



**HAL**  
open science

## ”Mot à mot, brin par brin” : les suites [Nom préposition Nom] comme indices de littéarité ?

Véronique Magri, Gérald Purnelle, Dominique Legallois

### ► To cite this version:

Véronique Magri, Gérald Purnelle, Dominique Legallois. ”Mot à mot, brin par brin” : les suites [Nom préposition Nom] comme indices de littéarité ?. JADT 2016 - Statistical Analysis of Textual Data, Jun 2016, Nice, France. pp.365-376. hal-01371226

**HAL Id: hal-01371226**

**<https://hal.science/hal-01371226v1>**

Submitted on 26 Sep 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ***Mot à mot, brin par brin :* les suites [Nom préposition Nom] comme indices de littéarité ?**

Véronique Magri<sup>1</sup>, Gérald Purnelle<sup>2</sup>  
& Dominique Legallois<sup>3</sup>

<sup>1</sup> Université Nice Sophia Antipolis, UCA, BCL– France – veronique.magri@unice.fr

<sup>2</sup> Université de Liège – Belgique – gerald.purnelle@ulg.ac.be

<sup>3</sup> Université Paris 3 Sorbonne Nouvelle, Clesthia – France– dominique.legallois@univ-paris3.fr

## **Abstract**

For the JADT 2012, we applied the text pattern theory to Noun-preposition-Noun (NpN) sequences, where both nouns surrounding the preposition are identical. Our working corpus was composed of texts from the four main literary genres, memoirs (109 texts), novels (740 texts), poetry (191 texts) and theater (210 texts), 85 million words, sampled out of *Frantext*, a categorized database. That exploratory research allowed us to pinpoint the specific place of the NpN sequence, between language and discourse, its blooming into a full discursive pattern, able to produce a stylistic effect linked to lexical innovation, as well as its possible use as a generic litterarity marker. This paper aims at pursuing the work by extending the corpus and proposing new contrastive studies, showing the role of the NpN sequence as a litterarity marker.

## **Résumé**

Lors des JADT 2012, la théorie du motif avait été évaluée à l'aune des séquences Nom-préposition-Nom, restreintes aux cas où les deux noms sont identiques. Le corpus était alors constitué d'un extrait de *Frantext*, base catégorisée, qui offrait un échantillon des quatre genres littéraires principaux, les mémoires (109 textes), les romans (740 textes), la poésie (191 textes) et le théâtre (210 textes), soit près de 85 millions de mots. Ce premier travail exploratoire a montré que le motif NpN se situe entre langue et discours et se développe comme motif discursif susceptible de produire un effet stylistique corrélé à l'innovation lexicale. Il intervient aussi comme possible marqueur générique.

La présente communication entend poursuivre cette recherche en développant le corpus de recherche et en proposant de nouvelles études contrastives, susceptibles d'évaluer la suite NpN comme marqueur de littéarité.

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

Lors des JADT 2012, la théorie du motif<sup>1</sup> avait été évaluée à l'aune des séquences Nom-préposition-Nom, restreintes aux cas où les deux noms sont identiques. Rappelons la définition du motif donnée par Longrée et Mellet (2013) :

Sur le plan fonctionnel, le motif est un « cadre collocationnel » accueillant un ensemble d'éléments fixes et de variables, susceptible d'accompagner la structuration textuelle et, simultanément, de caractériser des textes de genres divers, voire de permettre la détection, au sein d'un même texte, des passages de registres différents. (Longrée et Mellet 2013 : 38)

La suite NpN a pu être définie comme exemple de motif syntaxique. Cependant, NpN a également fait l'objet d'une analyse syntaxique précise dans le cadre de la Grammaire de Construction (R. Jackendoff, 2008) : les propriétés analysées constituent un défi pour la théorisation grammaticale, puisque la structure NpN est particulièrement idiomatique (ni

---

<sup>1</sup> Voir Longrée, Luong & Mellet (2008). Voir également, Quiniou, Cellier, Charnois & Legallois (2012), Legallois, Charnois & Poibea. (à par.) pour des approches sensiblement différentes.

l'emploi de noms massifs, ni l'emploi de déterminants, ni l'emploi du pluriel, ni la modification adjectivale ne sont possibles). NpN constitue, de plus, un groupe nominal dont on ne peut préciser la véritable tête. Cette tête, pour Jackendoff serait (plutôt) la préposition elle-même, mais avec une violation du schéma X barre...ce qui a pour conséquence, justement, de faire perdre la tête au syntacticien soucieux de dégager des règles générales du fonctionnement des structures linguistiques, alors même que la suite se rencontre, outre le français et l'anglais, dans des langues très différentes (mandarin, russe, allemand, japonais, italien, polonais).

