

### 3. 1. LA DIFFUSION DES TRAITS SEMANTIQUES

Le chapitre 2 nous a permis de montrer l'importance du contexte, qui "marqué" en quelque sorte par le discours et le champ générique, donnait des "couleurs" différentes à une lexie et à son voisinage. L'approche interprétative-herméneutique du texte nomme *isotopie*<sup>1</sup> l'itération d'une unité linguistique, engendrée par la diffusion ou propagation des traits sémantiques ; ce concept théorique sert à traiter d'une façon unifiée la question de la cohésion textuelle, aux différents paliers de l'interprétation (lexie, syntagme, énoncé, période ou texte) puisqu'il est indépendant des limites de la phrase et des structures syntaxiques. En comparant les sélections de cooccurrents autour de la lexie sous ses formes du singulier et du pluriel, et en tant que lemme (forme canonique) et quand on fait varier le corpus de travail en insérant la lexie *pied* dans deux syntagmes minimaux, *pied(s) de* et *au(x) pied(s) de*, nous allons mettre en évidence la cohésion textuelle, relative à la diffusion et à l'itération des traits sémantiques.

#### 3. 1. 1. Unités signifiantes et paliers d'interprétation

##### *La morphologie et la variabilité des cooccurrents de la lexie*

Dans ce chapitre et les suivants, nous ne traiterons plus que du corpus Roman 1830-1970 de Frantext<sup>2</sup>. Nous avons signalé au chapitre I qu'en fait il comporte un ensemble mixte de romans et nouvelles, et les raisons pour lesquelles nous avons conservé cette relative hétérogénéité, parce que le point de vue de la thématique multi-auteurs adopté était compatible avec cet aspect, et afin d'observer d'éventuelles différences<sup>3</sup>.

Les tableaux qui suivent proposent un ensemble de résultats statistiques obtenus en faisant varier le corpus de travail :

- sélection des cooccurrents du lemme *pied*, symbolisé par *pied(s)*, dans le corpus Roman, dans la fenêtre 10 mots avant et après
- sélection des cooccurrents de la forme *pied*, dans la fenêtre 10 mots avant et après
- sélection des cooccurrents de la forme *pieds*, dans la fenêtre 10 mots avant et après
- sélection des cooccurrents de chacune des formes, *pied* et *pieds*, dans la fenêtre 10 mots avant (le contexte gauche)
- sélection des cooccurrents de chacune des formes, *pied* et *pieds*, dans la fenêtre 10 mots après (le contexte droit)

<sup>1</sup> Sur l'extension de ce concept, d'abord proposé par Greimas pour qualifier certains composants sémantiques, v. Rastier F., 1987, 87-93.

<sup>2</sup> Ce corpus, constitué pour un ensemble d'expériences en 1994, a été figé à cette date et, comme Frantext s'est enrichie entre temps, il ne correspond pas tout à fait à celui que l'on pourrait sélectionner aujourd'hui dans la banque textuelle.

<sup>3</sup> De nouveaux travaux de contraste entre ces deux champs génériques ne pourront être entrepris qu'avec un corpus de nouvelles plus important en volume, pour relativiser la part de Maupassant, d'une part et représenter la production du XX<sup>e</sup>s., pour pouvoir y repérer des sous-genres comme dans le roman également ; d'autre part, il est nécessaire de disposer de corpus enrichis pour traiter des différences entre ces deux champs génériques. Cette question sera évoquée et illustrée à plusieurs reprises.

### CHAPITRE III : LES UNITÉS SÉMANTIQUES

Le tableau 1 montre ces variations d'association, pour le corpus Roman, selon qu'on a appliqué le test statistique au voisinage de la forme singulier de la lexie *pied*, de la forme pluriel ou bien en ne les distinguant pas.

<i>pied(s)</i>		<i>pied</i>		<i>pieds</i>	
score stat.	forme	score stat.	forme	score stat.	forme
166	nu	144	plain	212	nu
101	pointe	89	valet	101	pointe
98	plain	67	bot	93	fouler
79	fouler	58	étrier	61	foulé
59	valet	58	grue	50	chaussé
59	foulé	55	coup	45	chenet
57	bot	45	lit	43	sol
56	chaussé	39	pointe	43	pantoufle
48	sol	39	frappant	41	tête
46	étrier	37	mettre	36	sandale
42	pantoufle	36	terre	32	traînant
42	terre	34	pentamètre	32	conculatrice
39	grue	32	cul	31	soulier
39	coup	32	frapper	29	pouce
38	mettre	28	chaussé	29	savate
37	lit	28	mis	29	lié
36	soulier	26	fourchu	29	hauteur
34	chenet	26	poser	28	joint
31	sandale	24	arbre	27	chaussons
30	tête	24	mettant	27	main

Tableau 1 : les cooccurrents du lemme, et ceux des formes singulier et pluriel de la lexie

Sur les 20 premiers cooccurrents, 2 seulement sont sélectionnés en tête pour les deux types de contextes, singulier et pluriel de *pied* : *pointe* et *chaussé*. Mais la lecture des contextes révèle que la distribution ne doit rien au hasard et que des faits que l'on pourrait qualifier de "stylistiques", qui pour nous sont d'ordre sémantique au sens de la sémantique interprétative, président au choix des auteurs. Pour faire apparaître plus clairement la différence de distribution des cooccurrences, nous donnons dans le tableau 2, pour quelques lexies, la répartition en termes de rapport du nombre d'attestations dans le corpus de travail (les fenêtres de sélection autour de la lexie) au nombre d'attestations du corpus de référence (le corpus Roman), soit une partie des chiffres qui servent à calculer le score probabiliste<sup>4</sup>.

	cooccurrent de <i>pied(s)</i>	cooccurrent de <i>pied</i>	cooccurrent de <i>pieds</i>
<i>pointe</i>	446/2473	122/2473	324/2473
<i>chaussé</i>	69/199	24/199	45/199
<i>lit</i>	398/10534	307/10534	91/10534
<i>fouler</i>	103/226	15/226	88/226
<i>foulé</i>	56/119	14/119	42/119
<i>bot</i>	25/26	20/26	5/26
<i>étrier</i>	46/129	39/129	7/129

Tableau 2 : fréquences des lemmes cooccurrents dans les corpus de travail et de référence

<sup>4</sup> Ces chiffres apparaissent très clairement dans les listes et les tableaux synoptiques, à côté du score statistique, pour l'aide à l'interprétation (cf. documents du SAAS donnés en annexe du chapitre I).

**Sur la pointe du/des pied(s)**

Nous avons étudié les contextes d'association fournis par les concordances type "passages parallèles" du SAAS dont la plus grande partie figure dans les annexes du chapitre : les quelques exemples suivants montrent que la signification de l'expression *sur la pointe du/des pied(s)*<sup>5</sup> est, bien sûr, la même, si on envisage le "réfèrent", en utilisant nos "connaissances encyclopédiques" d'être humain, ou nos "connaissances d'univers", selon la terminologie consacrée.

a) *sur la pointe du pied*

. venue sur la POINTE du pied par une PORTE muette , sur des tapis de \*perse BALZ.cB1846  
à coucher et y entra sur la POINTE du pied , dans l' intention de se METTRE au LIT sans PONS.R1859  
à tourner autour de lui sur la POINTE du pied comme on MARCHE auprès de quelqu' un qu' on craint HUGO.m1862

b) *sur la pointe des pieds*

sur la POINTE des pieds , il alla vers l' ANTICHAMBRE , voulant s' assurer MONTH.pf36  
. doucement , sur la POINTE des pieds , il ouvrit la PORTE de la chambre , entra MART.Tmp29  
. elle allait sur la POINTE des pieds , PRENANT garde de ne faire aucun BRUIT ; mais ROL.JCad05

Ces extraits de textes viennent d'un tout cohérent et si on considère "le sens" en contexte, c'est-à-dire "ce que ces mots doivent engendrer dans l'esprit du lecteur", cette expression, sous ces deux formes, est employée pour signifier "avec discrétion, en prenant des précautions pour éviter qqc [réveiller qqn, se faire voir, etc.]"<sup>6</sup>. Mais c'est la contextualisation par le corpus qui donne la clé de la différence entre les deux formes : on peut observer que l'expression *sur la pointe du pied* est plus employée dans notre corpus dans les textes du XIX<sup>e</sup> s., jusque 1860 environ et qu'elle est très peu utilisée ensuite. Musset, Gautier, Stendhal n'emploient pas *sur la pointe des pieds*, tandis que Maupassant, Zola, Aragon, Queneau n'utilisent que cette expression. Il faudrait une étude détaillée pour comprendre tous les enjeux de cette distinction : *sur la pointe du pied*<sup>7</sup> est-il considéré comme plus "littéraire" que l'autre forme, ou bien l'influence du pluriel dans l'autre aire sémantique, c'est-à-dire dans le scénario "se hisser/dresser sur la pointe des pieds" se fait-elle sentir de plus en plus dans l'évolution de la "langue romanesque" ? En tout état de cause on peut voir ainsi que des raisons multiples, tenant au caractère pluri-sémiotique des textes, peuvent intervenir dans l'évolution des formes attestées, mais le lien au "réfèrent" n'est pas en cause<sup>8</sup>.

Les linguistes, comme les philosophes du langage se sont posé la question suivante : "pourquoi des auteurs emploient-ils au singulier un mot désignant une partie du corps qui est double, normalement chez l'être humain ?" Et c'est un des problèmes qui choque l'approche logique du langage et de la langue, marquée par l'impression référentielle<sup>9</sup>, qui souhaiterait une "langue parfaite", sans redondance ni "anomalie". Le concept de *classifiance-non*

<sup>5</sup> Ces unités figées ne rendent pas compte de tous les contextes de cooccurrence, mais elles sont très largement majoritaires dans les contextes (v. annexes) : au chapitre 2 nous avons observé qu'il y a plus de figements dans ce champ que dans la poésie.

<sup>6</sup> Parmi les lexies sélectionnées, on trouve : *précautions, bruit, regarder, (pieds) nus, endormi, (se) glisser, coller (son œil à), etc.*, v. les contextes en annexe.

<sup>7</sup> Cette forme semble plus ancienne, Littré en donne une attestation au XVIII<sup>e</sup> s. dans *La Nouvelle Héloïse* de J.-J. Rousseau (1761) et Frantext l'atteste, dans cette acception, depuis 1741. Mais l'interrogation différentielle des textes, en passant par les catégories pourtant assez frustes du système, permet de voir que les deux expressions appartiennent au roman (le pluriel étant 3 fois plus attesté que le singulier, ce qui s'explique en partie par ses deux aires sémantiques). En bonne philologie, il faut aussi vérifier quelle leçon est donnée dans la dernière édition revue par l'auteur, en particulier pour ceux qui, d'après notre corpus, emploient parfois les deux formes, Hugo, Ponson du Terrail, Flaubert, Céline, etc.

