

Statistiques et motifs séquentiels : le cas des suites [Nom préposition Nom], *mot à mot, pétale par pétale*

La contribution se propose d'évaluer l'outil statistique pour identifier et analyser les suites *Nom préposition Nom*, avec pour critère restrictif que les deux substantifs sont identiques¹.

L'enjeu essentiel de cette recherche est double : il est, d'une part, d'observer le rôle structurel des prépositions dans ces unités, à la fois sur le plan lexical et sémantique ; il est d'autre part d'évaluer le rendement de ces structures comme outils de classifications et, en particulier, comme marqueurs génériques. La base qui sert de corpus exploratoire à cette analyse permet en effet ce type de recherches, car elle est constituée d'un extrait de *Frantext* catégorisé qui comprend quatre groupes génériques ainsi répertoriés : deux ensembles narratifs d'une part, les mémoires (109 textes) et les romans (740 textes), la poésie (191 textes) et le théâtre (210 textes) d'autre part, soit près de 85 millions de mots. Les genres principaux sont ainsi représentés.

Le premier travail consiste à repérer les suites recherchées. Il s'avère que toutes les prépositions ne sont pas aptes à entrer dans ce type de constructions. Une réflexion pourra s'engager sur le statut des prépositions éligibles et sur leur portée sémantique. Le rendement des structures est évalué en observant ensuite les variations de substantifs de part et d'autre des prépositions : la préposition peut varier tandis que les noms restent identiques ; l'expression conserve globalement le même sens (*mot à mot, mot pour mot ; bouche à bouche, bouche contre bouche*) ; la préposition peut être la même tandis que les substantifs peuvent varier (*nuît après nuît, soir après soir* : le sens exprime ici l'idée de succession ; *rendre haine contre haine, corps à corps, nez contre nez* : le sens se module en l'idée d'échange, d'opposition, de contact).

L'analyse a aussi à composer avec les structures lexicalisées ; on s'intéressera aux critères qui autorisent une interprétation sur le degré de figement des structures, situées entre l'expression lexicalisée enregistrée en langue et la création lexicale. Entre phraséologie et style d'auteur, conçu comme appropriation des unités linguistiques, entre langue et discours, ces unités seront évaluées en fonction de leur productivité dans le corpus d'étude. L'observation des hapax – ou unités de fréquence 1 – sera

¹ Nos remerciements vont à G. Purnelle (Université de Liège ; LASLA) pour son aide quant à l'élaboration des résultats statistiques présentés dans ce travail, qui est un extrait d'une recherche plus ample en cours (en particulier sur le rendement de la théorie du motif, JADT 2012).

engagée comme possible critère distinctif de l'innovation lexicale. L'étude amorcera par ce biais une délimitation et une possible définition des unités lexicalisées.

Comme le corpus est constitué de quatre ensembles génériques, il sera loisible de voir enfin comment ces unités phrastiques se distribuent et si une répartition générique peut être proposée.

1. L'unité Nom préposition *Nom*

1.1. Observation du corpus

On commence par l'observation des données expérimentales fournies par le corpus d'étude afin de dégager quelles sont les prépositions éligibles, autrement dit susceptibles de relier immédiatement deux substantifs identiques, chacun étant dépourvu de déterminant.

Dans le corpus d'étude, des occurrences de la suite ternaire recherchée ne se rencontrent qu'avec les prépositions : *à, après, contre, par, pour, sur* ainsi qu'avec les structures plus complexes *de ... à* et *de ... en*².

<u>Structure =</u> <u>N préposition N</u>	<u>Total</u>	<u>Exemple d'ex-</u> <u>pression</u>	<u>Structure =</u> <u>N préposition N</u>	<u>Total</u>	<u>Exemple</u> <u>d'expression</u>
à	2997	mot à mot	d(e) ... en	4222	de page en page
après	247	Heure après heure	par	969	goutte par goutte
contre	253	cœur contre cœur	pour	398	coup pour coup
d(e) ... à	372	d'ami à ami	sur	663	pierre sur pierre
			<u>Total</u>	10121	

Figure 1- Les structures éligibles

Toute réalisation d'une structure NpN à partir d'un substantif particulier est appelée *expression ou unité* : *heure après heure, d'ami à ami*. Sur le plan terminologique, on pourrait réutiliser le terme de « motif »³ séquentiel ou de « cadre collocationnel »⁴ dans le sens d'un patron syntaxique fixe.