L'objectif de l'article des JADT de 2012 – ainsi que celui-ci – s'inscrivait dans un cadre différent, plus textualiste, dans la mesure où il s'agissait d'analyser la distribution de NpN selon les genres littéraires. Le corpus était alors constitué d'un extrait de *Frantext*, base catégorisée, qui offrait un échantillon des quatre genres littéraires principaux, les mémoires (109 textes), les romans (740 textes), la poésie (191 textes) et le théâtre (210 textes), soit près de 85 millions de mots. Ce premier travail exploratoire a montré que le motif NpN se situe entre langue et discours et se développe comme motif discursif susceptible de produire un effet stylistique corrélé à l'innovation lexicale. Il intervient aussi comme possible marqueur générique.

La présente communication entend poursuivre cette recherche en développant le corpus de travail et en proposant de nouvelles études contrastives, susceptibles d'évaluer la suite NpN comme marqueur de littérarité. Il est envisagé de confronter deux corpus, l'un composé d'articles journalistiques, l'autre de textes de fiction. Ces deux grands ensembles feront l'objet de partitions internes afin d'introduire des comparaisons plus fines. Le corpus médiatique se distribuera en deux sous-corpus, l'un constitué de deux quotidiens nationaux, *Le Monde* et *Libération*, l'autre de trois journaux régionaux, *L'Est Républicain*, *Ouest-France* et *Sud-Ouest*<sup>2</sup>, chacun des sous-corpus comptant environ 1,5 millions de mots. Le corpus fictionnel se compose de textes contemporains et se décline en sous-genres, comptant chacun entre 3 et 3,5 millions de mots ; il réunit 50 volumes de la collection *Harlequin* (publiés entre 2001 et 2012), 50 romans policiers (publiés entre 1980 et 2015) et une cinquantaine de romans dits sérieux (publiés entre 1980 et 2015) dont on trouvera la liste en annexe. Le premier corpus extrait de *Frantext* sera conservé en mémoire à titre de comparaison.

L'objectif est d'observer le comportement de la suite NpN dans ces différents corpus et de mesurer les variations d'emploi, le cas échéant, d'un corpus à l'autre, tout en posant la question de son fonctionnement comme indice éventuel de littérarité.

## 1. Répartition des structures

Il avait été constaté, lors du premier travail sur la suite NpN, que 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*. Cette même liste est conservée dans la présente recherche afin de pouvoir introduire des comparaisons.

Le tableau suivant présente les pourcentages dans les deux corpus, les journaux cumulés d'une part, les romans cumulés d'autre part :

---

<sup>2</sup> Les journaux ont été publiés entre 2010 et 2013.

Structure	Effectifs		Pourcentages	
	Journaux	Romans	Journaux	Romans
à	235	582	37,5%	44,1%
après	32	62	5,1%	4,7%
contre	7	11	1,1%	0,8%
de ... à	40	11	6,4%	0,8%
de ... en	155	478	24,7%	36,2%
par	86	81	13,7%	6,1%
pour	31	41	4,9%	3,1%
sur	41	53	6,5%	4,0%

Figure 1- Distribution des structures selon les deux partitions du corpus

Soumis à un test de Khi2, ce tableau révèle une probabilité quasi nulle pour l'hypothèse nulle ; les excès et les déficits statistiquement significatifs sont colorés (respectivement en jaune et bleu).

Ce tableau qui réduit le corpus à deux sous-ensembles permet de pointer les divergences d'emploi quant aux structures NpN entre journaux et textes fictionnels. Si les structures qui utilisent les prépositions « à », « après », « contre », « pour » sont à peu près équitablement réparties entre les deux corpus, des déséquilibres significatifs opposent l'emploi des structures qui utilisent « de ... à », « de ... en », « par » et « sur ». On constate en effet un sur-emploi, dans le corpus médiatique, des structures ordonnées autour de « de ... à », « par » et « sur » tandis que le privilège est accordé aux textes fictionnels pour celles organisées autour de « de ... en ». La structure « de ... à » est illustrée par *de clou à clou (L'Est Républicain)*, *de monastère à monastère*, *de peau à peau (Le Figaro)*, *de bande à bande*, *de gré à gré*, *de personne à personne*, *de baguettes à baguettes*, *de banlieue à banlieue*, *de cœur à cœur*, *de verrier à verrier (Libération)*, *de fibre à fibre*, *de ligne à ligne*, *de particulier à particulier*, *de poste à poste (Le Monde)*, *de gré à gré (Le Monde, Sud-Ouest, Ouest-France)*.

