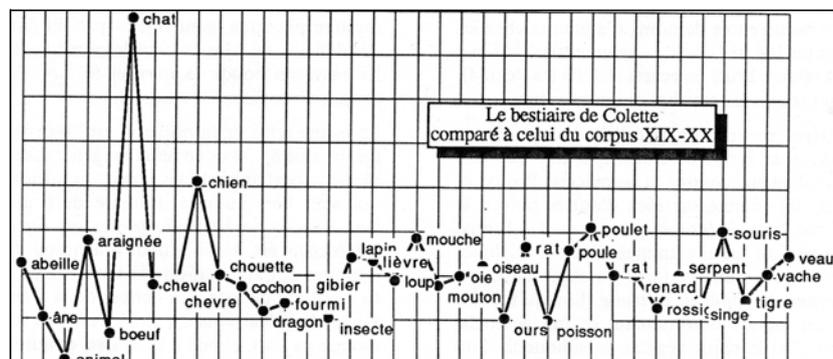
Graphique 1. Analyse factorielle du bestiaire de Colette

Le calcul a ici réparti les textes et parqué les animaux aux quatre coins du graphique, selon le principe : qui s'assemble se ressemble¹¹. Un premier enseignement peut être tiré de la succession des textes qui parcourt le graphique de haut en bas et qui reproduit fidèlement la suite chronologique. Il semble donc y avoir une évolution dans le sentiment de Colette à l'égard des bêtes, mais la « fourchette » de l'observation est un peu trop étroite pour qu'on en soit parfaitement assuré. Le second enseignement, par dessus la distance chronologique, est l'air de famille

11. La réalité du calcul est plus complexe, et l'interprétation, pour être sûre, doit prendre certaines précautions, à l'aide d'outils numériques qui accompagnent le graphique et dont le détail serait ici encombrant.

qui réunit, à droite du graphique, les textes relatifs à la maison natale : *Claudine*, la *Maison* et *Sido*. Le même paysage et le même village célébrés dans ces trois textes impliquent une même flore et une même faune, alors que l'univers intérieur et parisien qui est celui de *Dialogues* exclut nécessairement les animaux de l'étable et ceux des grands espaces. Et c'est pourquoi ce texte occupe un quadrant étriqué au haut du graphique, à gauche. Le dernier quadrant – en bas à gauche – est dévolu au paysage méditerranéen qui est assez large pour accepter les *poissons*, les *chevaux* et le *gibier*. Mais ici l'observation est un peu troublée par l'ambiguïté de la *Naissance du jour*, qui est dans le même élan un hymne à la lumière du midi mais aussi la célébration de la mère disparue dont Colette revendique l'héritage spirituel, le souvenir d'un jardin bourguignon et de ses bêtes familières. La difficulté de l'analyse vient aussi des constantes qui accompagnent Colette partout où elle s'installe : ce sont les *chats* et les *chiens*. Comme on les voit partout au premier plan, ils s'installent au centre du graphique sans manifester de préférence pour un texte ou pour un paysage particulier. C'est pourquoi il est difficile de solliciter davantage le tableau 1, les effectifs étant par ailleurs trop faibles pour permettre le déploiement de la logistique statistique.

Avant de quitter ce premier tableau, observons que les 661 occurrences se regroupent sur la marge droite pour dessiner le profil de Colette, tous textes confondus. Sur la toile de fond de l'ensemble de la littérature, de 1789 à nos jours, l'aimantation que Colette exerce sur les animaux se traduit par des écarts considérables en ce qui concerne le *chat* (59), et le *chien* (15), plus réduits mais généralement positifs à l'égard des autres espèces. Mais la préférence pour le chat est si excessive qu'il a fallu réduire l'échelle de la représentation dans le graphique 2 (grâce à une racine carrée).



Graphique 2. Le bestiaire de Colette

II – On peut estimer toutefois que la comparaison avec l'ensemble du corpus du T.L.F. reste un peu abrupte, car il y a là tant d'écrivains divers, tant de genres opposés et tant d'époques mêlées que la notion de « norme » ou de moyenne manque de consistance. On a donc constitué en corpus un lot d'écrivains qui pour des raisons supposées de ressemblance ou de contraste pouvaient être rapprochés de Colette. À vrai dire nous avons rencontré plusieurs d'entre eux dans nos études préalables et les données s'en trouvaient disponibles¹². Cette liste comporte un peu d'arbitraire, comme celle des animaux retenus (soit une dizaine d'écrivains et une cinquantaine d'espèces animales¹³), et l'on aurait pu sans doute étendre l'enquête ou la resserrer. La définition d'un corpus est nécessairement une décision personnelle, qu'aucune méthode ne pourra jamais formaliser et l'on peut saluer comme une bénédiction cette liberté retrouvée parmi tant de contraintes.

Le tableau 2 étant constitué d'effectifs absolus ne délivre pas d'informations directement exploitables, sinon la présomption que les écarts, s'ils se produisent, seront difficilement aléatoires, étant donné la taille imposante du corpus considéré. Les observations portent en effet sur 22.325 occurrences de noms d'animaux chez les auteurs retenus (et 104.087 dans le corpus entier)¹⁴. On a transformé les relevés bruts en écarts réduits, d'où sont issues toutes les courbes qui vont suivre.

12. En dehors de Giraudoux qu'on a déjà évoqué, il s'agit de Chateaubriand, de Hugo, de Zola et de Proust. Pour ces deux derniers auteurs, voir notre *Vocabulaire de Proust* et notre *Vocabulaire de Zola*, éditions Slatkine, Genève, 1983 et 1985.

13. C'est la même liste que celle du tableau 1. Si elle paraît plus abondante, c'est que le relevé des animaux de Colette ne mentionnait pas les espèces ignorées de Colette, comme l'aigle ou le cerf.

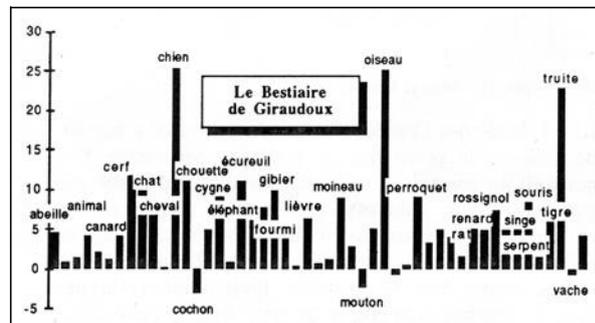
14. L'étendue relative des textes appartenant à chaque écrivain est la suivante (tous les mots, et mêmes les signes de ponctuation, entrant en ligne de compte) : Balzac 2239483, Chateaubriand 1.688.828, Claudel 1.301.385, Colette 245.603 mots, Giraudoux 885.258, Hugo 2.602.231, Maupassant 1.006.953, Pergaud 45.774, Proust 1.507.681, Renard 545.591 et Zola 3.609.064, soit plus de 15 millions de « mots » pour l'ensemble des auteurs (et 85 millions dans le corpus total XIX^e-XX^e siècles).