<sup>8</sup> C'est ce qui fonde l'autonomie de la langue et la possibilité de la description linguistique : on sait depuis Saussure qu'il n'est pas nécessaire, pour étudier les signes, de les rapporter aux opérations de pensée ou à la "réalité extérieure".

<sup>9</sup> V. chapitre 1 note 44.

*classifiance* a été forgé par J. C. Milner pour opposer l'emploi "comptable" et "logique", par rapport au référent, ici le pluriel, qualifié de classifiant à celui non-classifiant, que l'on considère généralement comme littéraire : "Le procédé qui consiste à recatégoriser un terme normalement comptable en substantif non-classifiant a été particulièrement apprécié des romanciers naturalistes. A sa manière, la littérature participait d'un mouvement d'ensemble qui, dans le code pictural, a pris le visage de l'"impressionnisme". Par ce procédé il s'agissait en effet de déplacer le centre d'intérêt de la description : l'essentiel n'est pas de découper le réel, mais de déployer un univers d'impressions"<sup>10</sup>. A ce propos, remarquons cependant que dans notre corpus, Zola n'atteste pas l'emploi non-classifiant : dès qu'on envisage le sens d'un signe en contexte, d'autres éléments d'interprétation que ceux issus de la logique et de l'impression référentielle se proposent à nous.

**le pied chaussé/les pieds chaussés**

a) le singulier

i)

de sa JAMBE fine et ronde et son JOLI pied CHAUSSÉ d' une PETITE PANTOUFLE bleue ... son beau BRAS SUE.A-G1831

POSA sur le TROTTOIR , comme craintivement , son pied CHAUSSÉ d' un PETIT SOULIER brillant , puis s' éloigna CHARD.e21

. elle FAILLIT METTRE son pied CHAUSSÉ d' une SANDALE - COTHURNE , les ONGLES tout TRI.pac1945

qui MAINTENAIT DROIT le COU , ASSISE , le pied très joliment CHAUSSÉ fixant au SOL , comme un CLOU VAL.dj1945

voyait circuler , à PETITS pas de leur PETIT pied , CHAUSSÉ de SOULIERS de toile , de SANDALES de VER.tm1873

ii)

. un pied CHAUSSÉ , les MAINS abandonnées , elle songeait maintenant ZOLA.pa1878

iii)

JAMBE TORDUE qui lui restait , terminée par un pied BOT et CHAUSSÉ d' une loque MAUP.cn884

. un pied , CHAUSSÉ d' un GROS SOULIER , parut enfin , MAUP.cn1887

b) le pluriel

ruban , TOMBAIENT sur ses éPAULES , et ses pieds NUS étaient CHAUSSÉS de SANDALES . ARL.o1929

. des pieds s' ALLONGEAIENT , CHAUSSÉS encore , BROYÉS et SAIGNANTS ZOLA.d1892

tourner son FAUTEUIL pour tendre à la flamme ses pieds CHAUSSÉS de feutre . MAUR.TD1927

nettoient dans les EAUX de vaisselle leurs pauvres pieds CHAUSSÉS d' une crasse squameuse ; d' autres se pouillent MILO.ai1910

de ses PETITES MAINS GANTÉES , de ses PETITS pieds CHAUSSÉS de BOTTINES à HAUTS TALONS , de sa GORGE ZOLA.PB1882

. ses pieds , CHAUSSÉS de BOTTINES LACÉES , DÉPASSAIENT le BORD de BOURGET.d89

. ses pieds étaient CHAUSSÉS de BOUE sèche . GON.mG1869

COIFFÉE de son bonnet BLANC , et ses PETITS pieds CHAUSSÉS de BRODEQUINS . MUSS.MP1845

il avait DURCIES , s' emmanchaient solidement ; ses pieds , CHAUSSÉS de GROS SOULIERS LACÉS , paraissaient forts , ZOLA.fr1871

inestimable au souvenir qui me restait de ses JOLIS pieds , CHAUSSÉS avec quelque recherche de PETITS BRODEQUINS TOE.ng1839

souvent plus de grâce que le naturel : ses pieds , finement CHAUSSÉS , étaient d' une EXIGUITÉ satisfaisante MURG.svb69

Dans ce corpus de romans, l'emploi dit "littéraire" de *le (son) pied chaussé* se rapproche plus de l'analyse de Maingueneau : plus précisément, c'est un ensemble de traits sémantiques /féminin/, /beauté/ /élégance/ qui caractérise les contextes autour du singulier. Dans les textes du corpus, on n'emploie pas le singulier pour les personnages masculins et on peut voir par les deux extraits de Maupassant, en iii), que le trait /masculin/ est au contraire attesté dans des

<sup>10</sup> Maingueneau D., 2000, p. 132.

contextes où les traits /beauté/ et /élégance/ ne sont pas validés<sup>11</sup>. Les contextes de *pieds chaussés* sont "simplement" descriptifs (y compris avec le trait /saleté/ qu'on ne trouve pas dans les exemples du singulier), et quand les auteurs emploient le pluriel dans la description d'une femme, l'adjectif *petit* qualifie *pied* pour propager les traits /beauté/ et /élégance/ du contexte (cf. les exemples de Zola et Musset), attestés également dans *joli* et *finement*<sup>12</sup>. On remarquera que dans l'exemple ii) la forme est différente, le signifié aussi : Zola emploie le déterminant *un* (= "un seul") dans un contexte qui signifie la distraction d'une femme "tourmentée d'une idée fixe" et qui s'interrompt dans sa toilette (*Une page d'amour*, 1878, p. 987).

### **Cooccurrence et phraséologie**

Les voisins sélectionnés en tête avec le singulier, *plain*, *valet*, *grue*, *coup* ne sont pas relevés près du pluriel, tandis que *tête* ou *chenet*, cooccurents de *pieds* ne voisinent pas de façon statistiquement valide avec le singulier. Quand l'association se rencontre avec les deux formes, on observe des différences importantes de score statistique : par exemple, le score de cooccurrence de *lit* avec le pluriel n'est que de 7, contre 81 pour le singulier, de *fouler*, de 16 avec le singulier au lieu de 93 (pour *foulé* on a 21 vs 61). De même, les scores de *bot*, et *étrier*, par exemple, sont d'un autre ordre de grandeur avec le pluriel : 15 vs 67, 9 vs 58. Trois textes seulement attestent le pluriel *pieds bots*, *César Birotteau* de Balzac, *Madame Bovary* de Flaubert (trois fois dont deux dans le même phrase), et *Bella* de Giraudoux<sup>13</sup>.

Pour *étrier* (v. les annexes), la répartition vient de l'alternance entre trait /physique/ ou /moral/ propagé par le contexte : on peut *avoir* ou *mettre le pied à l'étrier*, et, pour le francophone en tous cas, il n'y a pas d'ambiguïté dès lors qu'il a accès à un minimum de contexte. De même pour l'association avec *fouler*<sup>14</sup>, la répartition sémantique qui s'opère est celle-ci : si *fouler* voisine avec le pluriel, il s'agit du sens "mépriser, piétiner (au fig. pour la lexicographie, c'est-à-dire sur l'isotopie /moral/), alors que *pied foulé* (*se fouler le pied*) désigne une entorse, et que le sens "presser le sol en marchant", attesté dans des énoncés comme "la plus belle vallée que le pied de l'homme ait jamais foulée" et qualifié de "littéraire" par le dictionnaire (cf. TLF s.v. *fouler*), ne se trouvera que dans un nombre limité de contextes, au ton "noble", car l'expression est entrée dans "la langue" par la Bible et reste connotée ainsi.

<sup>11</sup> Le deuxième exemple met en scène un braconnier, caché sous un lit, qu'un gendarme tire en dehors de sa cachette.

<sup>12</sup> Nous verrons au chapitre 5 que le trait /élégance/ est souvent associé au trait /parisien/ dans le corpus Roman, comme dans cet énoncé des Goncourt (Charles Demailly, 1876, p. 22) : *Marthe avait de très-petits pieds, et les pieds d'une parisienne, de petits pieds remuants, coquets, presque spirituels.*

<sup>13</sup> *Larubanon, myope de l'œil droit et hypermétrope de l'œil gauche, légèrement bancal, délivré par la science à dix ans de deux pieds bots à cause desquels il avait déchiré toutes ses photographies d'enfant, était le fruit des amours cachées mais illustres d'un fondateur de la république et de cette cantatrice que ...* Giraudoux, *Bella*, 1926, p. 191 ; *Tout homme atteint d'un défaut de conformation quelconque, les pieds-bots, la claudication, les diverses gibbosités, l'excessive laideur, les taches de vin répandues sur les joues, les feuilles de vigne, l'infirmité de Roguin et autres monstruosités indépendantes de la volonté des fondateurs, n'a que deux partis à prendre : ou se rendre redoutable ou devenir d'une exquise bonté.* Balzac, *César Birotteau*, 1837, p. 147 ; *Il avait lu dernièrement l'éloge d'une nouvelle méthode pour la cure des pieds bots ; et, comme il était partisan du progrès, il conçut cette idée patriotique que Yonville, pour se mettre au niveau [en italiques dans le texte] devait avoir des opérations de stréphopodie. (p. 11) Redresser des pieds bots ! Est-ce qu'on peut redresser les pieds bots ?* Flaubert, *Mme Bovary*, 1857, p. 20.

<sup>14</sup> V. ci-dessous l'importance de ce verbe pour la forme *pieds*, dans le tableau 4.

Comme le montre le tableau 1, la phraséologie est bien mise en évidence dès qu'on ne lemmatise plus le pivot : *plain(-)ped, valet de pied, pied bot, mettre le pied à l'étrier, pied de grue, coup(s) de pied (au cul), pied de(du, de son, etc.) lit*<sup>15</sup>, (*frappant/frapper de la*) *pointe du pied, mettre/mettant pied à terre, des pieds à la tête (ou de la tête aux pieds)* sont bien représentés dans ce corpus Roman.

Les associations se repèrent encore plus clairement si l'on applique le test statistique séparément au contexte droit (les 10 signes après le mot pôle) et au contexte gauche (les 10 signes avant le mot pôle) ; dans les tableaux 3 et 4, les cooccurrents des colonnes 2 et 3 ne sont pas lemmatisés, il s'agit donc très souvent d'association figée entre telle forme et le singulier ou le pluriel de *pied*. Les scores sont plus importants que dans la sélection des colonnes 1 (où il s'agit de cooccurrents lemmatisés) : cependant, rappelons que le score de cooccurrence ne peut être le même puisque le volume du corpus diffère par le fait qu'on ne traite qu'une fenêtre de 10 signes pour chaque attestation de *pied* ou *pieds*.