² La préposition *selon* ne fournit qu'une seule occurrence : le « pli selon pli » de Mallarmé dans le sonnet *Remémoration d'amis belges* : « Que se devêt pli selon pli la pierre veuve ».

³ Le motif est défini de manière un peu abstraite dans certains travaux comme « un sous-ensemble ordonné [d'un ensemble] (E) formé par l'association récurrente de n éléments de l'ensemble (E) muni de sa structure linéaire ». D. Longrée, X. Luong, S. Mellet, 2008. « Les motifs : un outil pour la caractérisation topologique des textes », *9e JADT*, 733-744, <lexicometrica.univ-paris3.fr/jadt/jadt2008/pdf/longree-luong-mellet.pdf>.

⁴ A. Renouf et J. Sinclair, 1999. *English corpus Linguistics : Studies in Honour of Jan Svartvik*, Longman, Chapter Collocational Frameworks in English, 128-143.

Sur le plan formel, les structures NpN peuvent en effet entrer dans la catégorie du motif syntaxique, qui est une des catégories identifiées du motif, à côté de celles constituées par les suites phonologiques ou encore métriques. La structure étudiée se caractérise par le fait que les substantifs de part et d'autre de la préposition sont en théorie réversibles puisqu'ils sont identiques dans leur forme; cependant, l'ordre d'apparition des occurrences dans la phrase a forcément une incidence sur le sémantisme du lexème induisant une vectorisation du système. D'une occurrence à l'autre, le sens se module; la répétition induit une variation de sens. La disposition des unités dans l'espace du texte oriente la structure en l'inscrivant dans la temporalité de la lectureursive.

Le corpus d'étude fournit 10 122 occurrences de la structure ternaire choisie. Indépendamment de la partition générique, les expressions les plus fréquentes sont :

côte à côte	1107	nez à nez	165
goutte à goutte	359	jour par jour	150
coup sur coup	315	de main en main	133
de minute en minute	223	de porte en porte	124
corps à corps	192	tête à tête	108
de place en place	186		

Figure 2- Les expressions les plus fréquentes

1.2. Les unités et le paramètre sémantique

La plupart des prépositions sont polyvalentes et polysémiques. C'est le sens des substantifs de part et d'autre qui détermine le sens de la relation sémantique réalisée entre les noms et, par conséquent, le sens global de l'expression.

L'examen des expressions NpN du corpus permet d'établir une liste de huit expressions que le paramètre sémantique⁵ distingue :

Succession : décomposition du procès en éléments

Nous y secouons nos cigarettes: nous sommes lavés, nous avons le temps; le temps chaud et stagnant, le temps qui passe *minute à minute*, sans bruit, sans fièvre, en chuchotant. (A. Sarrazin, *L'Astragale*)

⁵ Sur 10121 occurrences, 11 n'entrent pas dans ces 8 catégories et sont donc soustraites des effectifs.

Accumulation : addition d'objets

La mère, toujours malade, comptait *sou à sou*. Elle tremblait devant lui, autant que nous. (R. Martin du Gard, *Les Thibault, l'Été 1914*).