	Colet	Balzac	Chate	Claud Girau	Hug	Maup	Perga	Proust	Renar	Zola	TOTA	corpus	
abeille	11	10	31	68	34	74	1	0	5	3	13	250	1513
aigle	0	68	104	44	24	303	4	0	15	11	33	606	1930
âne	3	34	21	73	24	80	15	0	8	34	63	355	1724
animal	25	142	127	201	201	99	114	8	68	42	118	1145	12839
araignée	12	16	7	30	16	96	11	0	4	36	39	267	904
boeuf	3	39	43	82	32	123	34	6	30	92	90	574	2438
canard	0	41	11	10	22	12	22	0	10	35	39	202	807
cef	0	18	25	19	34	20	11	1	4	5	24	161	558
chat	198	125	31	29	95	138	80	7	16	81	263	1063	3471
cheval	50	485	280	241	235	533	252	2	115	107	656	2956	14521
chèvre	5	25	30	26	14	124	11	0	6	11	50	302	1256
chien	97	185	103	180	318	268	289	42	41	155	384	2062	8180
chouette	3	4	5	7	24	19	11	3	6	10	13	105	335
cochon	7	22	3	42	5	10	48	0	5	69	254	465	1602
coq	1	25	11	37	29	40	30	6	11	46	46	282	1155
cygne	1	18	51	22	40	85	6	0	20	9	18	270	995
dragon	2	13	35	34	14	102	12	0	8	2	19	241	1024
écureuil	1	3	10	4	23	4	5	3	2	6	2	63	306
éléphant	1	6	17	16	29	66	7	0	0	7	3	152	926
fourmi	2	6	8	14	29	25	14	0	0	14	25	137	728
gibier	2	20	1	13	34	5	20	1	4	11	20	131	699
insecte	5	38	47	27	46	30	12	10	38	10	52	315	2496
lapin	11	14	10	31	14	8	65	4	7	30	101	295	1214
lièvre	8	6	16	14	27	10	20	35	1	36	29	202	824
lion	0	67	104	92	31	276	10	0	16	19	33	648	2580
loup	11	74	53	32	30	165	51	1	20	19	115	571	2301
moineau	1	20	6	5	24	43	3	2	2	37	53	196	447
mouche	21	61	19	35	34	124	35	3	12	57	84	485	1992
mouton	8	49	26	53	9	49	29	0	5	42	71	341	1768
oie	5	8	9	7	21	26	10	0	5	48	102	241	704
oiseau	45	107	279	214	358	652	126	42	105	119	158	2205	9928
ours	2	53	69	36	11	98	5	0	4	16	31	325	1259
papillon	13	15	14	46	14	77	3	0	21	15	28	246	1141
perroquet	1	13	5	9	25	6	9	0	3	11	1	83	472
pigeon	2	25	13	21	20	15	20	0	22	23	47	208	909
poisson	7	48	47	154	65	61	71	1	34	56	103	647	3414
poule	14	37	17	26	31	39	63	15	16	64	116	438	1442
poulet	14	27	6	11	13	7	43	0	15	23	69	228	792
rat	8	51	9	29	32	75	8	2	9	19	30	272	1270
renard	6	10	7	16	23	23	23	30	0	10	8	156	830
rossignol	2	19	43	22	30	21	6	0	3	5	2	153	810
serpent	11	44	113	40	43	79	35	4	23	21	28	441	1961
singe	4	46	16	14	34	39	19	0	9	17	24	222	1360
souris	14	19	17	29	34	99	13	1	12	19	31	288	861
taureau	0	21	39	23	21	43	2	0	3	28	23	203	1425
tigre	3	68	46	23	27	134	6	0	4	3	2	316	860
truite	0	2	2	5	37	4	8	0	6	2	10	76	224
vache	12	22	37	57	19	42	54	2	16	83	162	506	2084
veau	9	21	9	22	21	8	11	0	5	40	83	229	808
TOT2	661	2290	2032	2285	2370	4479	1757	231	794	1658	3768	22325	104087

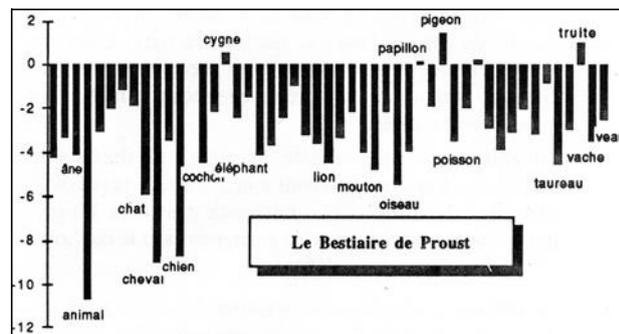
Tableau 2. Bestiaires comparés. Relevés des effectifs

a – La première met en parallèle, ou plutôt en opposition, deux écrivains contemporains que Colette a pareillement connus et appréciés : Proust et Giraudoux. Ici comme sur bien d'autres points, le contraste est accusé (graphiques 3 et 4). **Proust** s'intéresse peu au monde animal, sinon à quelques espèces poétiques dans certains passages de Combray : cygne, le *papillon*, la *truite*. Lorsqu'il prête attention au monde non-humain, ce qui accroche son regard c'est le règne végétal, la silhouette d'un arbre, la nuance de l'églantine ou du lilas, la crête des vagues à l'horizon ou la forme des nénuphars. Partout la courbe 4 descend largement dans la zone des déficits. La courbe 3 est l'image inversée de la précédente. C'est que **Giraudoux** se pose résolument en écrivain

animalier, à l'image de La Fontaine à qui il a voué un culte et un livre. Si la campagne de Giraudoux, et notamment celle de *Provinciales*¹⁵ ressemble assez à celle de Colette, la préférence de Giraudoux va au *chien* plutôt qu'au *chat*, au *cerf*, au *cheval*, à *l'écureuil*, au *gibier*, aux *oiseaux*, aux *moineaux* et aux *rossignols*. Manifestement Giraudoux admire la noblesse des animaux libres, même si cette liberté est menacée dans l'acte sacrificiel de la chasse, même si la liberté du *tigre*, du *singe* ou du *perroquet* est niée par les barreaux de la cage. Par contre la sympathie et l'attention de Giraudoux se détournent de la basse-cour et de l'étable, et généralement des animaux destinés à l'étal du boucher (*cochon*, *mouton*, *vache*).