<i>PIED</i>					
fenêtre 10 + 10		contexte droit (10 après)		contexte gauche (10 avant)	
score stat.	lemme	score stat.	forme	score stat.	forme
145	plain	107	bot	200	plain
88	valet	106	étrier	108	valet
65	bot	93	grue	75	coups
55	grue	63	lit	58	pointe
55	coup	55	fourchu	57	valets
50	étrier	49	pentamètre	56	frappant
44	lit	49	ergotée	41	coup
40	pointe	47	terre	40	mis
37	frappant	46	cul	35	mettre
37	terre	44	chaussé	34	mettant
36	mettre	40	arbre	29	frappa
35	pentamètre	35	mur	28	couvre
32	cul	34	astérior	27	mit
32	frapper	30	nu	27	poser
29	chaussé	29	pied	25	nasardes
28	fourchu	29	escalier	24	frappait
27	mis	28	égalité	24	pied
26	poser	26	gauche	22	brodequin
25	arbre	26	biche	23	arme
24	astérior	24	soulier	21	campèrent

Tableau 3 : la forme pied, son contexte gauche et son contexte droit

On observe que les scores vraiment atypiques se rencontrent avec les locutions qui constituent des unités de langue, celles dont on souhaite que le système automatique les reconnaisse pour les traiter comme telles. Une forme qui s'inclut presque exclusivement dans un syntagme figé se repère par son score élevé et sa présence seulement à droite ou à gauche de *pied* : 150 pour *pointe des pieds*, 200 pour *plain pied*, 108 pour *valet*, 353 pour *pieds nus*, etc.

<sup>15</sup> On peut lire dans les annexes le seul exemple de *pieds du lit* dans l'énoncé "*des pieds du lit où il était resté ...*" qui surprend chez Sainte-Beuve : cette lecture est conforme à l'édition saisie mais est-ce la bonne leçon ?

### CHAPITRE III : LES UNITÉS SÉMANTIQUES

Les sélections dans le contexte droit de  *pied*  concernent les entités verticales dont on désigne ainsi la base ( *pied de l'arbre, du mur, de l'escalier* ) ou les phraséologèmes<sup>16</sup> comme  *pied à l'étrier, pied d'égalité* , tandis que les contextes à gauche ont sélectionné plusieurs formes verbales :  *poser, mettre, mettant, mis, mit*  (qui se figent avec  *pied à terre*  ou  *pied à l'étrier* ),  *frappant, frappait, frapper*  en syntagme avec  *du pied* .

Pour le pluriel, on remarque que le contexte droit fournit des adjectifs ou participes qualifiant la partie du corps ( *nus<sup>17</sup>, chaussés, liés, joints, cambrés, gelés, endoloris, etc.* ), avec les types de chaussures ( *sandales, pantoufles* ), tandis que le contexte gauche atteste 10 formes de  *fouler* , ainsi que les verbes  *hausser, dresser, et traîner*  qui désignent des attitudes<sup>18</sup>.

PIEDS					
fenêtre 10+10		contexte droit 10 après)		contexte gauche (10 avant)	
score stat.	lemme	score stat.	forme	score stat.	forme
212	nu	353	nus	150	pointe
102	pointe	113	chaussés	83	fouler
89	fouler	82	joints	64	nu
60	foulé	77	liés	63	foulé
49	chaussé	70	pouce	49	plante
44	chenet	68	chenets	47	foula
42	sol	54	pantoufles	47	traînant
41	pantoufle	50	carrés	40	foulait
41	tête	46	bots	40	foulés
36	sandale	45	poings	36	mesurait
32	conculcatrice	45	cambrés	33	petits
31	traînant	44	sandales	31	hippopodes
31	soulier	43	enflés	31	foulerait
29	savate	42	mains	31	foulerai
29	hauteur	41	gelés	30	foulées
28	joint	41	endoloris	29	foulée
28	chaussons	40	sol	29	haussée
27	pouce	37	palmés	27	foulant
27	lié	34	saignants	26	dressa
27	main	32	meurtris	26	tête

Tableau 4 : la forme pieds, son contexte gauche et son contexte droit

On a vu que les différences de score statistique sont sous-tendues par des variations d'aires sémantiques<sup>19</sup>, notion liée à celle de "solidarités lexicales" (Coseriu) et qui peut correspondre parfois aux subdivisions des dictionnaires<sup>20</sup>. Comme le remarque plaisamment un fondateur de la sémantique linguistique : "Quand nous voyons le médecin au lit d'un malade, ou quand nous entrons dans une pharmacie, le mot  *ordonnance*  prend pour nous une couleur qui fait que nous ne pensons en aucune façon au pouvoir législatif des rois de France. Si nous voyons le mot  *Ascension*  imprimé à la porte d'un édifice religieux, il ne nous vient pas

<sup>16</sup>  *Ergotée, pentamètre, et astérion*  ne sont attestés que dans un texte.

<sup>17</sup> Quand  *nu*  est figé avec  *pieds* , c'est dans les mots-composés  *va-nu-pieds*  et  *nu-pieds* .

<sup>18</sup>  *Hippopodes*  n'est attesté que chez Flaubert, dans deux versions de La Tentation de Saint Antoine.

<sup>19</sup> Cf. Nyckees V., 1998, p. 193 : La polysémie (opposée à la monosémie) et l'homonymie concernent en effet ce qu'on peut appeler l' *aire sémantique*  d'un mot, c'est-à-dire l'ensemble des significations dont il est susceptible. On évitera en pareil cas d'utiliser l'expression de  *champ sémantique*  qui peut aussi désigner un ensemble de mots distincts partageant certains traits sémantiques communs.

<sup>20</sup> On a vu cependant que ce n'est pas toujours le cas, pour les sens "apeuré comme une biche" ou "résistant et gracile (comme une biche)" appliqués aux humains, au chapitre 2, ou pour les cooccurrents comme  *chaussé, étrier, fouler, etc.*

le moindre souvenir des aérostats, des courses en montagne, ou de l'élévation des étoiles. On n'a même pas la peine de supprimer les autres sens du mot : ces sens n'existent pas pour nous, ils ne franchissent pas le seuil de notre conscience"<sup>21</sup>.

L'application du test avec les différentes phases détaillées ici est une méthode pour repérer en premier lieu les figements d'un corpus, mais aussi les densités sémantiques d'origine diverse (normes de langue, de genre, d'époque, normes idiolectales<sup>22</sup>, etc.) et peut permettre d'améliorer un dictionnaire de formes importé d'une autre application pour l'adapter au corpus de la tâche. Les aires sémantiques conduisent vers ce que l'on nomme parfois des "scénarios" ou "scripts" qui sont le siège d'itérations sémantiques et de rôles complémentaires comme nous allons le voir ci-dessous.

### 3. 1. 2. Morphologie et sémantique

#### *Insertion du pivot dans un syntagme minimal*

Dans le paragraphe précédent, on a observé les sélections dans les cas de lemmatisation et non-lemmatisation du mot pôle. Qu'advient-il des cooccurrents si on inclut ce pivot dans un syntagme minimal, le type le plus fréquent en l'occurrence avec un substantif, soit N *de* ? Le tableau 5 illustre les sélections dans le voisinage des séquences '*ped(s) de*'.

<i>ped(s) de</i>		<i>ped de</i>		<i>peds de</i>	
score stat.	lemme	score stat.	lemme	score stat.	lemme
84	grue	108	grue	82	conculcatrice
67	lit	81	lit	36	opposite
51	conculcatrice	46	astérion	36	mesurer
36	astérion	42	arbre	30	argile
36	mur	41	mur	30	diamètre
34	arbre	38	égalité	29	célieri
29	égalité	32	escalier	28	bau
26	diamètre	29	mont	28	prosterné
24	escalier	28	montagne	27	profondeur
24	mont	27	croix	26	hauteur
24	prosterné	25	biche	24	longueur
23	montagne	24	muraille	24	large
23	biche	23	falaise	23	haut
23	muraille	22	rocher	23	jeter
23	mesurer	22	colline	23	triperie
22	opposite	21	échelle	23 r	alligator
22	célieri	21	autel	22	sol
22	croix	21	calvaire	20	foulées
21	échelle	20	assis	20	élévation
20	assis	20	contrefort	20	prosterner

Tableau 5 : les cooccurrents de la séquence-pivot *ped(s) de*

<sup>21</sup> Bréal M., Essai de sémantique, 1897, p. 145 cité dans Nyckees V., 1998, p. 198.

<sup>22</sup> Les grands auteurs se distinguent par des associations atypiques comme par exemple l'emploi de *ped de grue* chez Proust, qui détourne ce syntagme en le remotivant sémantiquement, l'enrichissant de traits propres à une acception de *grue* "femme entretenue, prostituée" et en l'insérant dans une apostrophe "*Grand ped de grue*", répétée trois fois (par Morel, La Recherche, La Prisonnière, 1922, p. 164). D'autre part, on a souligné aussi la présence de mots rares dans le corpus Roman et dans Frantext, par exemple dans des textes comme La Tentation de Saint Antoine. Ces informations sur la rareté sont importantes pour l'interprétation, c'est pourquoi on a fait en sorte que les documents du SAAS permettent toujours d'avoir sous les yeux en même temps la fréquence dans les deux corpus, pour relativiser une saillance importante ou au contraire ne traiter que les exemples rares d'association (cf. les extraits).

Au plan des cooccurrents attestés chez différents auteurs<sup>23</sup>, la répartition concerne d'une part des unités phraséologiques, comme *pied de grue*, *pied de (du, de son, etc.) lit*, *pied de l'arbre*, *du mur* (avec la variation : trait /physique/ ou /moral/), *de la muraille*<sup>24</sup>, *de la falaise*, etc., qui sont associées au singulier de la séquence-pivot et proviennent surtout des cooccurrents du contexte droit.

Pour les contextes autour du pluriel, on rencontre d'autres types de régularités : on a vu que *conculcatrice* se définit comme "celle qui foule aux pieds" et ce cooccurrent n'est donc associé qu'au pluriel, avec *foulées*<sup>25</sup>. La plus importante partie des cooccurrents réfère à l'isotopie /mesure/ car le sens "mesure ancienne" de *pied* est bien attesté dans ce corpus de romans (surtout dans des textes du XIX<sup>e</sup>s.) où les intervalles de temps représentés dans de nombreux textes évoquent la période antérieure à l'adoption du système métrique<sup>26</sup> : *mesurer*, *bau*, *diamètre*, *haut*, *hauteur*, *large*, *longueur*. On notera la présence d'un groupe de cooccurrents, *jeter*, *prosterné*, *prosterner*, qui introduisent une nouvelle aire sémantique, "qqn se prosterne ou se jette (/mouvement impétueux/) aux pieds de qqn".