Mon grand vizir est à la guerre où il remporte d'ailleurs *victoire après victoire* sur les Russes (M. De Grèce, *La Nuit du Sérail*)

Distance :

J'ai tendu des cordes *de clocher à clocher* ; des guirlandes *de fenêtre à fenêtre* ; des chaînes d'or *d'étoile à étoile*, et je danse. (A. Rimbaud, *Illuminations*)

Déplacement - mouvement

Je marche *de surprise en surprise, d'ahurissement en ahurissement*. (P. Benoit, *L'Atlantide*)

Relation :

D'où Nana tombait-elle ? Et des histoires couraient, des plaisanteries chuchotées *d'oreille à oreille*. (É. Zola, *Nana*)

Mille cris se croisaient *de baraque à baraque*. (P. Adam, *L'Enfant d'Austerlitz*)

Contact :

J'ai l'horreur de la fraternité qui établit des contacts *de peau à peau* (J. Genet, *Miracle de la rose*)

Disait Hérode à Sigognac, qui cheminait *botte à botte* avec lui (Th. Gautier, *Le Capitaine Fracasse*)

Opposition :

On se battit *corps à corps* sous les arbres (V. Hugo, *Les Misérables*)

J'ai lutté pendant quatre années jour et nuit, *corps à corps, astuce contre astuce*, avec ce génie infernal (Ponson du Terrail, *Rocambole*)

Échange :

Violence contre violence. Chacun donne ici sans mesure sa force. (Aragon, *Les beaux Quartiers*)

	à	après	Contre	de ... à	de ... en	par	pour	sur	Total
<u>Accumulation</u>	5	28	2	0	0	15	0	253	303
<u>Contact</u>	1790	0	190	2	0	0	0	51	2033
<u>Déplacement</u>	0	0	0	25	4222	0	0	0	4247

	à	après	Contre	de ... à	de ... en	par	pour	sur	Total
<u>Distance</u>	1	0	0	6	0	0	74	0	81
<u>Échange</u>	0	0	11	0	0	6	324	0	341
<u>Opposition</u>	171	0	50	0	0	0	0	0	221
<u>Relation</u>	30	0	0	329	0	0	0	0	359
<u>Succession</u>	994	219	0	5	0	948	0	359	2525
<u>Total</u>	2991	247	253	367	4222	969	398	663	10110

Figure 3 - Fréquences des structures et des sémantismes

En conservant soit la préposition, soit le substantif comme invariant d'une expression, et en faisant varier l'autre élément, on obtient les deux cas de figure suivants au service d'un processus interactif de constitution du sens :

Une structure qui combine une même préposition avec des noms différents peut prendre des sens différents : la préposition n'a donc pas un sens stable et invariant, donné en langue, mais un sens toujours modulable selon les combinaisons où elle entre, son sens s'ajustant au substantif répété de part et d'autre. La préposition « contre » par exemple peut exprimer quatre types de relations sémantiques : l'*Opposition* : « argument contre argument » ; le *Contact* : « bord contre bord » ; l'*Échange* : « donnant donnant, enfant contre enfant » ; l'*accumulation* « bloc contre bloc ».

La polysémie de la structure N *contre* N est ainsi démontrée.

De manière symétrique, des substantifs identiques reliés par des prépositions différentes peuvent constituer une structure de même sens : par exemple, entre les expressions *lutter âme à âme* et *lutter âme contre âme* ou encore entre *bouche à bouche* et *bouche contre bouche* peut être postulée une équivalence de sens. Un même sémantisme est réalisé par plusieurs structures.

Les statistiques établissant des pourcentages valideront-elles ces hypothèses de travail ?

Structure	à	après	contre	d(e) ... à	d(e) ... en	par	pour	sur
<u>Accumulation</u>	0,17%	11,34%	0,79%			1,55%		38,16%
<u>Contact</u>	59,85%		75,10%	0,54%				7,69%
<u>Déplacement</u>				6,81%	100,00%			

<u>Structure</u>	<u>à</u>	<u>après</u>	<u>contre</u>	<u>d(e) ... à</u>	<u>d(e) ... en</u>	<u>par</u>	<u>pour</u>	<u>sur</u>
<u>Distance</u>	0,03%			1,63%			18,59%	
<u>Échange</u>			4,35%			0,62%	81,41%	
<u>Opposition</u>	5,72%		19,76%					
<u>Relation</u>	1,00%			89,65%				
<u>Succession</u>	33,23%	88,66%		1,36%		97,83%		54,15%

Figure 4 – Quels sémantismes pour les structures identifiées ?