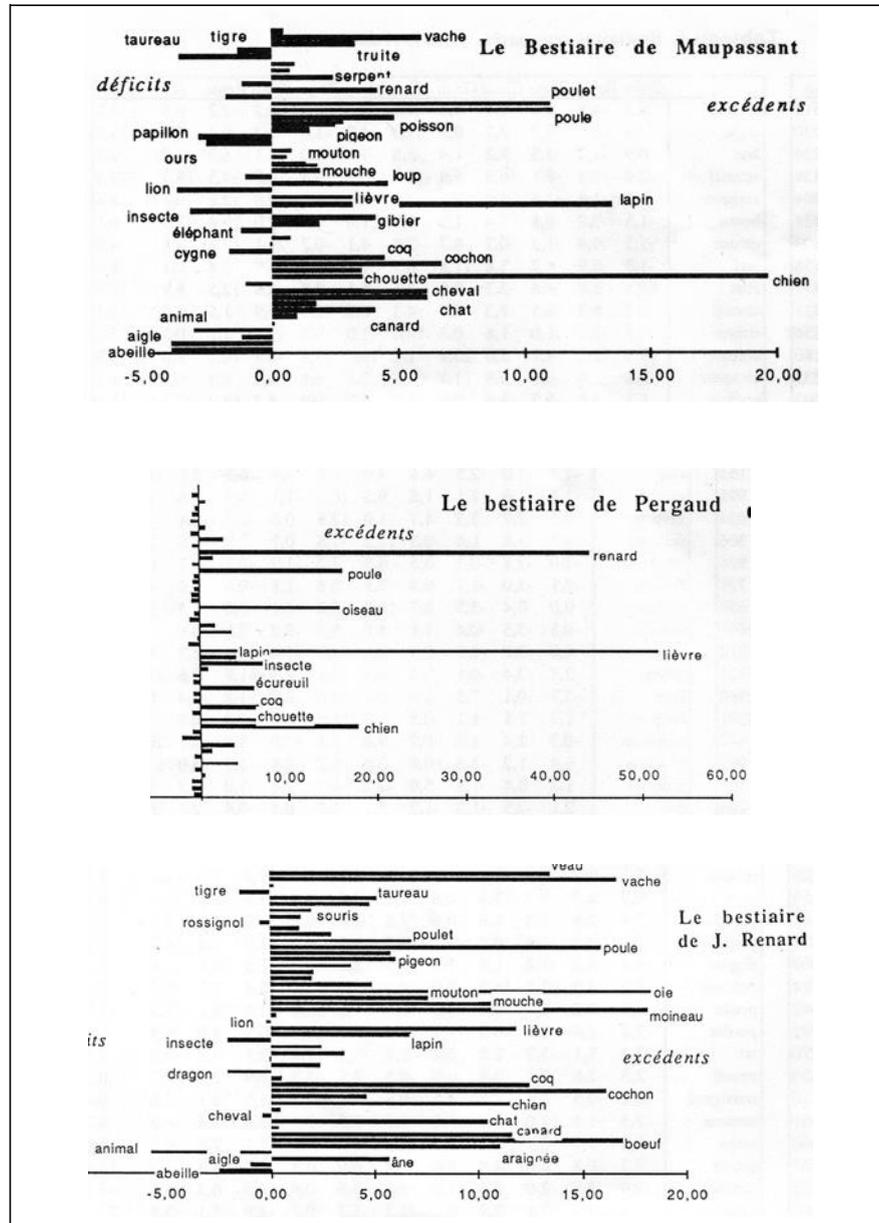


Graphique 3. Le bestiaire de Giraudoux



Graphique 4. Le bestiaire de Proust

15. C'est une phrase pittoresque des *Provinciales* qui valut à Giraudoux sa carrière diplomatique sinon sa fortune littéraire, en retenant l'attention de Philippe Berthelot qui régnait alors au Quai d'Orsay et qui aimait les animaux (c'est le même homme qui devait donner à Colette la fameuse panthère Bâ-Tou). La phrase, parmi beaucoup d'autres semblables, avait de quoi plaire à un amoureux des bêtes : « *Un cheval passa. Les poules suivirent, pleines d'espoir.* »



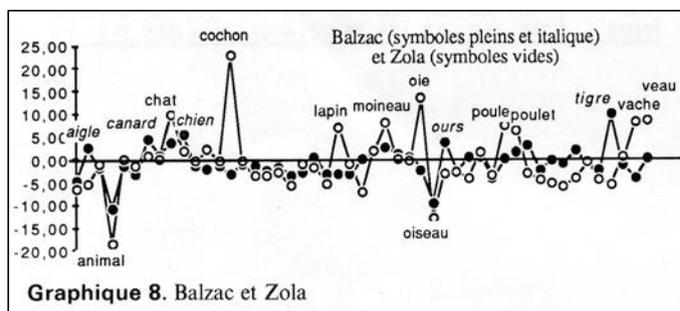
Graphique 5. Le bestiaire de Maupassant, Pergaud et Renard

b – Le graphique 5 superpose trois écrivains qui ne répugnent pas à la description réaliste et qui à ce point de vue s'écartent autant de Proust

que de Giraudoux. On ne s'attardera guère à **Pergaud** qui ne compte qu'un seul texte dans le corpus (*De Goupil à Margot*¹⁶) et dont les thèmes, étroitement liées à la chasse et à la basse-cour, sont ceux qu'on trouverait dans le *Roman de Renart* : *renard, lièvre, chien, lapin, oiseau, poule, coq*. L'univers de **Maupassant** est à peu près le même, mais les espèces y sont plus variées, qu'il s'agisse de gibier à poil ou à plume (*lapin, lièvre, renard, loup, gibier, canard, pigeon*), de poissons (*truite, poisson*) ou des animaux de la ferme (*cheval, cochon, chat, poule, poulet, vache*). Le goût de Jules **Renard** par contre ne va guère à la chasse, et toutes les bêtes de la campagne, surtout les plus humbles et les plus domestiques, ont droit aux égards, parfois humoristiques, de l'écrivain. Le *chien* tant aimé de Maupassant a perdu ses privilèges, et le *cheval*, sans doute trop noble, a quitté le devant de la scène, en cédant les premières places au *cochon*, à l'*âne* et au *mouton*, au *boeuf* et au *taureau*, à la *vache* et à son *veau*, aux *poules*, *oies*, *coqs*, *poulets* et *canards*, aux *mouches* et aux *araignées*. En réalité Jules Renard a tenté une comédie animale, et pour cela les êtres naïfs et rustauds offraient plus de ressources. C'est pourquoi tous les animaux chargés de symboles, de quelque dignité sociale ou d'un mystère poétique, sont rejetés par lui dans la zone des déficits, où Maupassant les avait aussi relégués. Le réalisme boude le *tigre* et le *lion*, l'*aigle* et le *dragon*, l'*abeille* et le *rossignol*.

