

# *Mot à mot, brin par brin :* **les suites [Nom préposition Nom] comme motifs**

Véronique Magri<sup>1</sup>, Gérald Purnelle<sup>2</sup>

<sup>1</sup> Université de Nice Sophia-Antipolis, BCL – Veronique.Magri@unice.fr

<sup>2</sup> Université de Liège – Gerald.Purnelle@ulg.ac.be

## **Abstract**

The purpose of this paper is to evaluate the relevance of the notion of “pattern” (“motif”), defined as “a recurring association of  $n$  items in a text with a given linear structure”. It will be conducted by singling out and analysing the sequence (pattern) “noun preposition noun”, both nouns of this pattern having to be identical: *mot à mot, jour après jour, erreur sur erreur*. The relevant prepositions for the “motifs” are *à, après, contre, par, pour, sur* as well as the more complex structures *de ... à* and *de ... en*.

The analysis will be carried out on a corpus composed by the addition of extracts from the Frantext textual database, for the following generic subsets: novel, memoirs, poetry and theater.

Does the sequence  $N + \text{prép} + N$  fit the definitional parameters of the “motif”, namely frequency, recurrence, memorization? The study will have to factor in original and variable expressions and the lexicalized units which are their models. It will also intend to show how these sequences take part in discursive organization, especially in generic differentiation, and what their linked semantic traits are.

## **Résumé**

On se propose d'évaluer le rendement de la théorie du motif, défini comme « l'association récurrente de  $n$  éléments du texte muni de sa structure linéaire » au travers de l'analyse de la séquence « Nom préposition Nom », avec pour particularité restrictive que les deux noms sont identiques : *mot à mot, jour après jour, erreur sur erreur*. Les prépositions rentables pour le motif sont : *à, après, contre, par, pour, sur* ainsi que les structures plus complexes *de ... à* et *de ... en*. Le corpus exploratoire est constitué d'extraits de la base textuelle Frantext catégorisée, correspondant aux ensembles génériques du roman, des mémoires, de la poésie et du théâtre.

La séquence  $N + \text{prép} + N$  vérifie-t-elle les paramètres définitionnels du motif que sont la fréquence, la récurrence, la mémorisation ? L'étude prend en compte les unités lexicalisées, ainsi que les expressions variables et inédites auxquelles elles servent de modèle. On tente de voir quels sémantismes sont attachés à ces structures productives et quel peut être le rôle qu'elles jouent dans l'agencement discursif, en mettant en évidence d'éventuelles spécificités génériques.

**Mots-clés :** motif, préposition, genre.

## **1. Introduction**

L'objectif de cette communication est de tester la validité du nouvel outil proposé pour la caractérisation topologique des textes, lors des JADT 2008 : le « motif », en observant les suites ternaires « N préposition N » (NpN), lorsque N est identique et sans article. Il s'agit de voir

si ces séquences peuvent postuler au rang de motif, défini comme « élément structurant de la texture discursive<sup>1</sup> » ; ce faisant, la réflexion théorique sur le motif est prolongée pour en tester la portée heuristique en vue de la caractérisation des textes dans un corpus contrastif.

Un corpus expérimental est constitué de toutes les occurrences qui répondent à ce schéma à partir d'ensembles génériques extraits de Frantext catégorisé ; quatre sous-corpus sont ainsi sollicités : les mémoires (109 textes), les romans (740 textes), la poésie (191 textes) et le théâtre (210 textes), tels qu'identifiés dans la base, soit près de 85 millions de mots.

On commencera par observer les données expérimentales fournies par le corpus d'étude afin de délimiter au préalable quelles sont les prépositions rentables pour le motif et comment ces prépositions fonctionnent en interaction avec un substantif pour composer une structure.

Ces données seront ensuite confrontées aux critères du motif pour en évaluer la pertinence dans ce cas d'étude particulier. La réalisation de la suite ternaire est donc d'abord envisagée dans le corpus d'étude indépendamment de la partition générique, aux deux niveaux de l'analyse du motif, syntaxique et sémantique. Enfin, la fonction textuelle du possible motif sera testée en observant les distributions selon les ensembles textuels qui reflètent la partition générique.

## 2. La suite ternaire NpN et le motif

### 2.1. Préalable méthodologique

Les prépositions rentables pour le motif, autrement dit, qui vérifient la possibilité du retour du même substantif de part et d'autre de la préposition, sont : *à, après, contre, par, pour, sur* ainsi que les structures plus complexes *de ... à* et *de ... en*<sup>2</sup>. Effectifs :

Structure = N préposition N	Total	Exemple d'expression	Structure = N préposition N	Total	Exemple d'expression
à	2997	<i>brin à brin</i>	d(e) ... en	4222	<i>de minute en minute</i>
après	247	<i>page après page</i>	par	969	<i>jour par jour</i>
contre	253	<i>épaule contre épaule</i>	pour	398	<i>mot pour mot</i>
d(e) ... à	372	<i>d'homme à homme</i>	sur	663	<i>coup sur coup</i>
			<b>Total</b>	10121	

Est appelée *expression* toute réalisation d'une structure NpN à partir d'un substantif particulier : *page après page, d'homme à homme*.