### ***Extension de la séquence-pivot***

Puisque l'insertion dans un syntagme de type *N de* met en évidence des régularités, on peut faire l'hypothèse que l'extension de la séquence pivot devrait en rendre certaines encore plus apparentes : nous avons appliqué le test probabiliste aux contextes de la séquence *au pied de*, d'une part et de la séquence *aux pieds de*, d'autre part.

<sup>23</sup> *Opposite* et *astérion* ne sont attestés que chez un auteur, de même que *triperie* et *alligator* ; *opposite* n'est en fait attesté qu'une fois (dans des doublons).

<sup>24</sup> La séquence *au pied du mur* est représentée dans le corpus technique (avec les traits /physique/ et /moral/) mais l'association avec *muraille* ne l'est pas : elle est plutôt "littéraire". *Muraille*, à la différence de *mur*, comporte toujours le trait /physique/ dans notre corpus.

<sup>25</sup> Les formes *foulé* (score 14) et *fouler* (score 13) sont sélectionnées, cf. tableau 1 ; deux exemples de *foule*, substantif ont été retenus par le système.

<sup>26</sup> Légal en 1799, ce système n'a été rendu obligatoire qu'en 1840 ; dans certains textes, *pied* peut désigner la mesure toujours en usage dans les pays anglo-saxons, surtout dans le domaine nautique (J. Verne).

<i>AU PIED (de, etc.)</i>		<i>AUX PIEDS (de, etc.)</i>	
score stat.	lemme	score stat.	lemme
84	lit	135	conculcatrice
56	astérion	46	prosterné
48	arbre	40	jeter
46	mur	36	argile
38	escalier	34	foulées
34	mont	34	prosterner
33	croix	26	agenouiller
31	montagne	25	colosse
26	colline	22	déposer
26	falaise	21	cénacle
25	calvaire	20	hommage
24	contrefort	19	idole
24	autel	18	tabouret
23	échafaud	17	expliquant
23	rocher	15	tomber
22	assis	15	madone
22	mamelon	15	petitesse
21	muraille	15	vierge
19	camper	15	inhumain
19	chêne	14	chaussons

Tableau 6 : comparaison entre les corrélats sélectionnés auprès des deux pivots, au pied de et aux pieds de dans le corpus Roman.

Il n'y a pas d'intersection entre les deux corpus pour les 20 premiers lemmes sélectionnés par le test : par rapport aux corpus de travail précédents, on peut dire que l'insertion dans le syntagme formé avec les prépositions *à* et *de* introduit des régularités sémantiques, une densité que l'on n'avait pas encore rencontrées. L'analyse des exemples montre qu'il existe dans ces séquences une répartition d'ordre sémantique, propagée par le contexte : si le trait /singulier/ est attesté, alors on a récurrence des traits /verticalité/ et /localisation/ /au sol/, compatibles avec le sens "à la base de (qqc de vertical)", mais avec le trait /pluriel/, les contextes valident les traits /humain/ et /sentiment/, diffusés dans différentes lexicalisations.

Exemples du corpus Roman pour la séquence *au pied de* :

*Le gamin, accroupi* (/position près du sol/) *au pied* /partie inférieure/ *du parapet* /verticalité/, *s'affairait à manier sa pelle*. Martin du Gard R., Les Thibault, 1940, t. 2, p. 825

*Vers le soir, Kai-Koumou accosta au pied* /partie inférieure/ *des montagnes* /verticalité/ *dont les premiers contreforts* /verticalité/ *tombaient* /verticalité/ /->le bas/ *à pic* /verticalité/ /->le bas/ *sur d'étroites berges*. Verne J., Les Enfants du Capitaine Grant, 1868, p.110

*Le chien, logé dans un baril* /au sol/, *au pied* /partie inférieure/ *d'un pommier* /verticalité/ *penchant, jappait avec fureur, en tirant sur sa chaîne*. Maupassant G. de, Contes et Nouvelles, 1884, p. 152-153.

Exemples du corpus Roman pour la séquence *aux pieds de* :

*Je suis écrasé d'épouvante. Si je me laissais aller, j'irais m'abattre et m'humilier pitoyablement aux pieds de mes bourreaux ... du courage, j'en aurai. Mais je ne sais pas ce que je ferais si une lâcheté pouvait me sauver à cette heure... qu'on ne fasse pas de moi un héros plus tard*. Van der Meersch M., L'Invasion 14, 1935, p.101.

*... cet abandon entier, cette désolation franche touchèrent Ra'hel, Tahoser s'avouait vaincue, et implorait sa pitié (...)* *Tahoser s'agenouilla aux pieds de Ra'hel et lui baisa la main ; Ra'hel la releva et lui entoura amicalement le corps d'un de ses bras*. Gautier Th. Le Roman de la Momie, 1858, pp. 296-297.

... une vie de **froid égoïste**, une vie de prêtre, **sans amour et sans femmes**. Et pourtant **le coeur me bat** dans la poitrine ... et la vue d'une femme me fait **tressaillir**, j'éprouve un **immense besoin de souffrir**, de **pleurer aux pieds** d'une femme, je n'ai plus de mère, moi ! ... **seul, isolé**, il faut bien que **j'aime** quelqu'un... qu'une bouche de femme **me console ou me plaigne**. Sue E., Atar-Gull, 1833, p. 31.

Bientôt la malheureuse vit, à travers un nuage qui se répandait sur ses yeux, approcher le brodequin, bientôt elle vit son pied emboîté entre les ais ferrés disparaître sous l'effrayant appareil. Alors **la terreur** lui rendit de la force. -ôtez-moi cela ! Cria-t-elle **avec emportement**. Et, se dressant **tout échevelée : -grâce !** Elle s'élança hors du lit **pour se jeter aux pieds du procureur du roi**, mais sa jambe était prise dans le lourd bloc de chêne et de ferrures, et elle s'affaissa sur le brodequin, plus **brisée** qu'une abeille qui aurait un plomb sur l'aile. Hugo.V, Notre-Dame de Paris, 1832, page 362

Ainsi, et malgré l'enseignement de la grammaire, dans le roman, *aux pieds de* n'est pas "le pluriel" de *au pied de*, car, comme nous l'avions observé avec les séquences *pied(s) chaussé(s)* ou *pointe du/des pieds*, le trait de nombre entre dans une *structure*<sup>27</sup> avec alternance d'un trait qui entraîne des régularités thématiques. Comme ces résultats s'interprètent dans l'unité sémantique sur-ordonnée au syntagme, un énoncé ou ensemble d'énoncés développant le même thème, que la rhétorique nomme *période*, nous allons les développer en détail.

---

<sup>27</sup> Ce terme est à prendre au sens de la linguistique structurale: «est *structure* une entité autonome de dépendances internes (...) *Structure* s'emploie ici pour désigner, par opposition à une simple combinaison d'éléments, un tout formé de phénomènes solidaires, tels que chacun dépend des autres et ne peut être ce qu'il est que dans et par sa relation avec eux.» Hjelmslev L., 1957, p. 109.

### 3. 2. UNE FONCTION NARRATIVE : L'IMPLORATION

#### 3. 2. 1. L'alternance /humain +/ /humain -/

Alors que les énoncés comportant la séquence *au pied de* réfèrent à la verticalité (/localisation/, /au sol/), ceux qui comportent *aux pieds de* sont indexés sur une isotopie générique /sentiment/ et lexicalisent le sentiment d'humilité, qui s'intègre à des scènes de pardon, de reconnaissance, d'aveu, etc.<sup>28</sup> Ce qui est régulier, c'est l'existence, dans ces contextes, d'un sème afférent, propagé par le contexte, issu de normes socialisées : si /position basse/ (pour ego<sup>29</sup>, près de X), alors /humilité/ et ce sème est compatible avec le signifié des lemmes sélectionnés en tête, comme *prosterné-prosterner, hommage, idole, pardon, sauveur*, etc., ainsi qu'avec le trait /perfectif/ validé dans le sémantisme des verbes et le temps grammatical, le passé simple, largement dominant dans les exemples. On peut voir dans le tableau 6 que le test a sélectionné en tête, avec *aux pieds de* autant de verbes que de substantifs qui se rapportent à la fonction narrative 'imploration' (verbes indiquant une action, un mouvement ponctuels), alors que près de la séquence au singulier on trouve une large majorité de substantifs qui forment un syntagme avec le pivot et seulement le participe passé *assis*, duratif et compatible avec une description de localisation. Comme le montrent les contextes classés par ensemble de traits discriminants dans l'annexe, l'agent dont on précise la position à la base de qqc peut comporter ou non le trait /humain +/, alors que dans les contextes de *aux pieds de*, l'acteur du cas ergatif est toujours animé, surtout humain (à 95% environ, quelquefois animal<sup>30</sup>), comme l'acteur du cas locatif.

Les contextes de description locale se présentent sous deux formes<sup>31</sup> :

1) : X /humain+/ Procès + *au pied de* /humain-/

Les agents sont lexicalisés par des noms propres, des pronoms personnels, des mots comme *les blessés, une négrillonne, un prêtre*, etc.

Les procès sont : *aller, s'arrêter, s'allonger, arriver, camper*, etc.

Pour le cas locatif, on rencontre : *arbre* (et les "sortes de" : *manguier, eucalyptus, figuier, pin, saule, érable*, etc.), *autel, butte, colline, cône, crucifix, côte, croix, escalier, fosse, falaise, gibet, haie, maison, mât, mur, montagne, monticule, monument, tombeau, tour, roc, rocher, dtatue, trône, échafaud, échelle, église, éminence* ...

2) : X /humain-/ Procès + *au pied de* /humain-/

Ce cas concerne principalement des descriptions de lieux, naturels ou construits par l'homme : une *cabane* à la base de la *falaise*, une *citée*, un *village* ou une *ville*, localisés au pied d'une *montagne*, de *contreforts*, d'une *colline*, etc. Plus rarement, il s'agit d'objets comme un *haillon* décrit au pied d'un *arbre*.

<sup>28</sup> Une grande partie des exemples est donnée en annexe, avec la distribution selon les traits discriminants.

<sup>29</sup> Nous empruntons à Hagège C. 1982 p.100 la notion d'*ego*, comme centre de déixis.

<sup>30</sup> Nous verrons que ce trait est compatible avec la structure repérée.