c – Le même refus de la noblesse est observé chez **Zola** (graphique 6). Mais ce refus est généralisé à l'ensemble de la gent animale, en exceptant quelques espèces qui sont liées au rôle utilitaire de fournisseur de viande (*veau, vache, poule, poulet, oie, lapin, cochon*). Le *cochon* est aussi probablement un fournisseur d'injures et c'est peut-être là la raison de sa promotion en tête de liste. En dehors de la *Terre* et de la *Faute de l'abbé Mouret*, Zola s'est peu intéressé au monde paysan, et peu à la chasse, et autour de lui il n'a guère supporté que le *chat*. Le désintéret de Balzac, dont la courbe est superposée à celle de Zola, est plus systématique encore. Peu d'animaux dépassent chez lui la ligne médiane, en dehors du *chat*, de l'*aigle*, du *rat*, de l'*ours* et surtout du *tigre*. L'un et l'autre romanciers semblent trop préoccupés par la comédie humaine pour prêter beaucoup d'intérêt à la coulisse animale.

16. C'est ce roman qui obtient le Goncourt, quand la *Vagabonde* obtient deux voix. Battue par les animaux de Pergaud, Colette a dû se consoler.



Graphique 6. Le bestiaire de Balzac et Zola

d – On sait que le romantisme, sensible aux thèses de Rousseau, a célébré le retour à la nature, la vertu des plantes et des simples et l’innocence des bêtes. **Hugo** défend le crapaud contre les garnements et le cochon contre les poètes trop délicats :

Je nommai le cochon par son nom. Pourquoi pas ?

J’ôtai du cou du chien stupéfait son collier

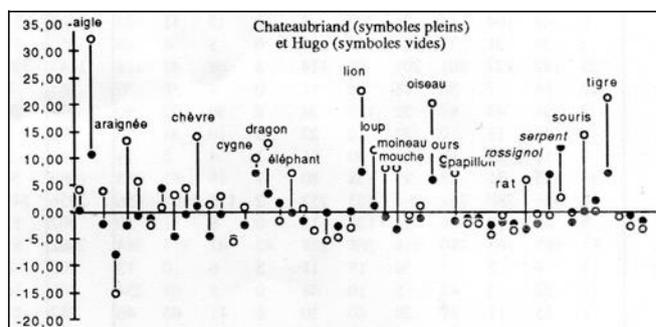
D’épithètes ; dans l’herbe, à l’ombre du hallier,

Je fis fraterniser la vache et la génisse...

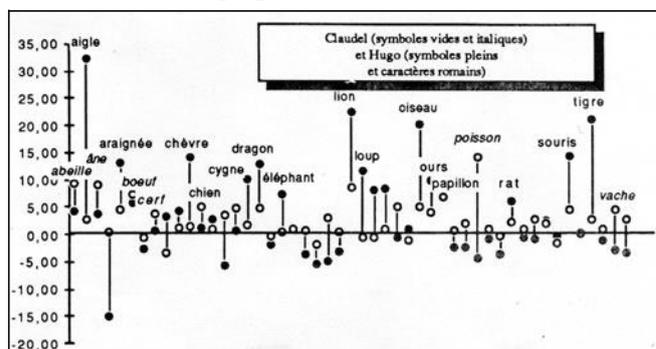
(Réponse à un acte d’accusation)

En réalité les intentions révolutionnaires de Hugo n’ont pas toujours été suivies d’effet. Et les animaux que sa plume chérit sont les plus nobles que la création (ou l’imagination) ait produits : l’aigle, le lion, le tigre, le loup, l’ours, le dragon, le cygne¹⁷. Il arrive à Hugo d’approcher les intouchables et de mettre en scène, en les idéalisant, une vache, un âne ou un crapaud. Mais le plus souvent il respecte les valeurs stylistiques établies, et son bestiaire s’écarte peu de celui de la poésie traditionnelle, sa préférence allant toutefois aux animaux qui symbolisent la force. **Chateaubriand** partage avec lui le sens de la grandeur et, sans avoir un bestiaire aussi abondant que celui de Hugo, il accueille volontiers dans sa prose, comme Hugo, les aigles, les cerfs, les cygnes, les lions, les ours, les tigres, les rossignols, les serpents. Le graphique 7, qui superpose les deux distributions, montre des mouvements assez parallèles. Plus éloigné dans le temps, **Claudé** se préoccupe moins des classes supérieures de la société animale : si le lion a encore une position forte, l’âne, le boeuf, la vache ne sont plus rejetés dans l’oubli et l’abjection (graphique 8).

17. La chèvre d’Esmeralda et l’éléphant monumental qui abrite Gavroche doivent leur fréquence – peu significative – à un accident thématique.



Graphique 7. Le bestiaire de Chateaubriand et Hugo



Graphique 8. Le bestiaire de Hugo et Claudel

III – L'étude des 11 colonnes du tableau 2 initial a permis de dessiner le profil des 11 écrivains considérés. On pourrait au contraire s'intéresser aux lignes du tableau, c'est-à-dire aux animaux, et suivre la démarche de chacun parmi les écrivains en observant lesquels il suit et lesquels il fuit. Mais comme il y a là 49 espèces différentes, c'est autant de « profils » qu'il faudrait établir, dont le foisonnement risque de provoquer la confusion. Cela ne se justifie que pour les espèces les mieux représentées, comme le *cheval*, le *chien*, ou le *chat*. Mais on peut imaginer des critères qui puissent permettre le regroupement des espèces en classes, et par exemple adopter la classification zoologique qui distingue les mammifères, les oiseaux, les reptiles, les insectes, etc., ou encore la division entre animaux sauvages et domestiques. Nous n'avons retenu, par curiosité, que des critères formels et plus grammaticaux que sémantiques : le sexe, le nombre et le genre grammatical. Le tableau 3 ci-dessous reproduit les effectifs des femelles et des mâles, du féminin et du

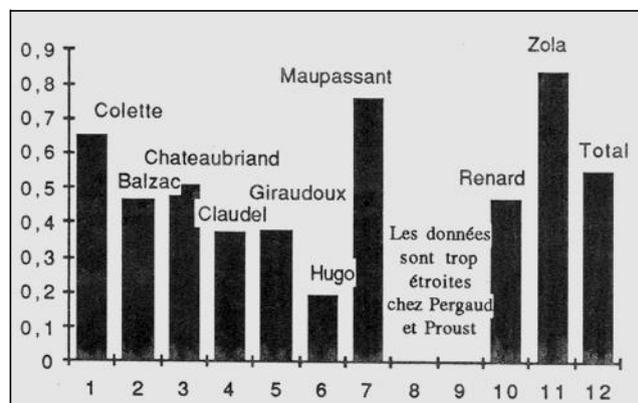
masculin, du singulier et du pluriel et dans les trois cas il établit le rapport entre les deux séries.