Le motif est défini 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 » (Longrée, Luong, Mellet 2008).

Sur le plan formel, les structures NpN peuvent entrer dans la catégorie du motif syntaxique, qui est une des catégories admises du motif, à côté de celles constituées par les suites phono-

1 D. Longrée, X. Luong, S. Mellet (2008). « Les motifs : un outil pour la caractérisation topologique des textes », JADT, 733-744.

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 ».

logiques ou encore métriques. La suite choisie présente la particularité que les noms sont théoriquement réversibles puisqu'ils sont identiques ; cependant, le processus de la répétition du même lexème à droite de la préposition et l'inscription de ce dernier dans la chaîne linéaire phrastique introduit *a priori* une variation sémantique d'une occurrence à l'autre. Associé à la conception topologique des textes, la réalisation d'un motif ne peut être indifférente à l'agencement des unités ; de fait, la position du nom, à gauche ou à droite de la préposition, donne une orientation à la structure et l'inscrit dans la temporalité, qui est d'abord celle de la lecture cursive de la structure.

La structure choisie se situe entre langue et discours. Le chercheur a à composer avec les lexies ou unités phraséologiques introduites dans le discours et les créations de séquences ternaires sur un patron linguistique préconstruit. L'analyse croise 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 une figure comme la synecdoque « nez à nez », admise comme équivalent approximatif d'une locution adverbiale telle « très près ».

Tout petit, encore sans geste et sans parole ; et deux mioches de deux et trois ans, le derrière dans la terre, *nez à nez*, jardinent de leurs mains maladroites et se jettent des poignées de poussière dans la figure. (Maupassant, *Contes et nouvelles*, 1884).

Cependant, la mise en contexte de la structure peut réactiver la figure et susciter un contraste susceptible de créer un effet. C'est à ce titre que la séquence « N préposition N », même si, dans certains cas, elle peut être envisagée comme lexie, se situe toujours entre langue et discours et, en tout cas, peut être investie d'un rôle discursif qui la constitue en motif.

Lorsque je me retrouverai, là-haut, *nez à nez*, avec l'âme de mon vieux fermier. (J. Sandeau)

## 2.2. Fréquence et récurrence des structures NpN

Les critères définis associent à des paramètres quantitatifs comme la fréquence des propriétés textuelles fondées sur la récurrence qui doit permettre la mémorisation du motif et son rôle dans l'agencement discursif susceptible de caractériser certains usages.

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		

## 3. L'unité sémantique

Les prépositions sont, pour la plupart, polyvalentes et polysémiques. C'est le sens des substantifs environnants qui détermine le sens de la relation sémantique instaurée entre les noms et, par conséquent, le sens de l'expression.

L'examen détaillé des expressions NpN du corpus permet d'établir une liste de huit structures discriminées par le paramètre sémantique<sup>3</sup> :

*Succession* : décomposition du procès en éléments

Et la rouille souderait lentement, *saison après saison*, leurs carrosseries immobiles (Le Clezio) ; [...] et je vois se ronger *atome par atome* le grès des pyramides (Flaubert).

*Accumulation* : procès d'addition d'objets

[...] il lit beaucoup, dévore *livre après livre* avec une avidité juvénile [...] (Gide) ; je faisais *bronchite sur bronchite* quand j'étais bébé (Queffelec).

*Déplacement* : (topographique/chronologique) / mouvement

*D'année en année*, cette âme s'est desséchée de plus en plus (Hugo) ; Il allait, indécis, *de merveille en merveille* [...] (Bouilhet).

*Distance* : topographique/chronologique / non-mouvement

Cinq mois, *jour pour jour*, après avoir été séparé de son fils [...]. (Dumas) ; Vie rapide, très rapide, *de jeudi à jeudi*, du 28 au 5 (Michelet).

*Relation* : entre sujets (dont synecdoque)

Nous parlons *cœur à cœur* de la condition misérable des écrivains vraiment catholiques [...] (Bloy) ; Il traite tout le monde *d'égal à égal* [...] (Huysmans).

*Contact* : entre parties de sujets

Nous nous parlions *bouche à bouche* et je sentais son haleine sur mes lèvres (Beauvoir) ; Et me voilà *visage contre visage*, avec cette jeune fille, *jambes contre jambes* [...] (Goncourt).

*Opposition* : entre parties de sujets

Rocamboles avait lutté *corps à corps* avec un ours [...] (Ponson du Terrail) ; [...] combattre *homme contre homme* ou *page contre page* [...] (Hugo).

*Échange* : échange d'objets

De là elle rendit *injures pour injures* à ses ennemies (Stendhal) ; [il] rend *haine contre haine* et *dédain pour dédain* (Musset).