<sup>31</sup> Nous classons à part (cf. les annexes) les exemples de *au pied du lit* (qui atteste les traits /localisation/ et /à l'opposé de la tête du lit/) et *au pied du mur* (au figuré) que nous considérons comme unité d'un autre type d'aire sémantique à cause du trait /moral/.

Ces régularités expliquent la sélection massive de substantifs pour ce corpus, ainsi que des verbes liés à l'action humaine vers ou dans un lieu particulier. On observe un autre fait dont l'explication ne vient pas tout de suite à l'esprit, car il s'agit, comme dans le cas de *piéd(s) chaussé(s)* d'une opposition entre les deux types de structures de contextes : le pluriel de *au piéd d'un arbre (d'un(e) montagne, colline, contrefort, maison, rempart, etc.)* n'est pas *aux piéd des arbres* mais *au piéd des arbres* (etc.). Là encore, on peut se demander pourquoi "la langue" n'utilise pas une régularité prescrite par la grammaire, mais c'est un fait avéré dans le corpus et auquel un seul exemple contrevient<sup>32</sup>. La raison en est sémantique : les traits de *piéd* "base [de qqc ayant le trait /humain -/]" ne sont pas identiques à *piéd* employé à propos d'un humain, le trait /partie du corps/ n'étant pas valide, et même si le trait /extrémité/ est retenu, c'est une notion de lieu qui est signifiée.

### 3. 2. 2. Logique sémantique et structure

#### *Le schéma actantiel de la fonction 'imploration'*

Le classement des contextes de la séquence *aux piéd de* permet de faire ressortir les traits discriminants, qui sont des traits afférents, propagés par le contexte, et dont on peut être amené à rechercher parfois la première mention dans le texte à une longue distance car elle concerne le rapport entre deux personnages, leur rôles respectifs dans le "scénario" : X ressent un sentiment d'humilité par rapport à Y, puisqu'il va se mettre dans une posture de supplication, de prière vis-à-vis de lui. Parallèlement, un trait /puissance sur X/ est donc attribué à Y, et ce trait peut être valide dans tous les univers (le roi, l'évêque, le juge sont des "puissants de la terre"<sup>33</sup>) ou seulement dans l'univers d'un acteur et dans un "épisode" particulier, comme la maîtresse, *l'inhumaine* qui refuse ses faveurs<sup>34</sup>.

La plus grande partie des énoncés se classe donc selon ce schéma :

X /humain +/ /sentiment d'humilité/ Procès = implorer Y /humain +/ /puissance (sur X)/

Ces exemples sont responsables de la sélection par le test de *cénacle, idole, inhumain* (cf. tableau 6) dans les premiers cooccurrents mais aussi de *sauveur, majesté, déesse, maîtresse*, (scores : de 13 à 8) par exemple.

Peu d'énoncés attestent le pluriel pour l'entité Y, et dans ce cas, bien sûr, la séquence ne change pas : on trouve dans le corpus *aux piéd des officiers, des saints en haillons, des Hubert*.

*Madone, vierge, agneau, ange* (score : 7) s'insèrent dans des énoncés où le trait /humain/ de Y est remplacé par /divin/

Les variantes les plus représentées peuvent se symboliser ainsi :

X /humain +/ Procès /respect, offrande, etc. / Y /humain +/

<sup>32</sup> C'est l'exemple de Fromentin, commenté ci-dessous en 3. 2. 2. 2. (*fouler aux piéd*). Une coquille est toujours possible et, comme nous l'avons dit, la vérification du texte dans une édition faisant foi s'impose ; cependant nous remarquerons qu'il y a toujours une petite proportion d'écarts, les régularités allant de 80 à 95 %. C'est peut-être un jeu nécessaire pour que la langue continue à rester vivante, et à évoluer grâce aux "trouvailles" individuelles.

<sup>33</sup> On a vu au chapitre 2 que des extraits de textes de sciences humaines attestaient ces lexicalisations de la fonction narrative : *la demande mise par l'avocat \*labori aux piéd de sa majesté ; des esclaves nu-tête aux piéd du roi*.

<sup>34</sup> Les concepts de rôles, de monde et d'univers qui concernent l'analyse du texte selon la dialectique et la dialogique (v. glossaire) seront explicités et illustrés aux chapitres 4, 5 et 6.

X /humain +/ *Procès (attitude de proximité, d'intimité respectueuse)* Y /humain +/  
Même s'il ne s'agit plus à proprement parler de scènes d'imploration<sup>35</sup>, le trait /respect/ est toujours validé et la relation entre X et Y entraîne une sorte de "puissance" de Y, reconnue par X, dans le rôle de mère, grand-mère ou d'amante.

#### Idole / idoles

Quand l'entité Y est qualifiée ainsi, on rencontre deux cas différents :

- soit *idole* comporte les traits /objet/ et /inanimé/

et dans ce cas, que le mot soit au singulier ou au pluriel, il est introduit par la séquence *au pied de*, conformément à ce que nous avons vu ci-dessus

- soit *idole* comporte les traits /animé/ et /humain/

dans des scénarios où l'être aimé est ainsi nommé, étant donnée l'adoration que lui voue X, et on retrouve donc la séquence *aux pieds de*, avec les traits repérés pour cette structure.

Deux exemples seulement semblent déroger à cette norme, mais un contexte plus large permet de comprendre l'apparente contradiction.

On lit dans *Lélia* (p. 170) :

*Mais ne sachant où dépenser ma vigueur intellectuelle, je la jetai rampante et tenace au pied d'une idole créée par mon culte ; car c'était un homme semblable aux autres, et, quand je fus lasse de me prosterner, je brisai le piédestal et je le vis réduit à sa véritable taille. Mais je l'avais placé si haut, dans mes pompeuses adorations, qu'il m'avait paru grand comme Dieu.*

*Lélia* (le foyer énonciatif) oppose le caractère d'artefact (= objet *créé*) de l'être chargé par son "culte", à l'homme réel dont elle découvre ensuite "la véritable taille", après avoir brisé le *piédestal* (/objet/). L'homme est opposé à l'idole, car il n'est pas digne des sentiments d'adoration que *Lélia* lui a voués : il ne s'agit donc pas d'un cas de sentiments respectueux envers un acteur ayant le trait /humain +/ comme dans les énoncés avec la séquence *aux pieds de*.

Dans *Eugénie Grandet* (pp. 228-29-30), le narrateur fait allusion aux habitudes que prend peu à peu l'héroïne éponyme d'être entourée d'une cour qui la flatte et l'adule :

*Chez elle, tous les soirs, la salle se remplissait d'une société composée des plus chauds et des plus dévoués Cruchotins du pays, qui s'efforçaient de chanter les louanges de la maîtresse du logis sur tous les tons. Elle avait le médecin ordinaire de sa chambre, son grand aumônier, son chambellan, sa première dame d'atours, son premier ministre, son chancelier surtout, un chancelier qui voulait lui tout dire. L'héritière eût-elle désiré un porte-queue, on lui en aurait trouvé un. C'était une reine, et la plus habilement adulée de toutes les reines. La flatterie n'émane jamais des grandes âmes, elle est l'apanage des petits esprits, qui réussissent à se rapetisser encore pour mieux entrer dans la sphère vitale de la personne autour de laquelle ils gravitent. La flatterie sous-entend un intérêt. Aussi les personnes qui venaient meubler tous les soirs la salle de Mademoiselle Grandet, nommée par elles Mademoiselle de Froidfond, réussissaient-elles merveilleusement à l'accabler de louanges. Ce concert d'éloges, nouveaux pour Eugénie, la fit d'abord rougir ; mais insensiblement, et quelque grossiers que fussent les compliments, son oreille s'accoutuma si bien à entendre vanter sa beauté, que si quelque nouveau venu l'eût trouvée laide, ce reproche lui aurait été beaucoup plus sensible alors que huit ans auparavant. Puis elle finit par aimer des douceurs qu'elle mettait secrètement **aux pieds de son idole**. Elle s'habitua donc par degrés à se laisser traiter en souveraine et à voir sa cour pleine tous les soirs*

Comme il s'agit de la représentation d'elle-même que se forge peu à peu Eugénie, la séquence *aux pieds de* rend compte du trait /humain +/.

<sup>35</sup> V. les contextes du verbe *se prosterner*, le trait /humilité/ est toujours attesté.

*Les autres associations***Fouler aux pieds**

Au plan de la forme linguistique, ce type est différent puisque *X foule aux pieds* se complète en principe par un cas accusatif comme dans : *on foulait aux pieds des lambeaux de chair*. Remarquons que l'agent ergatif de la construction passive (*je fus foulé aux pieds des chevaux*) peut occuper la même place que le cas accusatif : mais, en l'absence d'un analyseur morpho-syntaxique, le traitement automatique sélectionne ces contextes, avec ceux du schéma *X est aux pieds de Y*, parce qu'ils comportent la séquence *aux pieds de*.

Dans ce type de schéma, soit on rend compte de la position /au sol/ d'un acteur piétiné, soit c'est le trait /mépris/ qui est propagé sur l'isotopie /moral/: cependant, on observera qu'il existe un lien entre ces deux types et la structure de la fonction narrative 'imploration' : le trait /infériorité/ de *X/* par rapport à une entité qui a le pouvoir de le piétiner ou le mépriser.

On peut rattacher à ce type le seul exemple de *conculcatrice* "celle qui foule aux pieds, qui méprise", attesté dans le syntagme *jeter (qqc) aux pieds de la Vierge conculcatrice* (dont on a donné le texte développé au chapitre 2) : ce mot reçoit comme cooccurrent de la séquence pivot *aux pieds de* son score le plus important, 135 (contre 23 comme cooccurrent du lemme *pied(s)*, 32 avec la forme *pieds*, et 82 avec la séquence *pieds de*).

Dans l'exemple suivant :

*Deux femmes se dirigeaient vers nous, au bord de l'allée (...) L'une était jeune et remarquablement jolie (...) J'observais cette façon légère et contenue de fouler à petits pas le gazon qui s'étendait aux pieds des arbres, comme si elle eût marché sur les laines souples d'un tapis.* Fromentin, Dominique, 1863, p. 90

on remarque que contrairement à ce qui est réalisé dans le corpus, l'auteur met au pluriel la séquence pour une entité /humain -/ : *aux pieds des arbres* et non *au pied des arbres*. C'est le seul exemple, et on s'aperçoit qu'il se trouve en lien avec *fouler* au sens de "marcher" (qui hérite ici d'un trait /souplesse/ diffusé par le contexte). Si le texte est bien conforme à ce qu'a écrit l'auteur, on peut faire l'hypothèse que le trait /pluriel/ est venu plus loin, et comme par lapsus, pour le mot *pied*, sous la plume de l'écrivain qui venait d'employer *fouler*, parce que leur proximité évoquait le type figé<sup>36</sup>.