	Colette	Balzac	Chatea	Claude	Giraud	Hugo	Maupa	Pergau	Proust	Renard	Zola	TOTAL
<i>femelle</i>	87	82	60	63	63	66	104	16	39	116	293	627
<i>mâle</i>	133	176	118	169	164	335	136	19	53	246	347	1136
<i>fem/mâle</i>	0,654	0,466	0,508	0,373	0,384	0,197	0,765	0,842	0,736	0,472	0,844	0,552
<i>singulier</i>	456	1364		1523		2636	1046	119	462	1039	2344	7646
<i>pluriel</i>	184	812		658		1616	662	107	307	562	1338	4592
<i>sing/plur</i>	2,478	1,680		2,315		1,631	1,580	1,112	1,505	1,849	1,752	1,665
<i>féminin</i>	168	244	188	306	310	681	241	24	95	348	720	3325
<i>masculin</i>	493	2046	1844	1979	2060	3798	1516	207	699	1310	3048	19000
<i>fémin/masc</i>	0,341	0,119	0,102	0,155	0,150	0,179	0,159	0,116	0,136	0,266	0,236	0,175

Tableau 3. Le sexe, le genre et le nombre

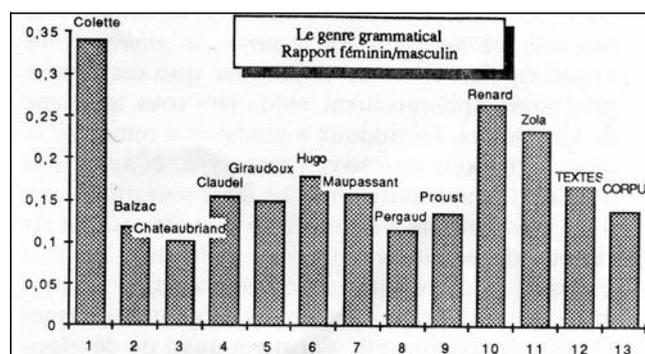
a – Le langage considère beaucoup d'espèces comme asexuées, et par exemple ne fait pas de différence entre mâles et femelles dans le cas des *mouches* ou des *serpents*. Les effectifs relatifs au **sexe** ne font entrer ici en ligne de compte que les espèces où la forme diffère selon qu'on désigne la femelle ou le mâle. Ainsi la *vache*, la *chatte*, et la *poule* se groupent du côté de l'évangile, quand le *taureau*, le *boeuf*, le *chat* et le *coq* se tiennent du côté de l'épître. Or ce rapport femelle/mâle subit d'amples variations suivant les écrivains (graphique 9). On pouvait s'attendre que Colette montrât quelque préférence pour la femelle – ce que les chiffres confirment. Elle reconnaît une *chatte* là où d'autres, moins connaisseurs ou moins curieux, verraient tout uniment un *chat*, c'est-à-dire un spécimen de l'espèce. Peut-être même y a-t-il chez Colette, à l'endroit des animaux femelles, une sorte de familiarité particulière ou d'intimité supplémentaire, une amitié plus prompte à franchir la barrière qui sépare les bêtes et les humains que celle qui oppose les hommes et les femmes. Malheureusement Colette est le seul écrivain femme de notre présent corpus et un sondage du côté de Georges Sand par exemple donnerait des informations utiles. Chez les hommes, l'opposition est tranchée entre Claudel et surtout Hugo d'un côté, Maupassant et Zola de l'autre. Le rapport est favorable au mâle chez les premiers, à la femelle chez les seconds. Faut-il rapprocher cette observation du goût très prononcé que Maupassant et Zola éprouvent pour la femme et dont les chiffres rendent compte avec la même netteté¹⁸ ?

18. Sur la féminisation de l'univers zolien, voir notre *Vocabulaire de Zola*, tome 1, p. 335. Les occurrences de la *femme* y sont plus nombreuses que celles de l'*homme*, qui habituellement – sur l'ensemble du corpus – occupe un volume double. Chez Hugo au contraire on compte 5.817 *hommes* pour 2.169 *femmes*, soit près de trois fois plus.



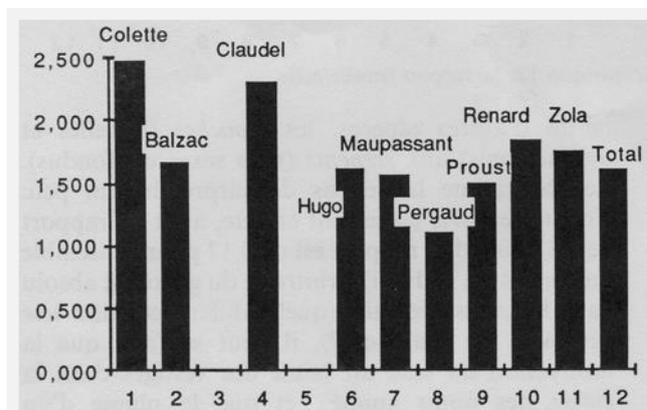
Graphique 9. Le rapport femelle/mâle

b – Le **genre** grammatical apparaît dans le graphique 10. Il n'est guère besoin d'insister sur l'arbitraire du genre, qui, pour la plupart des substantifs, est dégagé de la motivation du sexe. Afin qu'il n'y ait pas interférence des deux phénomènes, les effectifs précédents, qui concernaient expressément le sexe, ont été écartés du présent relevé, où l'on oppose, parmi d'autres espèces, les *mouches* (femelles et mâles réunis) aux *serpents* (tous sexes confondus). Le résultat ne laisse pas de surprendre un peu : Colette se porte largement en tête, avec un rapport de 0,37 quand ce rapport est de 0,17 pour l'ensemble du corpus du T.L.F. Si l'arbitraire du genre est absolu dans les mots abstraits (quelle différence entre *une* action et *un* sentiment ?), il peut se faire que la motivation du sexe ait laissé des vestiges dans la classe des noms animés, et que la plume d'un écrivain femme fasse parmi les noms animés un tri inconscient qui diffère de celui des hommes. Là encore une enquête plus systématique pourrait avoir certain intérêt.