La préposition « par » : *point par point*, *cas par cas (L'Est Républicain)*, *hôpital par hôpital*, *jour par jour*, *pays par pays (Le Figaro)*, *cas par cas*, *coup par coup*, *état par état*, *image par image*, *mètre par mètre*, *minute par minute*, *pièce par pièce*, *plan par plan*, *point par point*, *semaine par semaine*, *touche par touche (Libération)*, *cas par cas*, *entreprise par entreprise*, *point par point*, *brique par brique*, *coup par coup*, *heure par heure*, *phrase par phrase*, *secteur par secteur (Le Monde)*, *cas par cas*, *coup par coup*, *heure par heure*, *pied par pied*, *pierre par pierre*, *poste par poste*, *région par région (Sud-Ouest)*, *coup par coup*, *entreprise par entreprise*, *point par point*, *sujet par sujet*, *bateau par bateau*, *brique par brique*, *élève par élève*, *heure par heure*, *légume par légume*, *minute par minute*, *phrase par phrase*, *port par port*, *secteur par secteur (Ouest-France)*.

La préposition « sur » : *coup sur coup*, *film sur film*, *podium sur podium (L'Est Républicain)*, *coup sur coup*, *cigarette sur cigarette*, *rôle sur rôle (Le Figaro)*, *coup sur coup*, *nicorette sur nicorette*, *but sur but*, *cigarette sur cigarette*, *pièce sur pièce*, *ton sur ton (Libération)*, *coup sur coup*, *accident sur accident*, *colloque sur colloque*, *procès sur procès*, *ton sur ton (Le Monde)*, *coup sur coup*, *frites sur frites*, *nom sur nom*, *retard sur retard (Sud-Ouest)*, *coup sur coup*, *bêtise sur bêtise*, *bord sur bord*, *colloque sur colloque*, *mévente sur mévente*, *passé sur passé*, *procès sur procès*, *télé sur télé*, *ton sur ton*, *victoire sur victoire (Ouest-France)*.

Le détail des corpus présenté dans le tableau ci-après permet d'affiner les commentaires et de faire apparaître des résultats contrastifs au sein d'un même ensemble générique.

Structure	Sud-Ouest	Est Républicain	Figaro	Libération	Monde	Ouest-France	R-Harlequin	R-Policiers	R-Sérieux	Total général
à	41,0%	39,6%	35,9%	41,3%	28,7%	40,9%	70,3%	45,9%	29,7%	42,0%
après	3,8%	0,0%	3,9%	9,8%	6,2%	5,3%	3,7%	2,4%	6,3%	4,8%
contre	0,0%	0,0%	2,9%	1,4%	1,6%	0,5%	2,4%	0,7%	0,2%	0,9%
de ... à	1,9%	14,6%	1,9%	8,4%	7,0%	1,0%	0,0%	1,7%	0,8%	2,6%
de ... en	34,6%	31,3%	14,6%	14,7%	27,1%	22,1%	15,5%	34,8%	54,3%	32,5%
par	6,4%	5,2%	28,2%	12,6%	18,6%	18,3%	1,7%	6,9%	3,4%	8,6%
pour	5,1%	4,2%	5,8%	3,5%	6,2%	4,8%	5,1%	2,8%	1,5%	3,7%
sur	7,1%	5,2%	6,8%	8,4%	4,7%	7,2%	1,4%	4,8%	3,8%	4,8%

Figure 2- Distribution des structures selon les sous-genres

Pour les prépositions équitablement réparties, globalement, dans les deux ensembles génériques, on constate cependant que certains sous-corpus se distinguent par un emploi spécifique, excédentaire ou déficitaire. C'est la collection *Harlequin* qui concentre le plus fort effectif de structures avec préposition « à », tandis que *Libération* attire les structures organisées autour de « après ». La préposition « contre » assure un passage transgénérique en étant prisée aussi bien du *Figaro* que du roman *Harlequin*.

La primauté de la structure « de ... à » est confirmée pour l'ensemble médiatique et, en particulier, par trois journaux, l'un régional *L'Est Républicain*, les deux autres nationaux, *Libération* et *Le Monde*. Parallèlement, la structure « de ... en », apparue comme caractéristique de l'ensemble fictionnel, se trouve sur-représentée dans le *Roman sérieux* tandis que la collection *Harlequin* marque le pas dans son camp générique.

L'analyse arborée suivante permet de visualiser les affinités des partitions du corpus quant à l'emploi des structures NpN et l'analyse factorielle précise ces affinités en affichant les proximités entre les prépositions pivots des structures.