**Colosse aux pieds d'argile**

Cette expression figée est à l'origine de différentes sélections de *colosse*, ce qui montre encore que le test statistique met en évidence la densité sémantique : son score est de 25 avec *aux pieds de*, de 17 avec *pieds de*, de 11 avec la forme *pieds*, et de 10 avec le lemme *pied(s)*<sup>37</sup>. Au plan sémantique, la relation entre *pieds* et *colosse* est de type attributive, c'est une prédication intégrée, et même une argumentation de type enthymème : "c'est un colosse (trait /force/) et pourtant il a des pieds fragiles (*d'argile* /fragilité/), c'est-à-dire : "il y a une faille dans la cuirasse". On peut voir que quand une expression est bien connue et que le contexte

<sup>36</sup> La linguistique n'a pas pour objet les processus psychologiques mais elle ne peut les oublier ; dans l'activité sémantique les processus conscients et subconscients coexistent, comme les lapsus, et les jeux de mots spontanés le laissent penser.

<sup>37</sup> Son score est de 12 dans le "contexte gauche" de *pieds*, mais on ne peut comparer ce résultat aux autres puisqu'ici la fenêtre n'est que de 10 signes.

n'est pas ambigu, on peut y référer en la tronquant, ce que fait Balzac, dans *succès (...) aux pieds d'argile*.

#### **Le trait /animal/**

En ce qui concerne les acteurs comportant le trait /animal/, on remarque qu'il s'agit en majorité de chiens, et dans les contextes les traits /au sol/+soumission à Y/ sont validés ; dans la phraséologie, *vie de chien* réfère à la pénibilité, aux mauvais traitements que supporte cet animal, à sa position d'infériorité, et on parle aussi de sa fidélité à toute épreuve, des "bons yeux" de chien fidèle et soumis.

Un exemple concerne une chèvre qui peut être considérée dans ce contexte comme un animal domestique, un animal de compagnie (la chèvre "semble *penser*") : la scène décrite est pastorale, idyllique, et les animaux qui entourent le tailleur de pierre Claude pendant son sommeil (son chien, la chèvre avec ses trois chevreaux, quelques moutons et des lapins) forment avec lui une "*petite colonie*".

*Les grands yeux de la mère [la chèvre], vagues, lointains et rêveurs comme les yeux de la gazelle et du chameau, semblaient penser. Ils se portaient tour à tour du maître à ses petits, du chien aux moutons, des roches à l'herbe, comme si elle eût rassemblé voluptueusement dans son regard tout ce tableau de paix dont elle faisait partie. (...) On eût dit qu'un esprit de douceur et d'amitié avait mis la confiance et la paix entre toutes les choses et entre tous les êtres de cette petite colonie de la montagne. Lamartine, Le tailleur de pierre, 1851, pp. 418-419*

#### **X jette Z aux pieds de Y**

On a vu dans les exemples que, en fait, le type le plus fréquent consistait en propositions de type pronominal réfléchi : *X se jette, se prosterne*, etc. Mais quand le verbe n'est pas à la forme pronominale, le sémantisme n'est plus celui de la structure définie, même s'il ne s'écarte pas complètement de sa "logique sémantique". Si X met aux pieds de Y un acteur comportant le trait /animé/, que ce soit un humain ou un animal, il fait une offrande à Y, ce qui fait sens dans leurs rôles respectifs : chez Ponson du Terrail, X est Léon Rolland, contremaître qui capture le bandit Rocamboles (encore jeune vaurien) et l'amène devant le Comte de Kergaz pour qu'il avoue où se trouvent deux jeunes filles enlevées. Quand à l'exemple de Maupassant, il s'agit de la nouvelle intitulée *Le Loup*, racontant comment deux frères passionnés de chasse s'obstinent à traquer un loup particulièrement malin : le plus jeune tuera à son tour le loup qui a tué son frère, pour l'offrir à la dépouille de l'aîné. Dans ces deux cas, Y est supérieur à X.

Si Z comporte le trait /inanimé/, il s'agit de scènes où X dépose qqc au sol devant Y, et ce schéma n'est d'ailleurs que peu représenté dans le corpus.

#### **Autres contextes**

Cette subdivision (11 dans les annexes) ne comporte que peu d'exemples : le type *avoir, porter aux pieds* des [sortes de chaussures] est représenté deux fois, un énoncé atteste *aux pieds de l'homme* dans la position cas attributif, régie par *préparer* ; des décorations se trouvent *aux pieds de* meubles (lit, fauteuil) et le pluriel signifie "à chaque pied" en tant que support. Dans un énoncé, on trouve l'expression *précipiter qqn aux pieds de son juge* qui a le même sens que *traîner aux pieds des tribunaux* (où le trait /humain/ à propos des juges explique le pluriel de *pieds*).

Nous avons mis en évidence que des aires sémantiques différentes sont engendrées par l'alternance entre trait /pluriel/ et /singulier/ et trait /humain +/, /humain -/, parce que cette variation entre dans une structure. Si le même signe peut être attesté dans des aires sémantiques variées, c'est parce que les composantes (les traits sémantiques) de son signifié sont différentes, comme le montrent ses relations avec les autres signes du contexte. A ce titre, rappelons que même si les signes sont le point de départ obligé de la linguistique<sup>38</sup>, ils ne sont pas son objet dernier, la priorité est donnée aux relations plus qu'aux termes : depuis Saussure on sait que le système linguistique est un "tout en soi" et un "principe de classification", et que la valeur de chaque signe est fonction des valeurs de tous les autres.

Dans l'analyse des différentes aires sémantiques représentées autour des séquences *au pied de/aux pieds de*, on a vu se dégager une "logique sémantique", qui nous semble au fondement de la fonction de code de la langue. Au plan cognitif<sup>39</sup>, la régularité, c'est que si X humain est aux pieds de Y humain, on se situe sur un axe marqué par les traits /respect/, /humilité/ et /puissance/, que les contextes attestent, dans des proportions variées, de l'attitude intime et respectueuse, à l'imploration pour avoir la vie sauve. Les autres associations repérées autour des séquences *au pied de* et *aux pieds de* vont dans le même sens, que ce soit l'emploi de *jeter* qqc ou qqn devant qqn en signe d'offrande, la répartition des exemples d'*idole* selon le trait /humain/, comme le trait /soumission/ pour le chien près de son maître. De même, celui qui est foulé aux pieds est en position inférieure, a le trait /humilié/ si l'exemple se situe sur l'isotopie /moral/, et dans un certain nombre de contextes, si un acteur ayant le trait /humain +/ est au pied du lit de qqn, un trait /respect/ est propagé, comme dans une scène d'adieux à un mourant<sup>40</sup>. Dans l'approche sémiotique qui est la nôtre, on dira que l'ensemble de traits /humain +/ + /au sol/ + /humilité/ sert d'*interprétant* à différentes lexicalisations : "les faisceaux microsémantiques et microphonémiques qui fondent pour une part la textualité ne se laissent pas résumer par des mots : ils n'ont de nom dans aucune langue"<sup>41</sup>.

### 3. 2. 3. Les acteurs d'une fonction narrative

Nous avons vu que, dans l'interprétation des textes électroniques, il était important de procéder par étapes, parce que des éléments à vérifier émergent de la contextualisation par le corpus, comme la relativisation d'un score important par le fait qu'un signe est un hapax, ou des régularités liées à une époque. Notre travail se situe, d'autre part, dans le cadre d'une réflexion sur un concept de "station de travail" à fournir aux usagers des textes électroniques. Nous avons voulu vérifier si l'on pouvait aller plus loin dans l'emploi du test pour la recherche de différents types d'unités sémantiques.

<sup>38</sup> Pour nous, la linguistique est une partie de la sémiotique, c'est la sémiotique de la langue et des textes.

<sup>39</sup> Si des expressions signifiant l'imploration employaient alternativement *pieds* ou *tête*, la logique sémantique serait moins fondée et on aurait du mal, nous semble-t-il, à acquérir *le sens* d'une telle langue ; nous verrons ci-dessous qu'avec *jeter qqc à la tête de qqn*, c'est le trait /agression/ qui domine.

<sup>40</sup> C'est le cas de plusieurs exemple de la section *au pied du lit*, des variantes *au pied du divan* et *du matelas*, également. Avec *au pied de la croix* on peut se trouver dans des contextes de vénération, sur l'isotopie /religion/.

<sup>41</sup> Rastier F., 1989, p. 28.

*Variation de l'empan de la fenêtre*

Puisque dans les textes, la longueur de la période consacrée à la fonction narrative 'imploration' que nous avons mise en évidence peut être très variable, on a effectué le traitement statistique dans des fenêtres de sélection variant de 5 mots (avant et après le pivot), à 40 mots, afin de vérifier la consistance des résultats.

Les résultats figurant dans le tableau 7 montrent que, parmi les cooccurrents sélectionnés par le test en tête, il y a persistance de certaines sélections, quelle que soit la taille de la fenêtre : des mots particuliers désignent les acteurs de cette fonction narrative, dans le genre du roman, dans des lexicalisations validant le trait /puissance/ et cette régularité sémantique explique les résultats du test. Les contextes élargis de *au pied de* offrent toujours, comme cooccurrents sélectionnés, le même type d'entités marquées par le trait /verticalité/ (tableau 8).

On a choisi, pour les programmes du SAAS, d'éliminer de la sélection les noms propres : ceux-ci lexicalisant également les acteurs figureraient en tête des résultats, mais on ne peut les interpréter si l'on ne connaît pas les traits sémantiques attribués à tel "personnage"<sup>42</sup>, par le narrateur et aussi par la parole représentée des acteurs avec lesquels il est en interaction. L'utilisateur des programmes du SAAS aura le choix, par option, de les conserver ou de les traiter à part.

<i>Lemme</i>	<i>Aux pieds de + 5</i>	<i>Aux pieds de + 10</i>	<i>Aux pieds de + 20</i>	<i>Aux pieds de + 40</i>
abbé	5	3		
déesse	13	9	7	6
inhumain	22	16	12	10
majesté	16	11	11	11
mari	3			
maître	8	6		5
maîtresse	8	9	8	7
mère	8	7	5	5
père	4		4	
prêtre	5	5	3	3
reine	4	4	7	6
roi	9	8	8	7
saint	6	6	4	3
sainteté	11	8	6	10
vierge	18	16	15	12
évêque	9	6	6	5

Tableau 7 : les lemmes qui lexicalisent, dans les occurrences du Roman, les acteurs de la fonction narrative "imploration" et leurs scores statistiques.