Ce tableau présente les différents sémantismes associés à une même structure.

- Les structures les plus polysémiques sont *à* (6 sémantismes lui sont associés : l'accumulation, le contact, la distance, l'opposition, la relation, la succession), *de ... à* (5 sémantismes : le contact, le déplacement, la distance, la relation, la succession) et *contre* (4 sémantismes : l'accumulation, le contact, l'échange, l'opposition).
- En revanche, *De ... en* est monosémique et indique le déplacement.
- Chaque structure a un sémantisme nettement dominant, intrinsèque, sauf *à* qui cumule le *Contact* et la *Succession*, *contre*, associé au *Contact* et à l'*Opposition* et *sur*, associé à l'*Accumulation* et la *Succession*.

<u>Structure</u>	<u>à</u>	<u>après</u>	<u>contre</u>	<u>d(e) ... à</u>	<u>d(e) ... en</u>	<u>par</u>	<u>pour</u>	<u>sur</u>
<u>Accumulation</u>	1,65%	9,24%	0,66%			4,95%		83,50%
<u>Contact</u>	88,05%		9,35%	0,10%				2,51%
<u>Déplacement</u>				0,59%	99,41%			
<u>Distance</u>	1,23%			7,41%			91,36%	
<u>Échange</u>			3,23%			1,76%	95,01%	
<u>Opposition</u>	77,38%		22,62%					
<u>Relation</u>	8,36%			91,64%				
<u>Succession</u>	39,37%	8,67%		0,20%		37,54%		14,22%

Figure 5 – Par quelles structures sont réalisés les sémantismes ?

Ce tableau présente les différentes réalisations concrètes autrement dit les structures associées à un même sémantisme.

- Les sémantismes qui se réalisent au moyen des structures les plus variées sont l'*Accumulation* et la *Succession* qui sont toutefois sémantiquement proches, car ils sont liés en particulier par l'idée de succession chronologique qu'ils impliquent tous deux: il y a polyvalence des structures, à des degrés divers.
- Chaque sémantisme a une structure privilégiée, sauf la *Succession*, qui présente des affinités, à parts à peu près égales, avec les structures articulées autour de *à* ou de *par*; la primauté d'une structure est moins nette pour l'*Opposition*. Par exemple, le sémantisme du contact privilégie les structures avec *à*; ceux de la distance et de l'échange, les structures avec *pour*.

La synonymie des structures est vérifiée au travers des résultats statistiques, mais elle est plutôt limitée, si on prend en considération les sémantismes dominants associés à une structure, et ne porte de manière significative que sur deux paires de structures, comme le montre l'analyse arborée, figure 6.

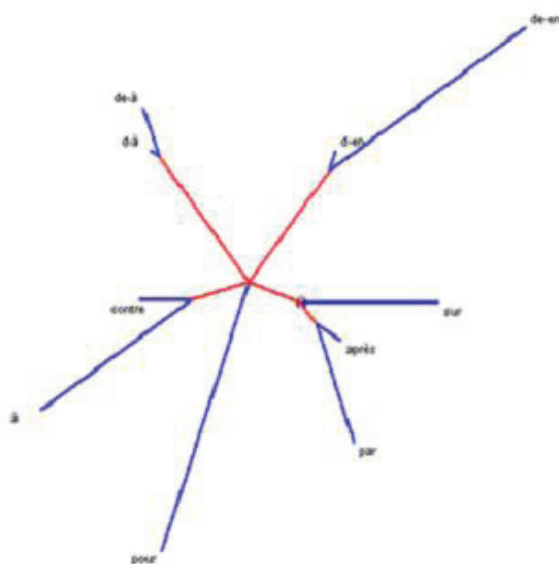


Figure 6 - Les structures évaluées selon les sémantismes associés

L'analyse arborée manifeste le rapprochement sur une même branche des prépositions *par* et *après* (succession), *à* et *contre* (contact). On peut en déduire que les expressions *N par N* et *N après N* d'une part et *N à N* et *N contre N* d'autre part se rejoignent par un sémantisme commun.