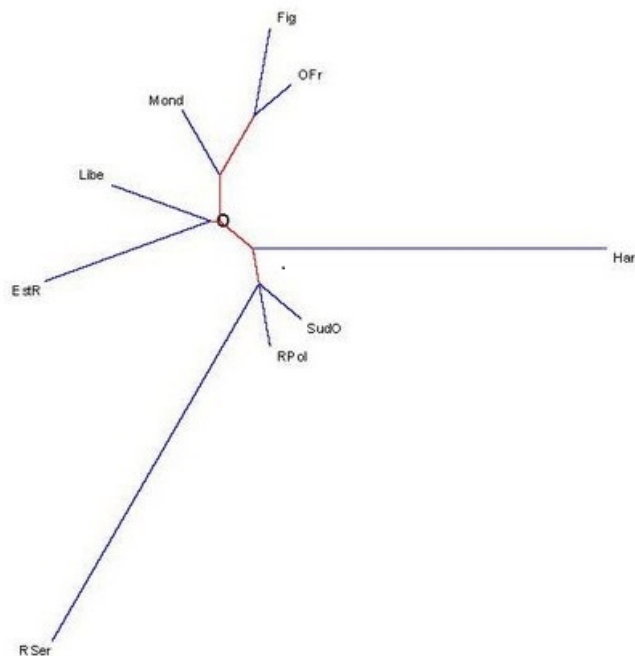


Figure 3- Analyse arborée de la distribution des structures selon les sous-genres

La figure 3 montre une faible structuration du corpus (deux journaux, *Libération* et *Est républicain*, sont proches de la racine) : un groupe relativement homogène sur le plan générique se dégage, regroupant deux journaux nationaux, *Le Monde* et *Le Figaro* mais aussi un journal

régional *Ouest-France*, tandis que le journal *Sud-Ouest* en est éloigné. Le sous-corpus littéraire est nettement moins homogène : *Harlequin* et les romans sérieux sont très distants des nœuds, et le premier est proche de la racine. Toutefois, on peut constater que si le roman sérieux est porté par la même branche principale que le roman policier, tout en affichant ses particularités par la longueur de la branche secondaire, la collection *Harlequin* se marginalise tant par rapport au corpus médiatique que par rapport au corpus littéraire.