Structure	à	après	Contre	de ... à	de ... en	par	pour	sur	Total
<b>Accumulation</b>	5	28	2	0	0	15	0	253	303
<b>Contact</b>	1790	0	190	2	0	0	0	51	2033
<b>Déplacement</b>	0	0	0	25	4222	0	0	0	4247
<b>Distance</b>	1	0	0	6	0	0	74	0	81
<b>Échange</b>	0	0	11	0	0	6	324	0	341
<b>Opposition</b>	171	0	50	0	0	0	0	0	221
<b>Relation</b>	30	0	0	329	0	0	0	0	359
<b>Succession</b>	994	219	0	5	0	948	0	359	2525
<b>Total</b>	2991	247	253	367	4222	969	398	663	10110

Tableau des fréquences selon les structures et les sémantismes

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

On observe la polysémie des structures ou leur équivalence sémantique.

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

a. Une structure qui combine une même préposition avec des noms différents peut se charger de sens différents, preuve que la préposition n'a pas un sens absolu mais toujours dépendant des combinaisons où elle entre. Son sens se module en interaction avec le substantif répété de part et d'autre. La préposition « contre » par exemple peut exprimer trois types de relations sémantiques : l'*Opposition* : « acier contre acier » ; le *Contact* : « bouche contre bouche » ; l'*Échange* : « rendre haine contre haine ».

La structure N *contre* N peut être dite polysémique.

b. 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 sont maintenant sollicitées pour valider ou non ces hypothèses de travail intuitives.

Structure	à	après	contre	d(e) ... à	d(e) ... en	par	pour	sur	Total
<b>Accumulation</b>	0,17%	11,34%	0,79%			1,55%		38,16%	3,00%
<b>Contact</b>	<b>59,85%</b>		<b>75,10%</b>	0,54%				7,69%	20,11%
<b>Déplacement</b>				6,81%	<b>100,00%</b>				42,01%
<b>Distance</b>	0,03%			1,63%			18,59%		0,80%
<b>Échange</b>			4,35%			0,62%	<b>81,41%</b>		3,37%
<b>Opposition</b>	5,72%		19,76%						2,19%
<b>Relation</b>	1,00%			<b>89,65%</b>					3,55%
<b>Succession</b>	<b>33,23%</b>	<b>88,66%</b>		1,36%		<b>97,83%</b>		<b>54,15%</b>	24,98%

Tableau des pourcentages des sémantismes selon les structures

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

– Les structures les plus polysémiques sont *à* (6 sémantismes lui sont associés), *de ... à* (5) et *contre* (4).

– En revanche, *De ... en* est monosémique et indique essentiellement le déplacement.

– Chaque structure a un sémantisme nettement dominant, intrinsèque, sauf *à* (*Contact* et *Succession*), *contre* (*Contact* et *Opposition*) et *sur* (*Accumulation* et *Succession*).

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

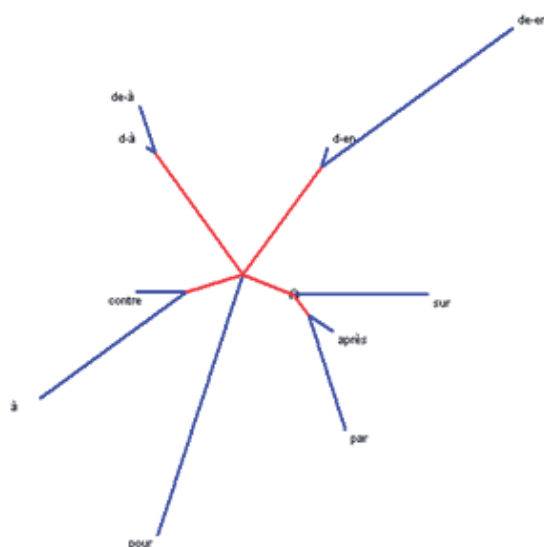
Tableau des pourcentages des structures selon les sémantismes

Ce tableau présente les différentes réalisations concrètes (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, liées en particulier par la succession chronologique qu'elles impliquent toutes deux : il y a une quasi-polyvalence des structures, à des degrés divers.

– Chaque sémantisme a une structure privilégiée, sauf la *Succession*, qui s'accommode à parts à peu près égales de la structure avec *à* ou avec *par* ; la primauté d'une structure est moins nette pour l'*Opposition*.