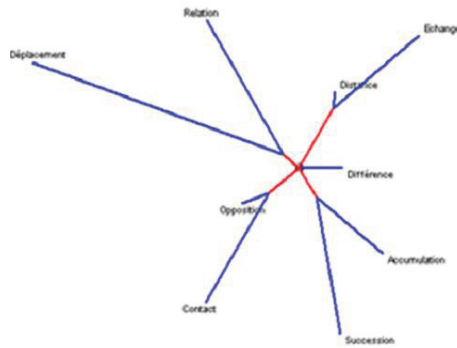


Figure 7 - Les sémantismes évalués selon les structures associées

On peut faire deux observations: la polysémie, partant la polyvalence partielle des structures d'un côté, la tendance à la spécialisation de l'autre. L'étude des structures NpN peut donc être envisagée soit à partir des 8 structures, soit à partir des 8 sémantismes.

La fréquence et la productivité des structures sont deux paramètres qui règlent l'appréciation de leur rendement.

2. Fréquence et productivité des unités

2.1. Les résultats contrastifs et relatifs

Pour évaluer la productivité des structures NpN, on peut calculer la proportion des occurrences des structures NpN sur les effectifs totaux de chaque préposition dans le corpus (*Mémoires, Roman, Poésie, Théâtre* de 1830 à nos jours). De ces proportions, trois observations peuvent être tirées: les prépositions les plus productives en NpN - dont on retient les effectifs - sont *contre* et *en*; la préposition *à* est si fréquente et présente une telle polyvalence qu'elle produit moins de structures NpN; enfin la productivité d'une préposition n'est pas identique selon les genres. On reviendra sur ce point en dernière partie de l'étude.

On peut aussi établir une comparaison entre la fréquence des prépositions et le nombre d'expressions différentes produites par les structures NpN. On calcule alors d'abord la proportion des effectifs de chaque préposition sur le total des effectifs des 7 prépositions, ensuite la proportion des différentes expressions générées par chaque préposition sur le nombre total des expressions.

Prépositions			Structures		
à	1341642	40,46%	à / de ... à	347	15,37%
après	75393	2,27%	après	122	5,41%
contre	44305	1,34%	contre	114	5,05%
en	772517	23,29%	de ... en	1083	47,98%
par	295617	8,91%	par	220	9,75%
pour	437191	13,18%	pour	137	6,07%
sur	349623	10,54%	sur	234	10,37%
Total	3316288			2257	

Figure 8 - Les proportions – tableau préparatoire au graphique de la figure 9

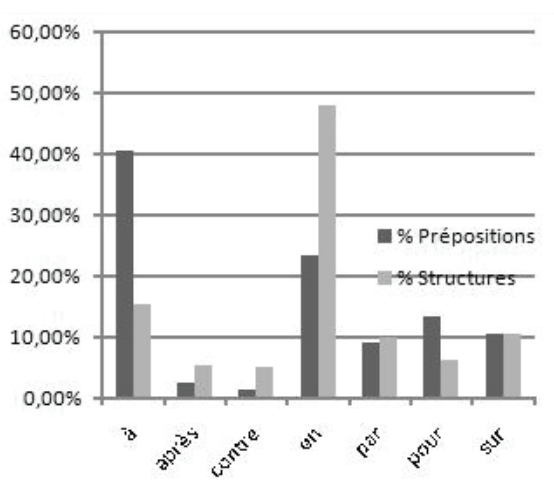


Figure 9 - La productivité des structures

À et *en* sont les deux prépositions les plus fréquentes dans le corpus, mais leur productivité est inversement proportionnelle : *à* génère beaucoup moins d'expressions différentes que *en* dans la structure *de ... en*. Comme pour *à*, les autres écarts sont le signe d'une spécialisation partielle des structures qui confine à la lexicalisation ou au figement : *œil pour œil*, *corps à corps*, ou, à l'inverse, d'une productivité : *après*, *contre*.