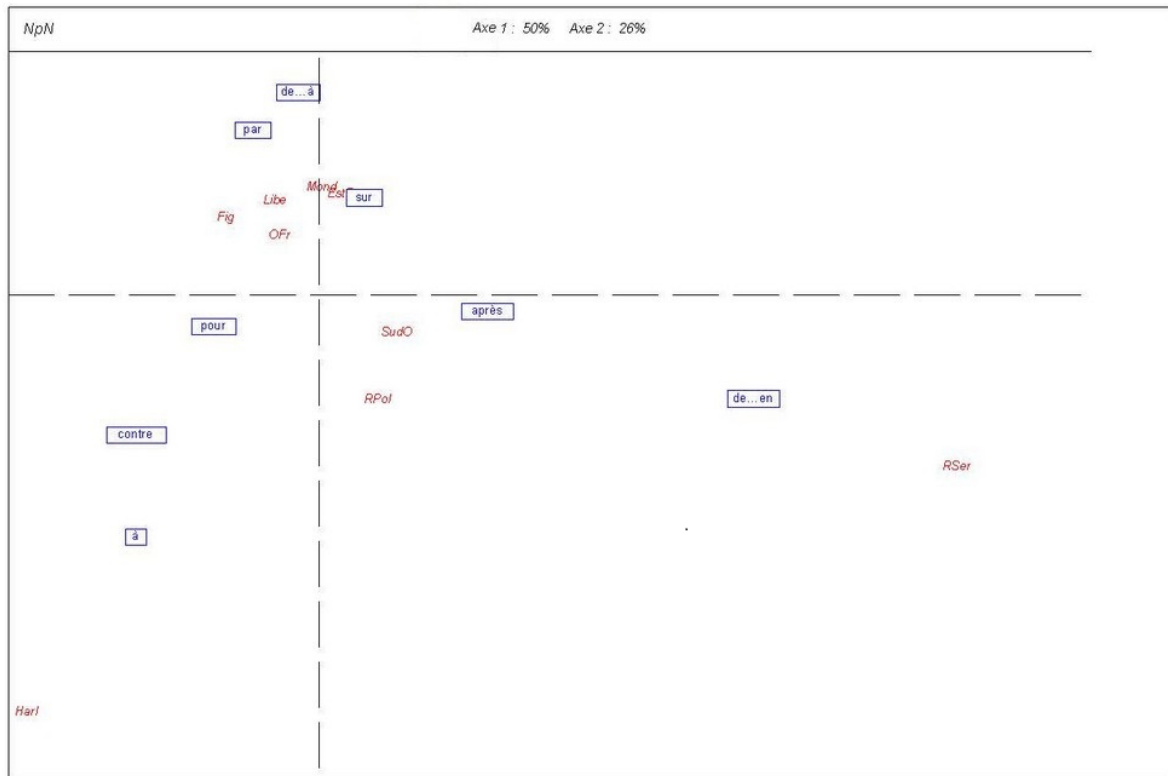


Figure 4- La répartition des structures selon les sous-genres

L'axe 1 de l'AFC montre clairement une partition du corpus entre sous-corpus journalistiques et sous-corpus littéraire, avec le dissident *Sud-Ouest* proche de la racine du graphique et présentant une certaine neutralité quant à l'usage des différentes structures NpN. L'axe 2 confirme l'éclatement du sous-corpus littéraire au regard de cet usage : *Harlequin* et les romans sérieux sont très fortement excentrés et opposés l'un à l'autre. La fréquence de telles structures apparaît comme représentative de styles différents : au plus simple à (*tête à tête*) s'oppose le plus complexe de ... en (*de génération en génération*) et le plus sophistiqué après (*feuille après feuille*).

Le corpus est fortement structuré par l'opposition de deux sous-genres littéraires, le *Roman sérieux* dans le quadrant sud-est et la collection *Harlequin* dans le quadrant sud-ouest. Les motifs NpN s'avèrent donc, dans un premier temps, nettement plus rentables pour distinguer les sous-genres au sein de l'écriture de type littéraire que pour distinguer les genres lourds tels que journalisme et roman.

## 2. Le 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 ?

### 2.1. La fréquence

Est calculée la proportion des occurrences des structures NpN sur les effectifs totaux de chaque préposition dans le corpus.

<b>Effectifs des prépositions (A)</b>									
	Sud-Ouest	L'Est Républicain	Figaro	Libération	Monde	Ouest-France	R-Harlequin	R-Policiers	R-Sérieux
à	39147	49642	31114	31879	31445	35467	81994	84677	81996
après	2037	1865	2369	2021	1845	2141	3932	3419	3071
contre	982	847	1666	1355	1375	1237	2618	1917	1686
à	39147	49642	31114	31879	31445	35467	81994	84677	81996
en	20731	19052	20444	21434	21678	19745	32435	32636	28632
par	9134	9380	8775	9268	9995	8644	8735	10760	10278
pour	1570	15839	13153	13454	12592	13528	27423	22954	20505
sur	9638	8693	3026	9484	9114	9136	18264	21238	16624
<b>Effectifs des structures NpN (B)</b>									
Structure	Sud-Ouest	L'Est Républicain	Figaro	Libération	Monde	Ouest-France	R-Harlequin	R-Policiers	R-Sérieux
à	64	38	37	59	37	85	208	133	156
après	6		4	14	8	11	11	7	33
contre			3	2	2	1	7	2	1
de ... à	3	14	2	12	9	2		5	4
de ... en	54	30	15	21	35	46	46	101	285
par	10	5	29	18	24	38	5	20	18
pour	8	4	6	5	8	10	15	8	8
sur	11	5	7	12	6	15	4	14	20
<b>Pourcentages des effectifs B dans A</b>									
Structure	Sud-Ouest	L'Est Républicain	Figaro	Libération	Monde	Ouest-France	R-Harlequin	R-Policiers	R-Sérieux
à	0,16%	0,08%	0,12%	0,19%	0,12%	0,24%	0,25%	0,16%	0,19%
après	0,29%	0,00%	0,17%	0,69%	0,43%	0,51%	0,28%	0,20%	1,07%
contre	0,00%	0,00%	0,18%	0,15%	0,15%	0,08%	0,27%	0,10%	0,06%
de ... à	0,01%	0,03%	0,01%	0,04%	0,03%	0,01%	0,00%	0,01%	0,00%
de ... en	0,26%	0,16%	0,07%	0,10%	0,16%	0,23%	0,14%	0,31%	1,00%
par	0,11%	0,05%	0,33%	0,19%	0,24%	0,44%	0,06%	0,19%	0,18%
pour	0,51%	0,03%	0,05%	0,04%	0,06%	0,07%	0,05%	0,03%	0,04%
sur	0,11%	0,06%	0,23%	0,13%	0,07%	0,16%	0,02%	0,07%	0,12%

Figure 5- Un rapport : Les effectifs des structures par rapport aux effectifs des prépositions

Le seuil de 0,19 % est choisi comme seuil au-delà duquel la proportion peut être dite significative. La préposition « après » peut être dite la plus productive car ce seuil est dépassé pour 7 sous-corpus sur 9. C'est la préposition qui mobilise le plus grand nombre de ses occurrences pour les faire entrer dans des structures NpN ; elle est suivie de la structure « de ... en » pour laquelle le seuil est dépassé pour 4 sous-corpus. En revanche, les prépositions « contre », « pour » et « sur » apparaissent comme les moins productives en structures NpN. Pour mémoire, dans le corpus littéraire extrait de *Frantext*, c'était les prépositions « contre » et « de ... en » qui étaient apparues comme les plus productives de NpN. Par comparaison avec ce premier travail exploratoire, « de ... en » reste majoritaire dans deux sous-corpus fictionnels (*Harlequin* se détache de son clan générique), mais aussi dans deux journaux régionaux (*Sud-Ouest* et *Ouest-France*). On ne peut donc parler d'un partage générique clair. « Contre », qui

était la préposition la plus productive dans le corpus littéraire *Frantext*, ne confirme sa primauté dans ce nouveau corpus que pour la collection *Harlequin*.