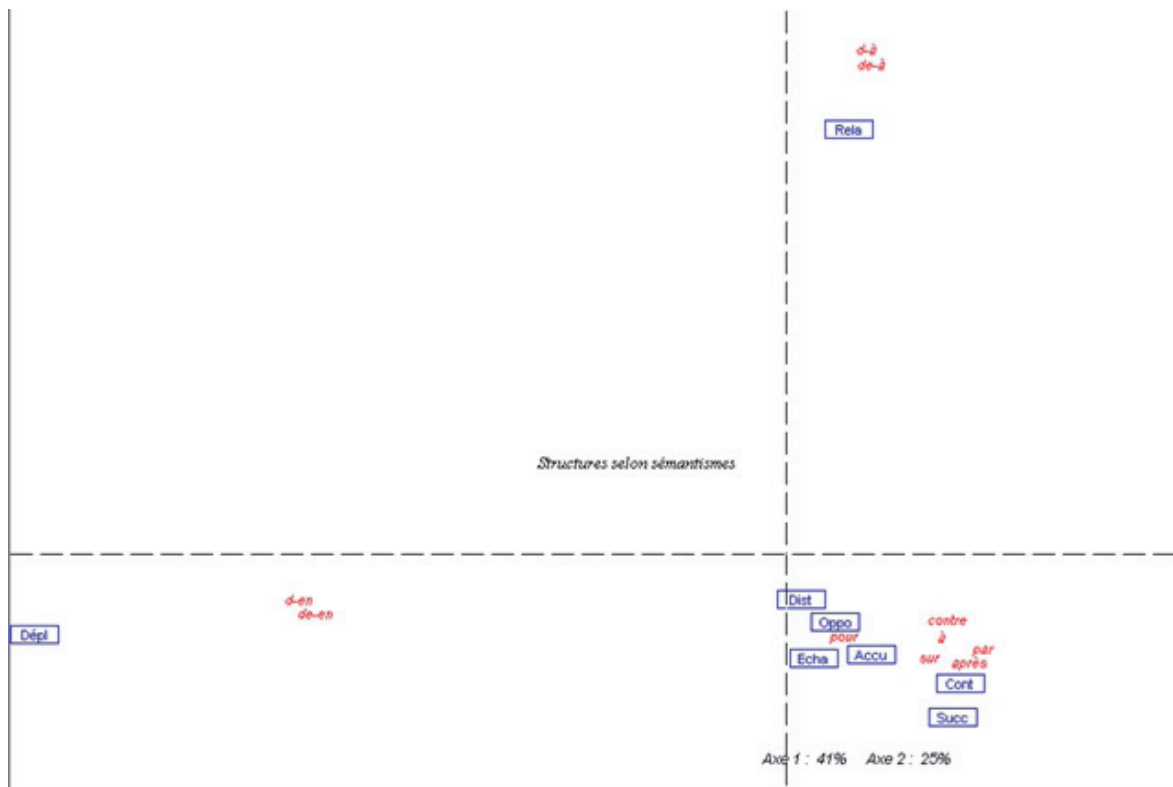
La synonymie des structures est vérifiée par les résultats statistiques mais elle est quantitativement limitée, eu égard aux sémantismes dominants associés à une structure, et ne porte de manière significative que sur deux paires de structures, comme le montre le graphique 1.



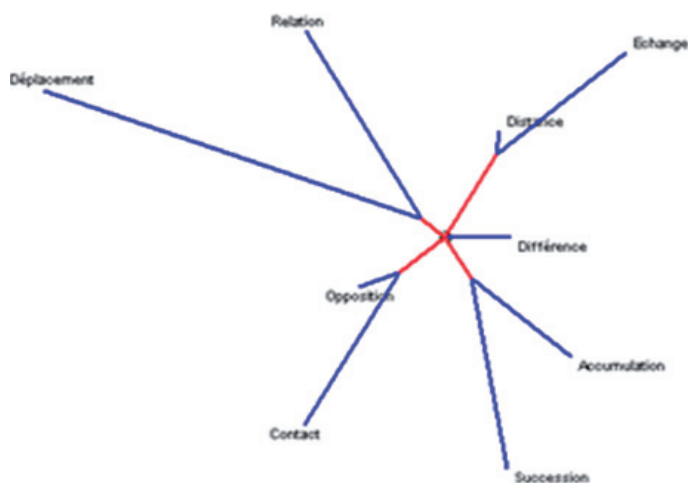
Graphique 1 : Les structures selon les sémantismes

L'analyse arborée montre le rapprochement sur une même branche des prépositions *par* et *après*, *à* et *contre*, signe que les structures 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.

La polysémie des structures est réelle, comme le montre le graphique 2 : seuls se dégagent les liens *de ... à* – *Relation* et *de ... en* – *Déplacement*. Mais elle est aussi limitée : le graphique arboré 3 montre que partagent surtout les mêmes structures *à* (*Contact* et *Opposition*) et *pour* (*Distance* et *Échange*).



Graphique 2 : Analyse factorielle des correspondances qui croise sémantismes et structures



Graphique 3 : Les sémantismes selon les structures

Il y a donc deux phénomènes à observer : d'une part, la polysémie et la polyvalence partielle des structures, d'autre part, la tendance à la spécialisation ; ces deux tendances sont bien marquées sur le graphique 2.

Il est donc possible d'envisager le motif NpN soit à partir des 8 structures, soit à partir des 8 sémantismes (cf. plus loin étude selon les genres).

#### 4. Rendement des structures

Le rendement des structures peut être évalué au travers de deux paramètres, leur fréquence et leur productivité : une structure est-elle plus ou moins propre à générer des expressions que les autres ?