Autour de la séquence *au pied de*, les résultats restent conformes à ce qui avait été observé et s'expliquent par la récurrence des traits /verticalité/ et /au sol/.

<sup>42</sup> V. chapitres 4, 5 et 6.

<i>Lemme</i>	<i>Au pied de + 5</i>	<i>Au pied de + 10</i>	<i>Au pied de + 20</i>	<i>Au pied de + 40</i>
lit	125	91	71	60
astérion		61	47	
arbre	48	38	31	29
escalier	53	39	30	26
mont	33	25	19	16
croix	45	35	28	25
montagne	27	25	20	19
colline	19	18	15	13
falaise	24	21	16	15
calvaire	39	28	21	18

Tableau 8 : scores statistiques des premiers lemmes associés à au pied de dans le corpus Roman, avec différents empans de sélection .

Cette expérience montre la consistance des résultats : les mêmes lemmes continuent à être sélectionnés, même si leur score diminue au fur et à mesure que le corpus de travail augmente en nombre d'occurrences. Les tableaux synoptiques marquant la fréquence dans chaque sous-corpus, on peut voir d'autre part que celle-ci ne varie qu'assez peu pour les mots sélectionnés, même si la fenêtre de travail s'agrandit, ce qui est lié à la densité sémantique et prouve qu'on reste dans les mêmes aires sémantiques.

#### *Variation du pivot*

On peut voir dans les exemples de l'annexe que le verbe *jeter*, en construction pronominale, est celui qui est le plus fréquemment attesté pour lexicaliser l'imploration, souvent au passé simple ou au participe présent qui valident le trait /perfectif/. Une rapide étude historique semble conforter l'idée que *qqn se jette aux pieds de qqn* pourrait être le syntagme en quelque sorte "canonique" à partir duquel les autres se seraient développés, au fil du temps et des textes : il est le plus anciennement attesté, dans l'un des premiers textes d'ancien français, *La vie de Saint Léger* (2<sup>e</sup> moitié X<sup>e</sup>s.)<sup>43</sup>.

Comme il constitue une variante de lexicalisation de l'imploration, on a voulu vérifier si, dans ses contextes, le test permettait de repérer les régularités structurelles. Le traitement statistique a été appliqué au sous-corpus constitué des énoncés comportant le verbe *jeter* pris comme mot-pivot (avec l'ensemble des formes représentées) à la condition que la séquence comprenne *aux pieds de (du, d')* dans le contexte droit, sans mot intercalé. Les exemples ci-dessous sont issus de ce corpus (de volume réduit, 643 occurrences) dans lequel les mots sélectionnés par le test sont : *pied* (puisque'il n'est pas en position de pivot), *supplier*, *sanglot*, *maîtresse*, *mari*, *crier*, *père* .

*Dona Isabelle prit d'abord ce langage pour une plaisanterie ; mais, quand elle vit que rien n'était plus sérieux, elle éclata en sanglots, se jeta aux pieds de son mari et le supplia de renoncer à un projet aussi déshonorant.* Gobineau J.-A. de, *Les Pléiades*, 1874, p. 216

*Elle (la marquise de Chamarante) jeûna, passa des heures en prières, s'abîma gravement les genoux ! Enfin, comme la retraite touchait à sa fin, elle se jeta aux pieds du moine et le pria de disposer de sa vie selon la volonté de Dieu qu'il connaissait, cela va sans dire.* Boylesve R., *La leçon d'amour dans un parc*, 1902, p. 244

<sup>43</sup> En latin sont attestés les syntagmes *ad pedes alicujus* ou *alicui accidere, procidere, jacere, se abjicere, se projicere, se prosternere*.

*Belle de douleur, les yeux brillants de larmes, les poignets ceints de lapis lazuli et comme enchaînée par sa foi, tout à coup elle se jeta aux pieds d'Hippolyte et lui embrassa les genoux, mourante, échevelée. Il céda presque.* France A., L'île des pingouins, 1908, p. 351

- *Je vais aller me jeter aux pieds de la reine, répondit la jeune fille, et nous verrons si elle sera sourde aux prières de deux enfants qui demandent la vie de leur père.* Verne J., Les enfants du Capitaine Grant, 1868, p. 35

- *Ami Thalcave, répondit Robert en se jetant aux pieds du Patagon, ami Thalcave, ne nous quitte pas !* Verne J., Les enfants du Capitaine Grant, 1868, p. 182

Ce traitement a mis également en évidence la régularité au niveau "formel" : en donnant au système la contrainte de sélection "avoir *aux pieds de* (du, d') dans le contexte droit", la machine n'a sélectionné avec *jeter* comme pivot que des énoncés où *jeter* est en construction pronominale, et ils consistent en scènes d'imploration<sup>44</sup>, comme nous l'avons observé dans l'étude de la structure.

L'étude sur la variation de la taille de la fenêtre a été menée également sur ce corpus, apportant le même type de résultats que ceux observés autour de la séquence *aux pieds de*.

### 3. 3. THEMATIQUE ASSISTEE : RECHERCHE DE METHODE DANS UN CORPUS MULTI-AUTEURS

#### 3. 3. 1. SAAS, un ensemble de documents d'aide à l'interprétation

Les résultats des différentes expériences menées pour vérifier si le test probabiliste, utilisé dans un corpus homogène quant au genre, dans une période de l'histoire de la langue suffisamment homogène également, permet de repérer des zones de densité sémantique, ont été probants. Comme le signe "isolé" est un artefact, nous devons chercher dans son entour les traits sémantiques qui contribuent à donner à son signifié sa "couleur" particulière, dans une aire sémantique donnée. Le test met en évidence les lexies qui sont en lien avec le mot pôle, ou la séquence pivot, sur l'axe paradigmatique ou sur l'axe syntagmatique<sup>45</sup> : dans les deux cas, les relations sémantiques sont à interpréter, elle ne sont pas des "évidences" et nous avons vu que les normes sociales ont leur place dans l'explication des régularités observées en corpus.

Au fur et à mesure que ces expériences étaient menées, nous avons revu et amélioré la présentation des "sorties" des traitements informatiques et statistiques. Les documents du SAAS (cf. les échantillons donnés en annexes du chapitre 1) ont été conçus pour permettre l'interprétation avec un maximum d'informations directement accessibles, puisqu'on ne peut

<sup>44</sup> Nous avons mené le même type de travail en constituant le corpus d'étude avec le verbe *jeter* et la séquence *à la tête* dans le contexte droit : on a affaire à une structure différente, où le trait /agression/ est primordial. Il s'agit surtout de disputes domestiques où on s'agresse avec (sur l'isotopie /physique/) une *bouteille*, un *flacon*, un *verre*, un *encrier*, des *assiettes*, un *gros Quicherat*, un *pavé*, et sur l'isotopie /moral/, on s'envoie des *insultes*, les *morts* et les *vivants*, sa *race*, son *portefeuille* [ministériel], ses petites *misères*, des *vérités*, ses hypothèses catholiques, sa *famille*, sa *démission*, etc. (en italiques les mots sélectionnés par le test). Le trait /agression/ (au plan social) est validé dans les contextes de l'expression *se jeter à la tête de qqn* "chercher à se faire épouser" attesté comme un topos dans ce corpus roman, où l'acteur /féminin/ *jette son dévolu* (autre lexicalisation du topos) sur un acteur /masculin/, sans respect des "règles implicites" : la femme séduit, l'homme conquiert, mais la femme n'engage pas l'action de façon "agressive". Le topos est attesté également dans la version : 'une mère jette sa fille à la tête d'un bon parti'.

<sup>45</sup> Ces notions de la linguistique du signe prennent un autre relief dans le travail sur corpus : le corpus entier sert de "contexte" et les relations sémantiques peuvent être mieux interprétées grâce à la variété des lexicalisations et des rôles, cf. chapitres 4 et 5.

prévoir quelles types de recherches seront menées : sur les signifiants, les signifiés, les hapax, la phraséologie, l'aspect morphologique, syntaxique, etc.

1) Les listes de cooccurrents sélectionnés les présentent selon différents classements (ordre par score, par fréquence et sous-listes alphabétiques combinant des critères de scores et de fréquence) : on peut ainsi aller d'un signifiant à un autre, puisque les listes alphabétiques permettent de retrouver rapidement un mot de la famille morphologique ou un mot associé sur l'axe paradigmatique ou syntagmatique, sur les traits sémantiques duquel on fait une hypothèse.

2) Les documents contextuels renouvellent l'antique présentation en "concordances"<sup>46</sup> qui a largement fait la preuve de son utilité dans les sciences du texte, en distinguant les cooccurrents sélectionnés en majuscules, afin que la lecture non-linéaire du texte soit enrichie de la mise en évidence des associations non aléatoires, probablement en lien avec la tâche : les processus interprétatifs sont sollicités par ce soulignement, et la discrimination des traits pertinents facilitée par cette contextualisation pondérée<sup>47</sup>.

3) Les tableaux synoptiques regroupent les informations sur la fréquence dans les corpus de travail et de référence et les scores statistiques, pour faciliter l'interprétation d'une éventuelle densité sémantique, engendrée par la consistance des résultats des différentes colonnes ou au contraire par des constats de divergences.

A l'aide de ces documents, l'interprétation en termes de traits sémantiques, qui consiste, comme on l'a vu, en des démarches alternées de recherche onomasiologique (on part du "sens" pour rechercher dans quelle lexie on peut le "couler") et de recherche sémasiologique (quel est le signifié en contexte de ce signe ?) peut s'effectuer aisément et même relativement rapidement puisque les accès au contexte sont divers et que le nombre restreint de cooccurrents sélectionnés découpe les textes en passages denses.

C'est une conception sémiotique du texte et du corpus qui a présidé à l'élaboration de ces programmes : on part du texte pour aller vers ses éléments, en le considérant comme une structure de type "gestalt", où la classe détermine l'élément, et où le global détermine le local, ce qui justifie que toute "information" doit être contextualisée<sup>48</sup>.