Si on observe le comportement des sous-corpus relativement aux structures générées par les différentes prépositions, c'est-à-dire si on observe les colonnes plutôt que les lignes, on constate que *Ouest-France* est le journal pour lequel le plus grand nombre de prépositions entrent dans des structures NpN : « à », « après », « de ... en », « par ». Il est suivi par *Sud-Ouest* avec les prépositions « après », « de ... en », « pour » puis par un texte fictionnel, *Harlequin* avec les prépositions « à », « après », « contre ». En revanche, *L'Est Républicain* se distingue par le fait qu'aucune préposition ne se démarque de l'ensemble pour lui.

## 2.2. La 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

Pourcentages de chaque préposition sur toutes les prépositions										
	<i>Sud-Ouest</i>	<i>L'Est-Républicain</i>	<i>Figaro</i>	<i>Libération</i>	<i>Monde</i>	<i>Ouest-France</i>	R-Harlequin	R-Policiers	R-sérieux	Total
à	47,03%	47,14%	38,63%	35,86%	35,72%	39,45%	46,75%	47,68%	50,37%	44,44%
après	2,45%	1,77%	2,94%	2,27%	2,10%	2,38%	2,24%	1,93%	1,89%	2,16%
contre	1,18%	0,80%	2,07%	1,52%	1,56%	1,38%	1,49%	1,08%	1,04%	1,30%
en	24,91%	18,09%	25,38%	24,11%	24,62%	21,96%	18,49%	18,38%	17,59%	20,61%
par	10,97%	8,91%	10,89%	10,43%	11,35%	9,62%	4,98%	6,06%	6,31%	8,08%
pour	1,89%	15,04%	16,33%	15,13%	14,30%	15,05%	15,63%	12,92%	12,60%	13,41%
sur	11,58%	8,25%	3,76%	10,67%	10,35%	10,16%	10,41%	11,96%	10,21%	10,00%

Figure 6-La proportion de chaque préposition

« À » est la préposition la plus représentée parmi toutes les prépositions et cela quel que soit le sous-corpus observé. Une harmonie dans l'emploi des prépositions se dégage, quel que soit le sous-corpus. L'ordre décroissant de l'emploi des prépositions majoritaire est régulièrement constaté : « à », « en », « pour », « sur », « par », « après », « contre ». Seuls *Sud-Ouest*, *L'Est Républicain* et *Le Figaro* introduisent de légères variations, tout en conservant toujours en tête de liste « à » et « en » et en queue de liste la préposition « contre ».

Pourcentages sur le nombre d'expressions différentes										
Structure	Sud-Ouest	L'Est Républicain	Figaro	Libération	Monde	Ouest-France	R-Harlequin	R-Policiers	R-Sérieux	
à	31,25%	50,00%	31,11%	39,13%	31,34%	27,03%	26,67%	20,54%	17,16%	
après	4,69%	0,00%	8,89%	7,25%	4,48%	4,05%	11,11%	6,25%	15,67%	
contre	0,00%	0,00%	6,67%	2,90%	1,49%	1,35%	4,44%	1,79%	0,75%	
de ... en	39,06%	33,33%	11,11%	21,74%	32,84%	25,68%	26,67%	48,21%	43,28%	
par	10,94%	4,17%	26,67%	15,94%	13,43%	18,92%	8,89%	11,61%	8,96%	
pour	7,81%	6,25%	8,89%	4,35%	8,96%	9,46%	13,33%	3,57%	4,48%	
sur	6,25%	6,25%	6,67%	8,70%	7,46%	13,51%	8,89%	8,04%	9,70%	

Figure 7-La productivité des structures selon les sous-genres



Structure	Nombre d'expressions	%
à/de ... à	69	19,77%
après	33	9,46%
contre	10	2,87%
de ... en	135	38,68%
par	47	13,47%
pour	17	4,87%
sur	38	10,89%
Total général	349	