##### 4.1. Fréquence

Est calculée 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).

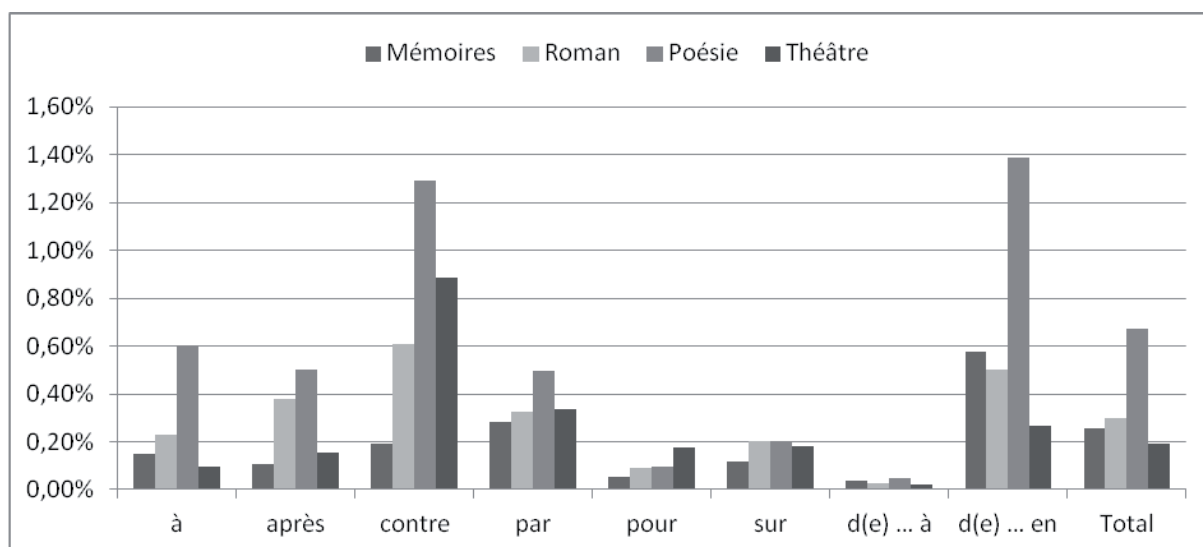
Effectifs des prépositions (A)						Effectifs des structures NPN (B)					
Préposition	Total	Mémoires	Roman	Poésie	Théâtre	Structure	Total	Mémoires	Roman	Poésie	Théâtre
à	1341642	235867	976759	59063	69953	à	2997	354	2219	356	68
après	75393	13171	56448	2590	3184	après	247	14	215	13	5
contre	44305	8392	32101	1781	2031	contre	253	16	196	23	18
par	295617	56121	213080	14505	11911	par	969	160	697	72	40
pour	437191	75488	315216	21746	24741	pour	398	42	292	21	43
sur	349623	52075	255874	26589	15085	sur	663	62	520	54	27
à	1341642	235867	976759	59063	69953	de ... à	372	84	246	27	15
en	772517	124192	570613	39593	38119	de ... en	4222	718	2854	549	101
Total	3316288	565306	2420091	165867	165024	Total	10121	1450	7239	1115	317



Pourcentages des effectifs B dans A					
Structure	Total	Mémoires	Roman	Poésie	Théâtre
À	0,22%	0,15%	0,23%	0,60%	0,10%
après	0,33%	0,11%	0,38%	0,50%	0,16%
contre	0,57%	0,19%	0,61%	1,29%	0,89%
par	0,33%	0,29%	0,33%	0,50%	0,34%
pour	0,09%	0,06%	0,09%	0,10%	0,17%
sur	0,19%	0,12%	0,20%	0,20%	0,18%
de ... à	0,03%	0,04%	0,03%	0,05%	0,02%
de ... en	0,55%	0,58%	0,50%	1,39%	0,26%
Total	0,31%	0,26%	0,30%	0,67%	0,19%

On voit :

- que les prépositions les plus productives en NpN (effectifs) sont *contre* et *en* ;
- que *à* est si fréquent et polyvalent qu’il ne produit pas autant de NpN ;
- que la productivité d’une préposition n’est pas uniforme entre les genres.



Graphique 4 : Les structures selon les genres

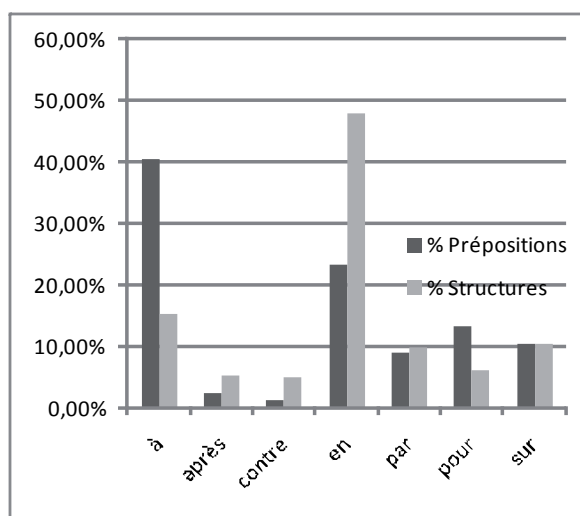
Le graphique 4 montre que la poésie est le genre qui mobilise la plus grande partie des effectifs de ses prépositions pour des structures NpN : c’est vrai pour *à*, *après*, *contre*, *par* et *en*.