La structure choisie se situe entre langue et discours. La part doit être faite en effet entre les lexies ou unités phraséologiques introduites dans le discours et les créations de séquences ternaires sur un patron linguistique préconstruit. L'analyse croisée des structures qu'on pourrait dire figées puisqu'elles sont constituées en langue, comme syntagmes adverbiaux. La lexicalisation de la structure se vérifie notamment lorsqu'elle repose sur un transfert normalisé du sens propre au sens figuré :

- (1) Ce n'est pas que Combeferre ne fût capable de combattre, il ne refusait pas de prendre *corps à corps* l'obstacle et de l'attaquer de vive force et par explosion. (V. Hugo, *Les Misérables*)

Dans d'autres exemples, le jeu sur l'emploi du sens concret ou figuré de la structure procède au défigement discursif de la structure, susceptible de créer un effet.

- (2) Tout le monde descend l'escalier ; toi, tu déchireras tes draps de lit, tu en feras, *brin à brin*, une corde, puis tu passeras par ta fenêtre, et tu te suspendras à ce fil sur un abîme. (V. Hugo, *Les Misérables*)
- (3) Aux quatre vents, quelle tempête s'amasse sur le lac du genre humain ! N'est-ce pas la création sans foi qui se détache *brin à brin* des mains du créateur, et tombe dans l'abîme, comme le chapelet d'un prêtre d'Arménie tombe à ses pieds. (É. Quinet, *Ahasverus*)

La contextualisation de la structure permet d'appréhender celle-ci : elle est susceptible d'être envisagée comme lexie quand l'emploi métaphorique est neutralisé par l'usage comme dans l'exemple (2). Sa place entre langue et discours est cependant préservée dans tous les cas, car le contexte est toujours à même de remotiver une figure lexicalisée.

2.2. La création lexicale

La création de nouvelles expressions est un fait de discours susceptible d'être interprété comme fait de style. La création peut se réaliser par analogie à partir de syntagmes figés en langue. Par exemple, si *d'égal à égal* est identifié comme locution adverbiale par le *TLF*, *groin à groin* est attesté en discours, mais n'apparaît pas dans l'article *groin* du dictionnaire. Le sens de l'expression se comprend par référence à d'autres expressions structurellement équivalentes et usuelles avec lesquelles elle forme un paradigme et par sa contextualisation ; mais l'expression n'est pas grammaticalisée. La fréquence des hapax dans un corpus donné peut être proposée comme indice de la productivité des structures.

Il est difficile d'établir facilement et avec certitude quelles suites existent en langue et lesquelles sont des créations ; un critère statistique robuste peut être choisi comme préalable :

- les structures les plus fréquentes atteignent un degré de figement élevé jusqu'à se grammaticaliser pour certaines et à fonctionner le plus souvent comme locutions adverbiales ;
- inversement, moins une expression est attestée, plus elle échappe à la lexicalisation.

De ce postulat, on déduira que les hapax peuvent fonctionner comme des indices de productivité; pour une structure, compter le nombre d'expressions de fréquence 1 dans le corpus établit une nouvelle mesure entre le nombre d'hapax et le total des expressions de la structure. Le décompte se porte sur le « vocabulaire » pour déterminer la proportion des hapax sur le nombre d'expressions différentes, autrement dit pour calculer combien de structures sont hapax et combien ne le sont pas.

On pourrait supposer que la proportion des expressions qui sont hapax est directement liée à la fréquence des structures : plus une structure est fréquente, plus le nombre de ses occurrences qui sont des hapax est important. C'est un fait avéré, mais avec les exceptions significatives qui indiquent une plus grande capacité à produire des hapax pour une expression et le phénomène inverse pour une autre. Il s'avère que la structure avec *sur* produit plus d'hapax que ne le laisserait prévoir sa fréquence relative, confrontée aux autres structures. La structure avec *à* manifeste un déficit en hapax ce qui signifie qu'elle entre davantage dans des structures figées.