---

<sup>46</sup> Une évolution importante des instruments de travail s'opère, à partir du XII<sup>e</sup>s., faisant adopter des *indices* avec classement des références par ordre alphabétique qui renouvellent les recueils d'*exempla* de l'Antiquité gréco-latine, fondés sur le classement par rapports réciproques : du XIII au début du XIV<sup>e</sup>s., les concordances des textes latin, grec et hébreu de la Bible connaissent un succès très rapide et se diffusent, souvent sous forme de "*species*", c'est-à-dire de cahiers séparés qu'on peut louer séparément. Les auteurs sont partagés entre un désir d'exhaustivité et le caractère pragmatique et commode de la consultation des manuscrits : la question des mots "grammaticaux", celle de la référence au verset et celle de la présence de notices (pour l'amplification) partagent les auteurs. La première concordance de la Bible traduite en français apparaît en 1535 et tout au long du XVI<sup>e</sup>s., la question de la conformité du texte avec les Saintes Ecritures se pose, pour des raisons religieuses et philologiques également : l'Eglise fixe le texte de la Bible au Concile de Trente. Sur l'évolution, la "modernité et l'universalité de cet outil", v. Sekhraoui M, 1995.

<sup>47</sup> Les dernières versions d'Hyperbase ont adopté ce procédé.

<sup>48</sup> Ce point est particulièrement important pour l'enrichissement des corpus par les balises de type TEI (en-têtes et renseignements de type "bibliographiques"), comme nous l'avons signalé au chapitre 2.

### 3. 3. 2. La méthode

#### *Recherche de régularités thématiques autour d'un mot pôle*

Sur le modèle des analyses conduites autour du mot pôle *pied* (en tant que lemme ou forme par forme), qui ont mené vers des unités sémantiques allant du syntagme à la période, nous avons testé une méthode de thématique assistée dans un corpus multi-auteurs : l'hypothèse est que le test probabiliste sélectionne, dans un ensemble de cooccurents attestés dans une fenêtre de sélection de 10 signes avant et après le mot pôle, ceux qui ont un rapport sémantique dense avec celui-ci, et que l'étude des contextes de ces cooccurents sortis par le test parce que leur probabilité d'être là n'est pas due au hasard, permet d'interpréter les traits sémantiques faisant "structure", si cette structure existe pour un ensemble d'auteurs donné.

Le chapitre 4 rend compte de l'expérience menée autour des "mots de la peur" : il n'était pas possible de rechercher s'il existe des régularités thématiques autour du mot *peur*, sans se demander si elles rejoignaient éventuellement celles qu'on pouvait trouver auprès de *crainte*, *terreur*, etc. Comme en sémantique on ne privilégie pas de classes de signes<sup>49</sup>, nous avons donc étudié non seulement les substantifs, mais aussi les adjectifs et verbes qui comportent "en langue" le trait /peur/, soit 26 mots<sup>50</sup> : cet ensemble de mots pôles a permis de constituer les sous-corpus de travail (chaque ensemble de fenêtres de 10 signes avant et après l'un des mots de la peur) et le corpus de travail global (l'addition de ces sous-corpus). Pour l'interprétation des résultats, nous avons suivi la double démarche de la sémantique<sup>51</sup>, en procédant en deux étapes :

- démarche sémasiologique : étude des cooccurents sélectionnés dans chaque sous-corpus (autour de *terreur*, de *faire peur*, d'*effrayé*, etc.) en ne gardant que les cooccurents attestés chez des auteurs différents, avec une fréquence égale ou supérieure à 4, et au moins dans deux sous-corpus
- dans un deuxième temps, nous avons repris l'étude de chaque sous-corpus pour vérifier si les traits sémantiques repérés lors de la première analyse étaient validés dans les contextes avec d'autres lexicalisations que celles relevées avec les "contraintes fortes" de la première étape (démarche onomasiologique).

Ce type de travail a été réalisé également autour des mots pôles *parure* et *parer*, à la recherche d'un thème 'Parure' dans le corpus roman : il est l'objet du chapitre 5, qui complète l'illustration de la méthode parce que les difficultés rencontrées étaient différentes (importance des formes homographes de *parer*, en particulier).

<sup>49</sup> L'approche logico-grammaticale se caractérise au contraire par l'importance accordée aux noms (et surtout aux substantifs), par exemple : "la description lexicale des noms dans un réseau comme WordNet est plus riche et plus structurée" Habert et alii, 1997, p. 115.

<sup>50</sup> Ce sont : *crainte*, *effroi*, *épouvante*, *frayeur*, *horreur*, *panique*, *peur*, *terreur*, *crainitif*, *peureux*, *épouvanté*, *horrifié*, *terrifié*, *effrayant*, *effroyable*, *épouvantable*, *horrible*, *terrible*, *craindre*, *redouter*, *avoir peur*, *effrayer*, *épouvanter*, *terrifier*, *terroriser*, et *faire peur*.

<sup>51</sup> Comme Swiggers, P. 1983, p. 437, nous considérons que l'onomasiologie et la sémasiologie sont "deux méthodes de recherche" et non pas deux disciplines, mais nous ne le suivons pas quand il écrit : "ces deux méthodes s'opposent par la nature de la corrélation établie : celle-ci est homogène dans le cas de la sémasiologie (qui met en jeu deux entités linguistiques), alors qu'elle est hétérogène dans le cas de l'onomasiologie, qui se base sur des entités extra-linguistiques afin d'analyser les données linguistiques." Ce point de vue récuse le statut sémiotique des traits sémantiques et du texte.

***Caractérisation d'un texte contrasté sur un corpus de genre***

Des études littéraires et stylistiques très fines et très stimulantes ont été menées par le passé par des chercheurs qui consacraient leur vie à l'approche de l'œuvre d'un auteur, ou d'un mouvement littéraire, d'un thème, etc. Ils travaillaient par fiches manuscrites en prenant des notes à la fois sur le contenu et sur la forme. Comment utiliser l'outil informatique pour renouveler ces recherches, et bénéficier de la rapidité d'accès aux formes lexicales pour pratiquer de nouveaux modes de lecture ? Peut-il nous permettre d'appréhender "ce qui fait sens" dans un texte, quels faits langagiers, thématiques, stylistiques le caractérisent par rapport aux productions de son époque, de son genre textuel ?

Nous avons tenté l'expérience de contraster un roman, *Le Père Goriot* de Balzac, sur son corpus générique à l'aide du test probabiliste en faisant l'hypothèse que celui-ci nous fournirait des pistes sur des faits caractéristiques de cette œuvre, en lien avec sa ou plutôt ses thématiques (articulées autour des personnages de Rastignac et de Goriot), mais aussi d'autres éléments du "tissu" de l'œuvre, un peu comme les techniques statigraphiques apportent des éléments pour l'analyse des œuvres picturales (datation, composition, retouches, emploi de tel pigment, etc.) . Le plan de ce qui est quantifiable, les signes, tels groupes de signes particuliers, les scores de pondération, etc. ne sont que des indices de ce qui est à déchiffrer au plan qualitatif, par la mise en parallèle des éléments pertinents, et dans le questionnement auquel ils confrontent le chercheur.

On a appliqué le test probabiliste sur *Le Père Goriot*, un roman de taille moyenne (près de 114000 signes), en le contrastant sur le corpus constitué de 350 romans environ de 1830 à 1970, c'est-à-dire que l'ensemble de ses signes a constitué le corpus de travail, contrasté sur le corpus de référence de 40 millions d'occurrences environ. Bien que Frantext n'ait pas de jalon de chapitre, nous avons tronqué le texte en quatre sous-parties, les chapitres, pour les contraster à leur tour chacun sur le corpus Roman, comme un tout thématique, afin de mettre en parallèle les résultats obtenus dans les différentes zones de localité. L'hypothèse de travail est que les cooccurrents sélectionnés par le test sont surdéterminés par différentes "dimensions" du texte et doivent pouvoir s'interpréter à l'aide des quatre composantes textuelles, *la thématique, la dialectique, la dialogique et la tactique*<sup>52</sup> qui représentent, en sémantique textuelle, les modes d'approche des différentes facettes du texte considéré comme un objet sémiotique complexe, et non comme un "tas de mots": cette étude constitue le chapitre 6.

**3. 4. RESULTATS, PROBLEMES, PROPOSITIONS**

1) Comme différents travaux dans le domaine du TAL l'ont montré, la recherche sur les figements peut utiliser le test probabiliste (en association avec d'autres méthodes comme le calcul de similarité) : mais il faut souligner que les résultats seront meilleurs si l'on traite des corpus homogènes en discours, domaines et genres, puisque, en faisant varier seulement le trait de nombre dans le corpus Roman, les résultats permettent de distinguer des "unités

---

<sup>52</sup> Sur ces notions, v. glossaire, et chapitres 4, 5 et 6.

sémantiques" : *le pied chaussé vs les pieds chaussés, les pieds nus, nu pieds, va nu pieds, mettre le pied à l'étrier, mettre au pied du mur, X pieds de profondeur, etc.*

- Il est probable que des seuils de scores qui ont trait au figement puissent être déterminés, si on opère sur des corpus homogènes, et en pratiquant sur les contextes droit et gauche séparément, en décalant d'une place à chaque fois : la collaboration avec des statisticiens pour la détermination de seuils permettra d'automatiser certaines opérations
- la comparaison de résultats obtenus dans des corpus différents doit contribuer à fixer la taille des "patrons d'association" entre des cooccurrents, en précisant les types d'expansions possibles

2) Dans une banque textuelle, le moteur de recherche doit permettre à l'utilisateur de filtrer le voisinage avec un test probabiliste pour pouvoir caractériser les contextes : les résultats bruts sur les voisinages (fréquences absolues) ou les fréquences relatives (rapport au nombre d'occurrences du texte, comme le propose Frantext) sont à proscrire car ininterprétables.

- Variation du pivot, variation de l'empan : on a vu que certains contextes sont le siège d'une densité sémantique due à la lexicalisation d'une fonction narrative. L'utilisateur doit pouvoir sélectionner automatiquement des contextes dont il fixe l'empan, autour de mots pôles ou séquences pôles variées (ex. : *implorer, prier, se jeter aux pieds de, s'agenouiller aux pieds de*) afin de déterminer les régularités d'une forme sémantique aux lexicalisations variées. Contrairement à ce qui est pratiqué dans Frantext, toute sélection doit pouvoir être affinée autant de fois que nécessaire par l'utilisateur, pour limiter le bruit qui provient d'aires sémantiques différentes. La possibilité de se constituer des sous-ensembles de corpus (sur la base de traits discriminants comme /pluriel/ + /humain/, par exemple), doit être ménagée afin de pouvoir pratiquer des traitements (syntaxiques, statistiques, etc.) à des niveaux d'homogénéité maîtrisée.

La recherche thématique intéresse également les textes techniques, dans lesquels le degré de complexité est moindre : dans un texte à thématique fixe comme l'ouvrage sur le football, le nombre d'acteurs ('ballon', 'surface de réparation', etc.) et de rôles ('donner un coup de pied de coin', 'sortir de la surface de réparation') sont à la fois plus réduits, et lexicalisés de manière plus stéréotypée, donc plus faciles à mettre en évidence.