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

#### 4.2. Productivité

Pour évaluer la productivité des structures, on établit une autre comparaison entre la fréquence des prépositions et le nombre d'expressions différentes produites par les structures NPN. On calcule alors d'une part la proportion des effectifs de chaque préposition sur le total des effectifs des 7 prépositions, d'autre part 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	



Graphique 5 : La productivité des structures

À et *en* sont les deux prépositions les plus fréquentes dans le corpus, mais leur productivité est inverse : à génère nettement 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*.

#### 4.3. La création lexicale

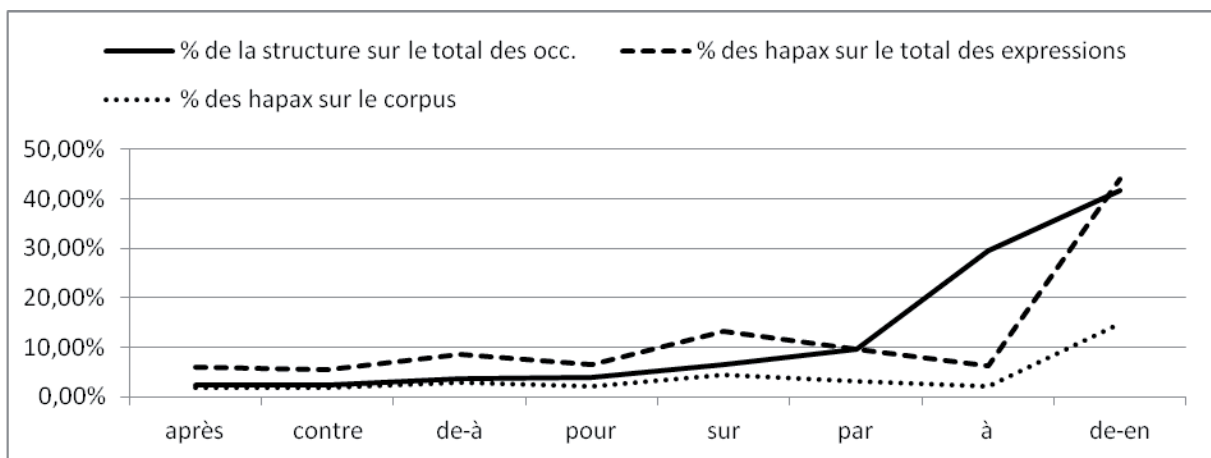
La création de nouvelles expressions est un fait de discours susceptible d'être interprété comme fait de style. Elle peut se faire par analogie à partir de syntagmes figés en langue. Par exemple, si *côte à côte* est identifié comme locution adverbiale par le TLF, *frange à frange* est attesté en discours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

Faute de pouvoir établir facilement et avec certitude quelles structures existent en langue et lesquelles sont des créations, partons d'un critère statistique sévère :

- les plus fréquentes sont forcément figées jusqu'à se grammaticaliser pour certaines et à fonctionner le plus souvent comme locutions adverbiales ;
- moins une expression est attestée, plus elle échappe à la lexicalisation.

De ce postulat, on déduira que les hapax sont des indices de productivité ; pour une structure, compter le nombre d'expressions de fréquence 1 dans le corpus établit une nouvelle proportion entre le nombre d'hapax et le total des expressions de la structure. Le comptage se fait sur le « vocabulaire » : combien d'expressions différentes sont hapax et combien ne sont pas hapax ?

On pourrait penser que la proportion des expressions 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 grand. C'est effectivement le cas, mais avec les exceptions qui indiquent une plus grande productivité (autrement dit une plus grande capacité à produire des hapax) pour une structure et le phénomène inverse pour une autre :