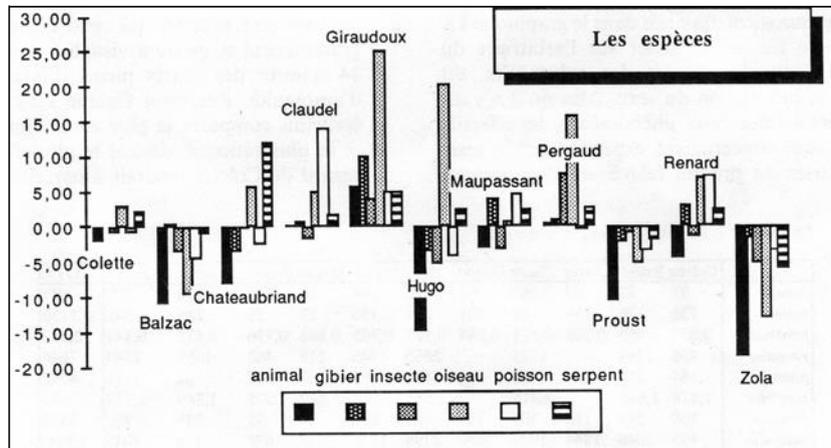
Graphique 10. Le rapport féminin/masculin

c – Le troisième rapport qui rend compte du **nombre** grammatical et qu'on a visualisé dans le graphique 11 montre des écarts pleins d'intérêt mais aussi d'ambiguïté. Pourquoi Colette est-elle de tous les écrivains comparés la plus accueillante au singulier et la plus réticente devant le pluriel ? Est-ce que le regard de Colette tendrait à parcelliser et à actualiser, plutôt qu'à globaliser et à abstraire ? Le doute subsiste parce que l'abstraction s'accommode du singulier comme du pluriel et par exemple s'exprime pareillement dans les deux variantes : *le chien est l'ami de l'homme* et *les chiens sont les amis de l'homme*. N'est-ce pas plutôt l'indice de la relation personnelle que Colette entretient avec chacune de ses bêtes, et de l'intimité qui la lie à un individu, non à l'espèce ?



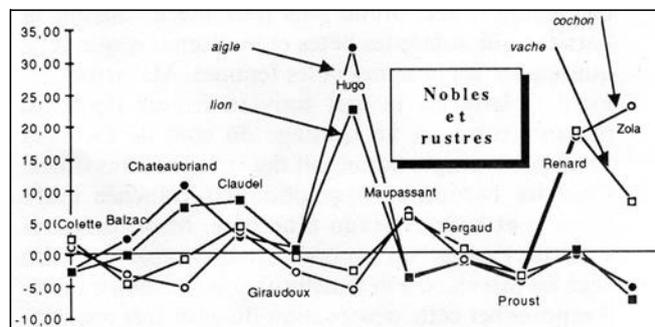
Graphique 11. Le rapport singulier/pluriel

La confirmation de ce trait peut être trouvée précisément dans la distribution des noms d'**espèces**. Certains termes de zoologie représentent un plus haut niveau de généralité, comme le mot *animal* lui-même, et, dans une mesure moindre, les *oiseaux*, les *insectes*, les *poissons*, les *serpents*, le *gibier*. Nous avons eu l'occasion de remarquer que ces termes génériques apparaissaient volontiers sous la plume de Giraudoux. Giraudoux a tendance à remonter la généalogie pour retrouver l'archétype. N'ayant pas le souci du naturaliste (dans les deux sens du terme), il est plus sensible à l'essence de l'espèce, qu'il n'est curieux de la morphologie des individus. Le graphique 12 corrobore ces observations. Il montre aussi que Colette ne partage pas ce goût des termes génériques, comme elle s'abstient aussi de développer les théories sur le statut de l'animal où Giraudoux se complait parfois. L'animal n'est pas pour Colette un objet de réflexion, mais un objet de tendresse.



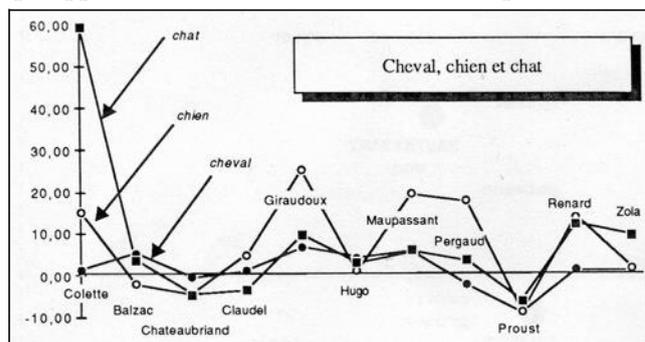
Graphique 12. Les espèces

d – Beaucoup des analyses qui précèdent laissent soupçonner que la société animale est régie par des clivages anthropomorphiques qui sont le reflet des classes sociales dans la société humaine. Dans la société des animaux les nobles consentent difficilement à se mêler aux manants. Et la révolution chez les bêtes a suivi d'assez loin celle de 1789 chez les hommes. Le graphique 13 met en relief la persistance dans la première partie du XIX^e siècle de la ségrégation animale. 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 rustaude 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, Zola ou Pergaud, bovins et porcins envahissent alors la littérature, tandis que se retirent les seigneurs. Au XX^e siècle la cause est entendue et l'égalité sociale acceptée : ni Colette, ni Giraudoux, ni Claudel, ni Proust ne font la différence entre nobles et rustres, et à leur niveau les courbes se rejoignent dans le graphique. Cela ne veut pas dire que toute distinction est abolie, mais les critères de choix ont changé.



Graphique 13. Les nobles et les rustres

Ainsi les animaux bourgeois, ni manants, ni seigneurs, qu'on a groupés dans le graphique 14, sont loin d'être logés à la même enseigne. En particulier le *chien*, le *chat* et le *cheval* sont traités très différemment par Colette. Le *chat* a chez elle des privilèges exorbitants, au détriment du *cheval*. Quelques restes de sang bleu valent au contraire à ce dernier la faveur relative de Hugo et de Balzac. Quant au *chien*, son succès est net chez ceux qui apprécient la chasse : Giraudoux, Maupassant et Pergaud.