Figure 8- La productivité des structures dans le corpus

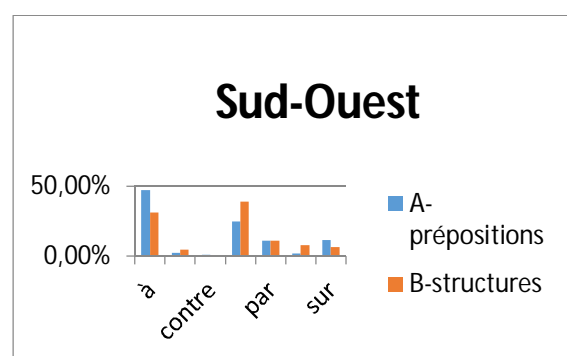
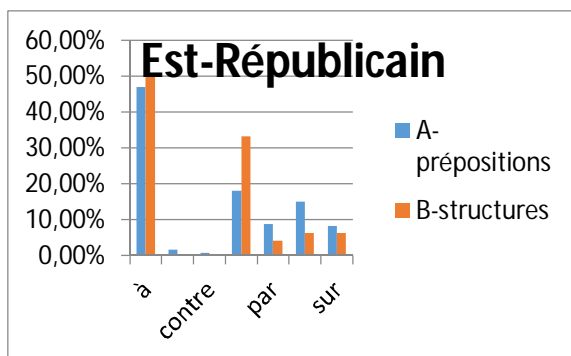
« De ... en » entre dans le plus grand nombre d'expressions différentes, tandis que « contre » ne sert qu'à la construction de 10 expressions différentes. Il avait été constaté des résultats similaires dans l'étude précédente sur l'extrait de *Frantext* : on avait en effet observé que « à » et « en » étaient les prépositions les plus fréquentes mais que leur productivité était inverse car « à » génère moins d'expressions différentes que « en » dans la structure « de ... en ». Une variété moindre des structures témoigne d'une spécialisation partielle des structures qui peut confiner à la lexicalisation ou à la productivité : par exemple, la préposition « après » qui se trouve reléguée en fin de liste dans l'ensemble des prépositions remonte juste après « de ... en » et « à » dans l'ensemble des structures générées par les prépositions. « Après » est une préposition productive. C'est le même constat qui avait été fait lors de la précédente recherche, ce qui paraît bien confirmer un phénomène de langue : la productivité de la préposition « après » dans les structures NpN. Liée essentiellement au sémantisme de la succession et, dans une moindre mesure, à celui de l'accumulation, « après » s'accommode de divers substantifs, notant la succession chronologique stricte (*année, mois, semaine, jour, matin, minute*), la succession dans l'espace (*millimètre*), ou événementielle (*match, solstice, photo*). C'est le *Roman sérieux* qui présente la panoplie la plus large de structures avec « après ». Outre les indications chronologiques, on trouve des noms plus insolites comme *barreau, grain, maille, pavé, vitrine*. On a vu plus haut que la spécificité et l'isolement des romans sérieux était liés aux deux structures « après » et « de ... en », dont la productivité vient ici confirmer leur plus grande littérarité.