Graphique 6 : La productivité des structures et les hapax

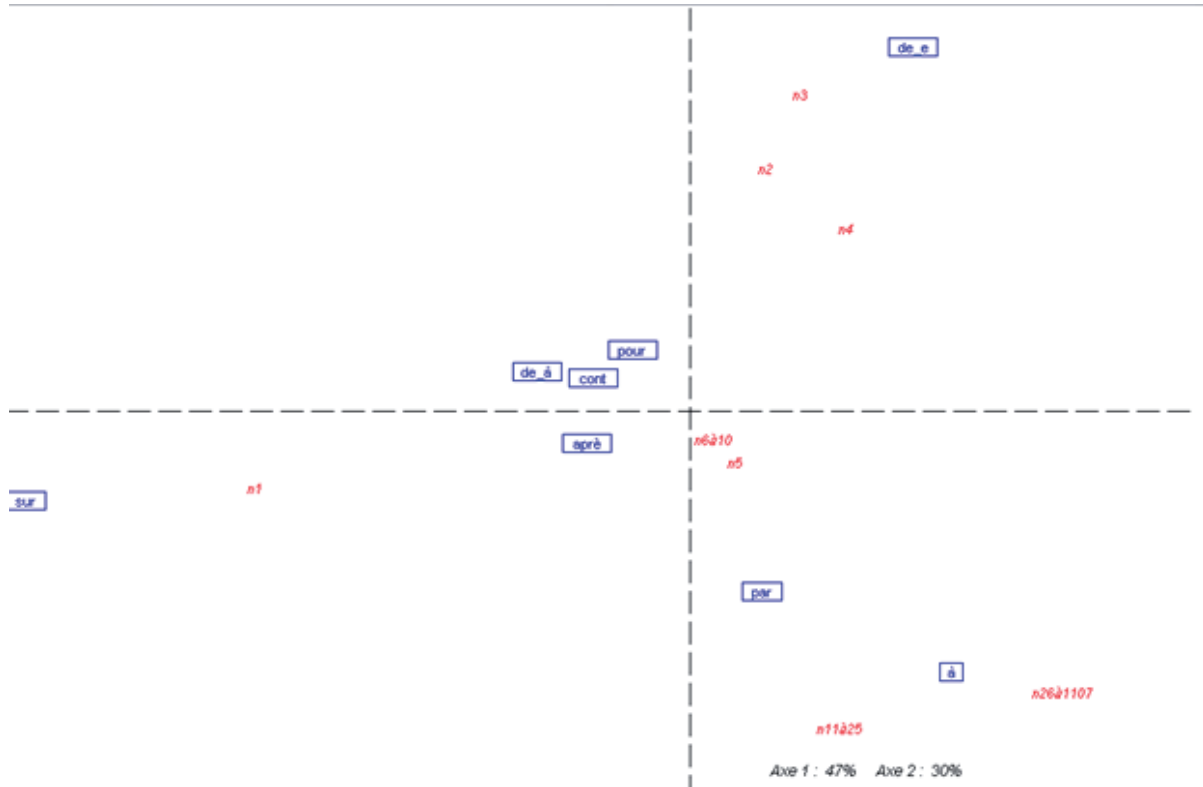
Le graphique 6 est calculé à partir de :

- la proportion de chaque structure sur le total des occurrences des NpN, ce qui donne un ordre de grandeur, de la moins fréquente (*après*) à la plus fréquente (*de ... en*) ;
- la proportion des hapax d'une structure sur le total de ses expressions, sans tenir compte de leurs effectifs : chaque expression est comptée pour une, et le calcul porte sur les hapax parmi celles-ci ;
- la proportion des hapax de chaque expression dans tout le corpus des NpN.

On voit que les courbes sont assez parallèles, mais :

- la structure *sur* produit plus d'hapax que ne le laisse attendre sa fréquence absolue, comparée aux autres structures ;
- la structure *à* accuse un déficit en hapax : elle est moins productive et plus riche en expressions figées ;

Le graphique AFC ci-dessous confirme ces observations :



Graphique 7 : AFC des hapax et non-hapax par structures

Les hapax (n1) sont du même côté que *sur*, la structure *à* est dans le même quadrant que les expressions à haute fréquences.

## 5. Les genres

Les paramètres observés jusqu'à présent, sémantisme et productivité, méritent d'être affinés par une approche de la population des structures NpN selon les genres.

Les effectifs globaux de chaque structure sont donnés dans les premiers tableaux.

### 5.1. La productivité des structures selon les genres

Certains genres entretiennent des affinités avec certaines structures (graphique 8) :

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*.

L'observation vaut surtout si l'on détaille les données en distinguant hapax et non-hapax (graphique 9) :