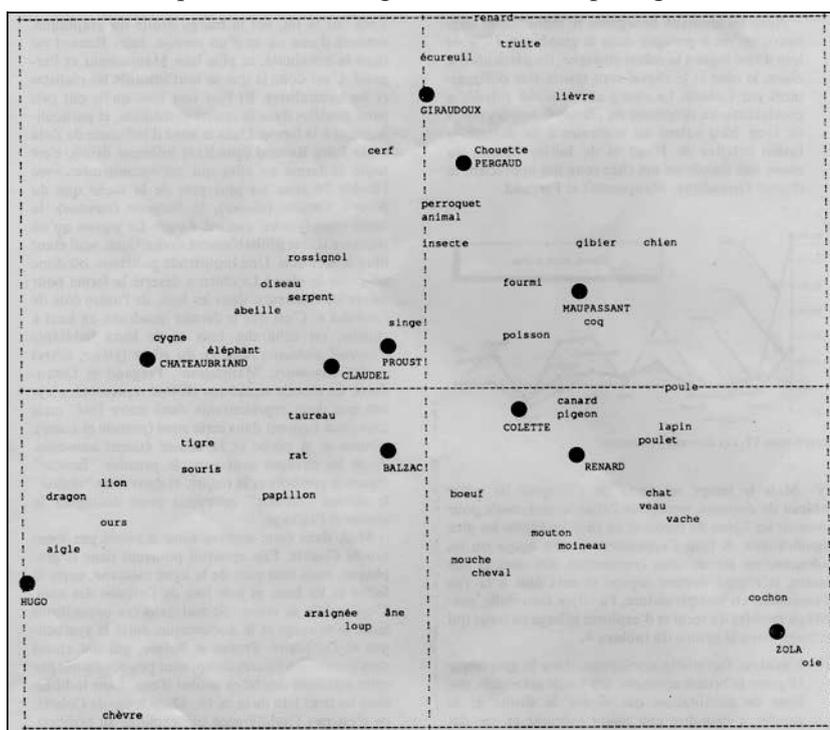


Graphique 14. Les animaux « bourgeois »

IV – Mais le temps est venu de s'éloigner de notre tableau de données, non pour l'abandonner, mais pour en saisir les lignes de forces et en tirer les leçons les plus significatives. A trop s'appesantir sur les lignes ou les colonnes, ou sur certains croisements des unes et des autres, le regard devient myope et cela nuit à la vue d'ensemble. Or la procédure que nous avons utilisée une première fois pour survoler le corpus propre à Colette va nous servir derechef pour explorer le large paysage qui se cache dans la brume du tableau 2.

a – L'analyse factorielle représentée dans le graphique 15 permet de percer la brume aisément. On y voit nettement une ligne de décantation

qui sépare la droite et la gauche, c'est-à-dire 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 symbolisme 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.



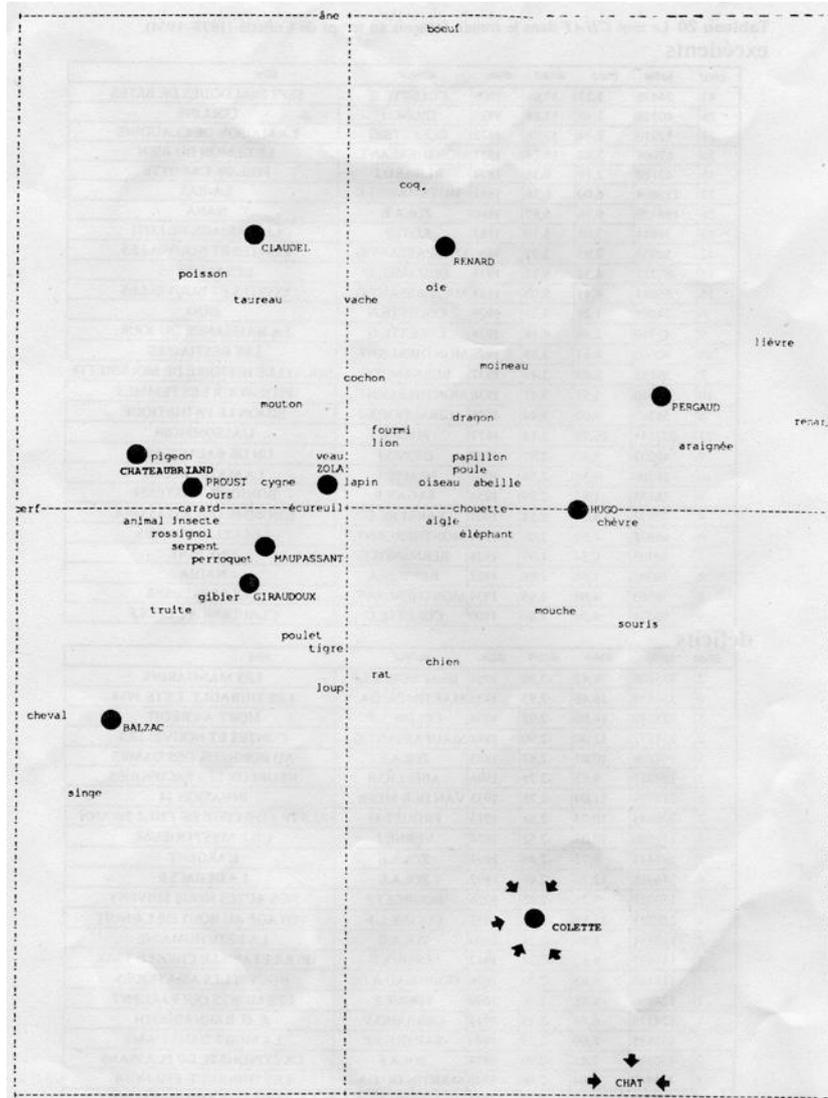
Graphique 15. Analyse factorielle du bestiaire de quelques écrivains (facteurs 1 et 2)

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 Jules Renard (quadrant inférieur droit), c'est toute la ferme en effet qui est reconstituée : avec l'étable (le *veau* est plus près de la *vache* que du *boeuf*), l'écurie (*cheval*), la bergerie (*mouton*), la basse-cour (*poulet, canard, lapin*). Le *pigeon* qu'on rencontre ici est probablement domestique, seul étant libre le *moineau*. Une inquiétude pourtant : où donc est passé le *chien* ? Le chien a déserté la ferme pour suivre les chasseurs, dans les bois, 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*) et des chasseurs : Maupassant, Pergaud et Giraudoux. Le monde aquatique est mal représenté, n'ayant que deux représentants dans notre liste, mais tous deux figurent dans cette zone (*poisson et truite*), comme si la pêche et la chasse étaient associées. Ainsi les clivages sont nets : le premier facteur sépare le symbole et la réalité, et dans cette « réalité » le second facteur intervient pour distinguer la chasse et l'élevage.