On peut coupler cette observation des hapax avec une distribution par genres pour avancer sur l'idée proposée au départ que la structure NpN est un possible marqueur générique.

Les paramètres observés jusqu'à présent, sémantisme et productivité, peuvent être utilisés pour une approche de la distribution des structures NpN selon les genres.

3. Les structures et les genres

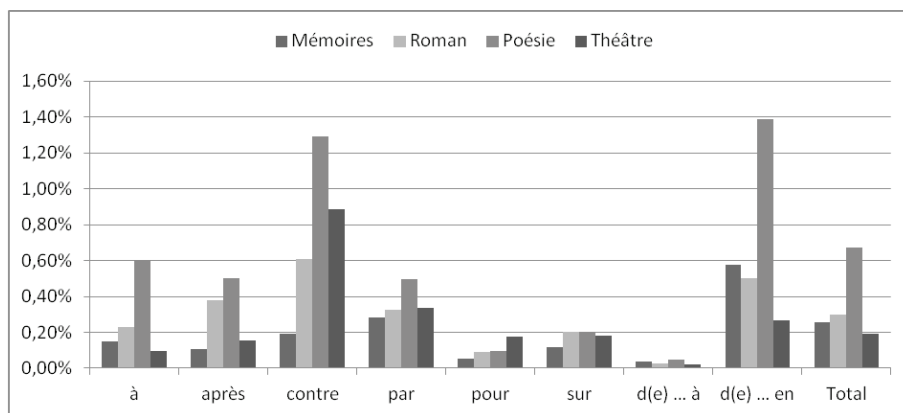


Figure 10 – Distribution des structures selon les genres

La figure 10 montre que la poésie est le genre qui mobilise la plus grande partie des effectifs de ses prépositions pour des structures NpN : cette affirmation est vraie pour *à*, *après*, *contre*, *par* et *en*.

Ceci est donc un premier indice d'une « fonction » générique de la structure NpN comme possible motif : il est plus spécifique à la poésie.

L'analyse factorielle permet la visualisation à la fois des structures et des genres : on voit ainsi sur le graphique 11 que certains genres affichent des affinités avec certaines structures en particulier.

Autour du roman gravitent les structures *après*, *sur* et *contre*, tandis que la poésie attire *de...en*, les mémoires *de...à* et le théâtre *pour* et *par*.

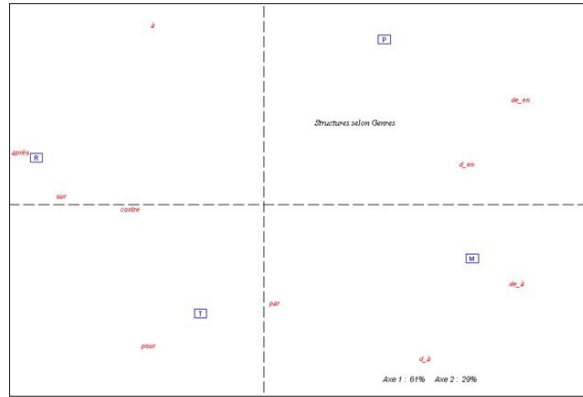


Figure 11 – Les structures selon les genres

L'observation peut être précisée en détaillant les données et en distinguant hapax et non-hapax (graphique 9) :



Figure 12 – Hapax et non-hapax selon les genres

On voit que la plupart des structures, dans les occurrences d'expressions hapax et non-hapax, ont le même comportement par rapport aux genres, à l'exception de la poésie, qui attire les occurrences hapax de plusieurs structures : *contre*, *sur*, *par*, *à*, et *de ... en*.

Ceci constitue un deuxième indice de l'affinité d'un genre parmi les autres pour le motif NpN : la poésie est le genre le plus propice à la création d'expressions nouvelles à partir des structures *contre*, *sur*, *par*, *à et de...en*.