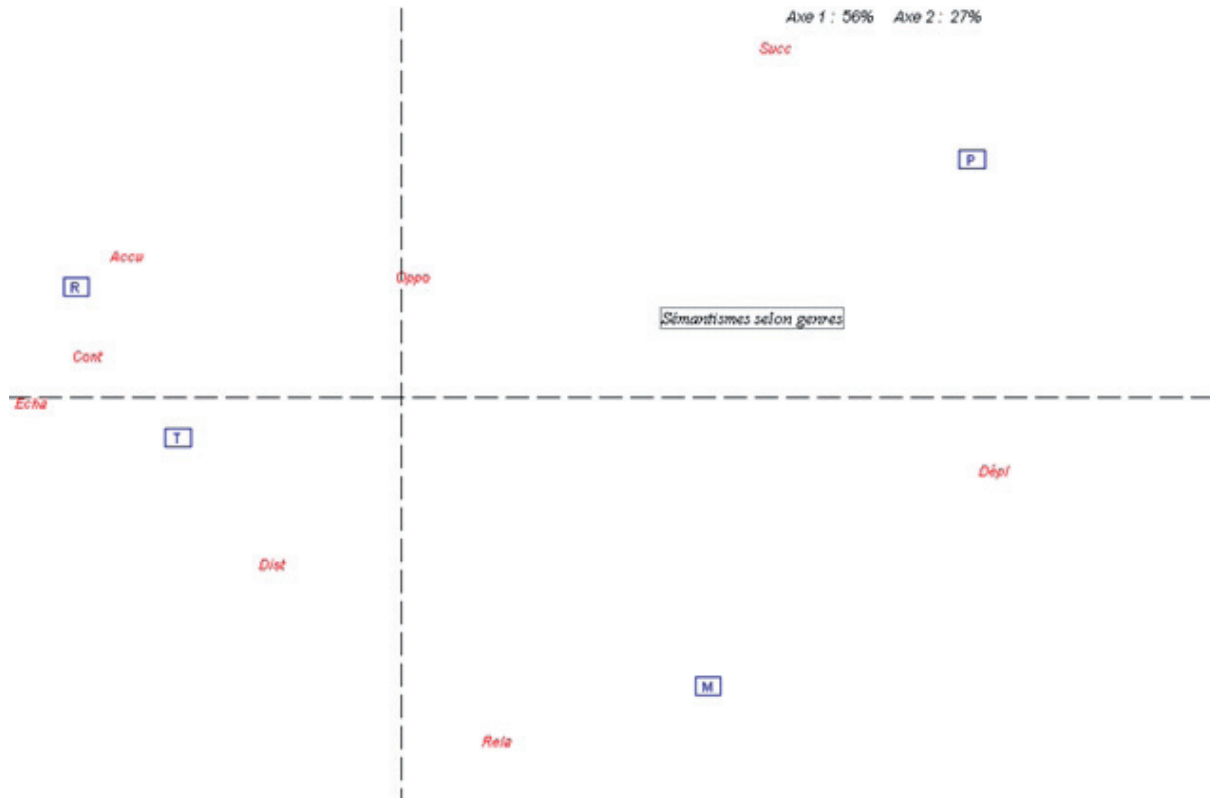
Ceci constitue un deuxième indice de l'affinité d'un genre parmi les autres pour le motif NpN : la poésie est le genre le plus propice à la création d'expressions nouvelles à partir des structures *contre*, *sur*, *par*, *à* et *de...en*.



On voit que la plupart des structures, dans les occurrences d'expressions hapax et non-hapax, restent en place 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*.

### 5.2. Les sémantismes et les genres

L'affinité des genres et des sémantismes est représentée dans le graphique 10.



Graphique 10 : Les sémantismes selon les genres

Cette AFC ne double qu'en partie celle des structures, en fonction des sémantismes privilégiés de chacune d'entre elle. 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.

## Conclusion

À la question qui était posée en introduction, à savoir la recevabilité de la candidature de la structure NpN au rang de motif, la réponse est positive. Les critères de fréquence en langue et de récurrence dans le corpus sont vérifiés. La mémorisation est aisée, soit que l'expression soit lexicalisée, soit que le lecteur soit apte à reconnaître un pattern syntaxique par analogie avec les expressions figées. La productivité à partir de schémas préconstruits est variable mais puissante, tout en évoluant autour de huit sémantismes identifiés. De fait, ce motif se situe entre langue et discours et se développe comme motif discursif susceptible de produire un effet stylistique

corrélé à l'innovation lexicale, dont les hapax sont un indice. Le motif peut jouer un rôle dans la caractérisation générique, puisqu'a été observée une répartition claire, même si elle ne peut être complètement interprétée ici, entre les genres. La poésie, notamment, se caractérise par les observations suivantes : c'est le genre qui mobilise la plus grande part des effectifs 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. Ces remarques strictement factuelles sont des amorces à d'autres études plus linguistiques qui pourront être développées afin d'affiner et d'interpréter ces résultats.

## Bibliographie

- Longrée (D.), Luong (X.), Mellet (M.), 2008 : « Les motifs : un outil pour la caractérisation topologique des textes », *JADT*, Lyon, pp. 733-744.
- Magri-Mourgues (Véronique), 2010 : « Distance intertextuelle et connexion lexicale : outils de catégorisation générique ou stylistique ? Approche expérimentale d'un corpus inédit : le corpus aragonien », *JADT, Sapienza University of Rome – Statistical Analysis of Textual Data Proceedings of the 10th International Conference*, édité par Sergio Bolasco, Milan : Edizioni Universitarie di Lettere Economia Diritto, p. 333-340.
- Magri-Mourgues (Véronique), 2008 : « Sémantique générique et statistique », *JADT*, Lyon : PUL, p. 753-764.
- Magri-Mourgues (Véronique), 2006 : « Stylistique générique et statistique », *JADT*, Besançon : « Les Cahiers de la MSH Ledoux », p. 655-666.