Mais dans cette analyse nous n'avons pas mentionné Colette. Elle apparaît pourtant dans le graphique, mais tout près de la ligne médiane, entre la ferme et les bois, et non loin de l'origine des axes. C'est qu'elle se reconnaît mal dans ces oppositions entre le sauvage et le domestique, entre le symbolique et l'utilitaire. Proust et Balzac, qui ont choisi sans conviction l'autre camp, sont peu concernés par cette agitation des bêtes autour d'eux. Leur indifférence les tient loin de la mêlée. Dans le cas de Colette, ce n'est pas l'indifférence qui explique sa position, mais bien au contraire la tendresse qu'elle voue à des espèces variées, dispersées à divers endroits du graphique. Mais Colette, partagée et hésitante dans le premier graphique, impose sa loi dans le second (figure 16), qui poursuit la même analyse en envisageant les facteurs 3 et 4. A elle seule Colette contribue pour les trois quarts au facteur 4 et visuellement son influence prépondérante est facile à constater, au bas du graphique. Là Colette marque nettement ses préférences, accaparant son favori, le *chat*.

b – Cette préférence pour le *chat*, quoique non exclusive, est si marquée qu'elle nous a engagé dans une recherche supplémentaire, afin de vérifier si d'autres écrivains n'ont pas tenté de disputer à Colette le cœur du *chat*.

Car notre sélection, n'étant pas exhaustive, pouvait laisser passer le doute.



Graphique 16. Analyse factorielle du bestiaire de quelques écrivains (facteurs 3 et 4)

Or le doute n'est plus permis à la lecture des romans publiés à l'époque de Colette, de sa naissance à sa mort. Sur 179 romans complets

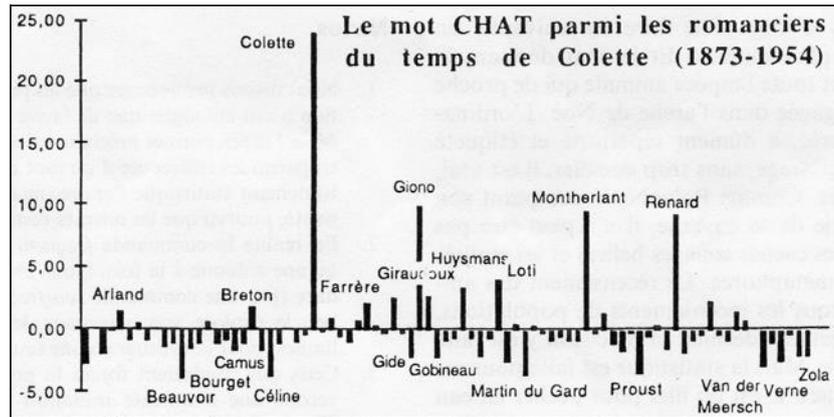
ainsi sélectionnés de 1873 à 1954, Colette arrive en tête avec les *Dialogues de bêtes* et la *Maison de Claudine*. *Sido* (douzième rang) et la *Naissance du jour* (treizième) confirment cet avantage, que ne dément pas *Claudine à l'école* (rang 27)¹⁹. Aucun autre écrivain ne concentre toute sa production dans cette zone habitée par les *chats*, même si certains se plaisent en leur compagnie, comme Giono dans deux de ses romans : *Colline* et *Un de Baumugnes*, Montherlant dans cinq des siens : *Le Démon du bien*, *Les Bestiaires*, *Pitié pour les femmes*, *Les Célibataires* et *Les Lépreuses*, ou encore Giraudoux (*Simon le Pathétique*, *Suzanne et le Pacifique*), Zola (*Nana*, *L'Assommoir*), Bernanos (*Nouvelle histoire de Mouchette*, *Une Nuit*). Encore faut-il prendre garde que le même auteur peut se trouver en tête et en queue de cette même liste, suivant les variations thématiques de sa production. Et c'est le cas de Zola (une occurrence de *chat* dans *Au Bonheur des Dames*, quand *Nana* en contenait 28)²⁰

c – C'est pourquoi le graphique 17 regroupe tous les textes qui appartiennent à un même auteur. Les 84 romanciers recensés trouvent ainsi une place moyenne qui correspond mieux à leur sentiment à l'égard du *chat*. Ainsi Zola s'inscrit dans la zone négative, ce qui s'explique par la préférence donnée à la *chatte*. Dans cette zone des déficits on trouve également Camus, Céline, Gobineau, Gide, Martin du Gard, Proust, J. Verne. Les excédents les plus accusés sont le fait de Giono, Montherlant, Renard, Huysmans, Loti et Giraudoux. Mais c'est Colette qui se distingue de toute la production romanesque en élevant à son animal favori, au-dessus de toutes les têtes, une colonne triomphale et en nous conviant au couronnement du *chat*²¹.

19. Notre enquête porte ici sur la forme *chat*, à l'exclusion du pluriel et du féminin. C'est pourquoi les effectifs et les écarts ne correspondent pas tout-à-fait à ceux qu'on a produits précédemment et qui prenaient en compte le vocable, non la forme seule.

20. On remarquera que c'est une femme, Simone de Beauvoir, qui occupe la position symétrique de celle de Colette, dans le classement des déficits : deux occurrences seulement du mot *chat*, pour une étendue de 355.108 mots, dans les *Mandarins*.

21. On peut élargir l'enquête à l'ensemble du corpus du T.L.F., de 1789 à nos jours, en comparant la distribution du *chat* à celle du *chien* et du *cheval*. L'influence de Colette est alors noyée dans la masse. Mais l'évolution chronologique y est sensible qui accorde au *chat* une faveur grandissante au cours du XIX^e siècle, alors qu'un destin inverse est réservé au *cheval* : succès jusqu'en 1865, déclin depuis lors. Quant au *chien*, son sort est parallèle à celui du *chat*, et les deux courbes sont torsadées comme dans une guirlande. Visiblement les *chats* et les *chiens* sont faits pour s'entendre, comme Toby-Chien et Kiki-la-Doucette.



Graphique 17. Le mot *chat* parmi les écrivains contemporains de Colette

Nous retrouvons donc notre point de départ, la préférence avouée de Colette pour les chats. La statistique ici confirme le sentiment de l'évidence. Certes il ne s'agit pas d'une découverte et l'on n'a avancé que médiocrement, en prouvant, sinon que la méthode est juste, du moins qu'elle n'est pas controuvée. Mais le premier pas nous a engagé à faire les suivants, en abordant des cas plus épineux. Et le chat donnant la patte au chien, c'est toute l'espèce animale qui de proche en proche s'est engagée dans l'arche de Noé. L'ordinateur, posté à l'entrée, a dûment répertorié et étiqueté chaque couple au passage, sans trop démêler, il est vrai, les faux et les vrais. Comme Polyphème comptant son troupeau à la sortie de sa caverne, il n'a peut-être pas discerné les hommes cachés sous les béliers et les réalités sous-jacentes aux métaphores. Le recensement des animaux est comme tous les recensements de populations, en attendant le jugement dernier : la précision y est faite avec de l'incertitude. Mais la statistique est faite pour ces situations équivoques. C'est un filet pour pêcher en eau trouble.