L'affinité des genres et des sémantismes est représentée dans le graphique 13.

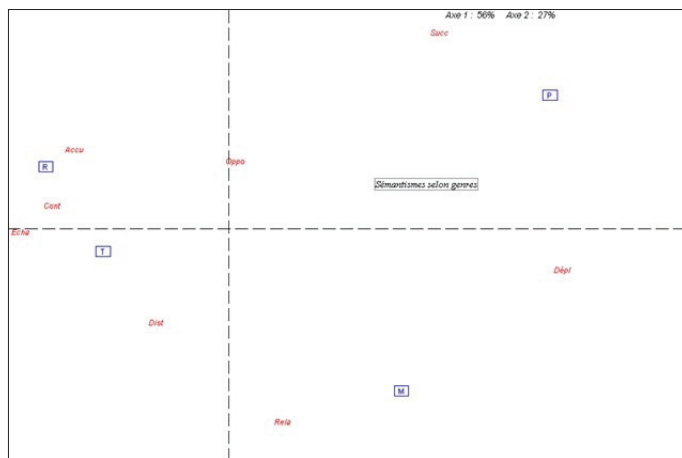


Figure 13 – Les sémantismes selon les genres

Cette AFC permet de visualiser les sémantismes privilégiés par chaque préposition. On peut supposer que le roman est caractérisé par le sémantisme du *Contact* en raison de son affinité avec la structure *contre*. La poésie se rapproche de la *Succession* – sémantisme réalisé par le plus grand nombre de structures sans qu'aucune n'affirme sa prédominance. C'est bien un sémantisme propre à la poésie.

4. Conclusion

Cette analyse sur les structures NpN a permis d'amorcer une réflexion sur le statut des prépositions et leur rôle structural dans l'unité lexicale et sémantique étudiée.

Un processus interactif de constitution du sens de l'unité a pu être démontré. Le sens de la préposition est modulé en fonction des noms environnants ; les structures sont polysémiques et, symétriquement, un même sémantisme peut être exprimé par plusieurs structures.

Toutefois, la plupart des structures ont un sémantisme dominant, affirmant une spécialisation d'emploi. Chaque sémantisme a une structure privilégiée.

La productivité des structures a été évaluée à l'aune des hapax qu'elles produisent. *À* accuse un déficit en hapax, entrant dans la constitution d'expressions figées tandis que *sur* entre dans le plus grand nombre d'expressions nouvelles. *Sur* est plus productif quant à l'innovation lexicale.

La suite NpN peut être proposée comme un possible marqueur générique puisqu'a été observée une distribution nette entre les genres. La poésie, en particulier, se distingue par les traits suivants: c'est le genre qui mobilise le plus grand nombre de ses prépositions dans des structures NpN; c'est le genre le plus propice à la création d'expressions nouvelles à partir des structures *contre, sur, par, à, de...en* et se trouve associée au sémantisme de la succession. Pour poursuivre l'analyse de ces suites NpN, il serait opportun d'établir un autre corpus non littéraire afin d'observer le comportement de la suite NpN dans l'un et l'autre ensemble et de voir si la suite peut fonctionner comme indice de littérarité⁶.

Université. Nice Sophia Antipolis
CNRS, BCL, UMR 7320

Véronique MAGRI-MOURGUES

Bibliographie

- D. Longrée, X. Luong, S. Mellet, 2008. « Les motifs : un outil pour la caractérisation topologique des textes », 9^e JADT, p. 733-744. <<http://lexicometrica.univ-paris3.fr/jadt/jadt2008/pdf/longree-luong-mellet.pdf>>.
- A. Renouf et J. Sinclair, 1999. *English corpus Linguistics: Studies in Honour of Jan Svartvik*, Longman, Chapter Collocational Framaeworks in English, 128-143.

⁶ Idée intéressante suggérée par David Trotter, lors de la présentation orale de ce travail.