### 3. Le partage générique ?

#### 3.1. Les sous-corpus

A : pourcentages de chaque préposition sur toutes les prépositions

B : pourcentages des structures sur le nombre d'expressions différentes



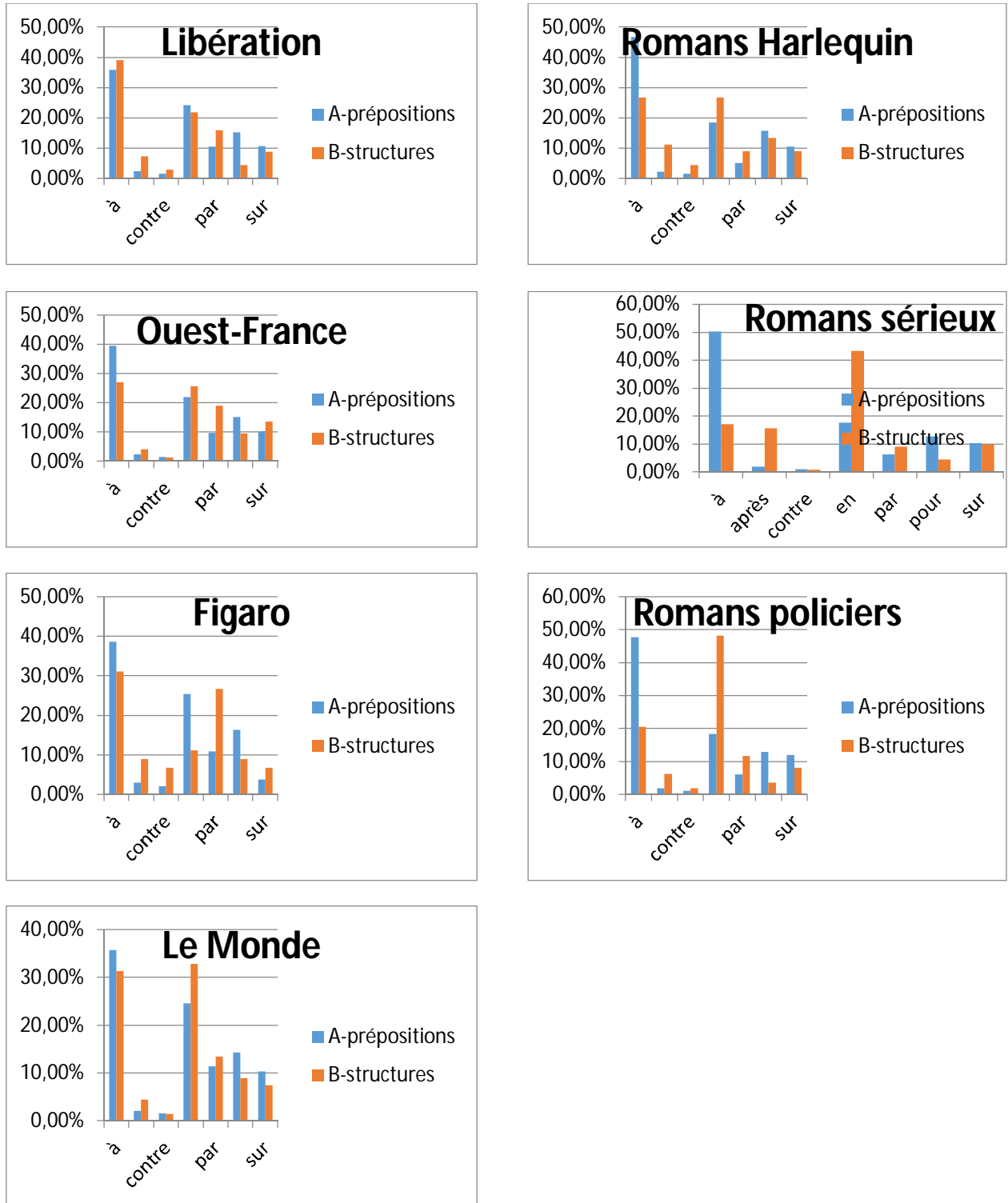


Figure 7- Histogrammes contrastifs par sous-genres : la place des prépositions et des structures

Ces histogrammes affectés à chaque sous-corpus permettent de visualiser à la fois la part de chaque préposition sur l'ensemble des prépositions et la part des structures composées avec telle préposition sur l'ensemble des structures différentes. Des tendances majoritaires se dessinent et des particularités se dégagent :

On observe, pour la préposition « à », la faiblesse du renouvellement des structures pour tous les sous-corpus sauf pour deux journaux, *L'Est Républicain* et *Libération*. Le déséquilibre est même accentué pour les textes fictionnels entre le pourcentage des prépositions et le

pourcentage des structures. Il semble bien que ce n'est pas dans les structures avec « à » qu'un quelconque indice de littérarité puisse être trouvé. Son affinité avec la collection *Harlequin* constitue un indice de figement stylistique propre à caractériser l'écriture mise en jeu dans ce genre romanesque.

En revanche, la préposition « après » propose la configuration inverse : une faible représentation parmi les prépositions et une variété plus grande dans les structures NpN, avec une légère accentuation dans les textes fictionnels. C'est la même tendance à la variété des structures qui peut être constatée pour la préposition « en », sauf pour *Libération* et *Le Figaro*. La tendance à la lexicalisation est esquissée pour la préposition « pour » sauf pour le *Sud-Ouest*. Enfin, « sur » ne présente une variété dans ses structures que pour *Ouest-France* et *Le Figaro*. La préposition « contre » ne paraît intéressante que pour deux sous-corpus : *Le Figaro* et *Harlequin*. On y trouve les occurrences « bloc contre bloc, gel contre gel, projet contre projet » pour le journal et des expressions plus sensuelles pour *Harlequin* « peau contre peau, poitrine contre poitrine ».

### 3.2. Les expressions les plus fréquentes

C'est quand on observe le détail des structures réalisées en contexte que des spécificités génériques s'affirment plus clairement. On constate en effet que deux expressions ont la faveur des sous-corpus médiatiques : « tour à tour » et « cas par cas », la première pour *L'Est Républicain*, *Libération*, *Sud-Ouest*, la seconde pour *Le Figaro*, *Le Monde*, *Ouest-France*. On retrouve « tout à tour » en tête de liste pour le *Roman policier* tandis que *Harlequin* préfère l'expression « tête à tête » et le *Roman sérieux*, de façon plus neutre, « de temps en temps ».

Il avait été constaté, lors de notre précédente étude, que 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.

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.

De manière symétrique, des substantifs identiques reliés par des prépositions différentes peuvent constituer une structure de même sens. Un même sémantisme est réalisé par plusieurs structures.

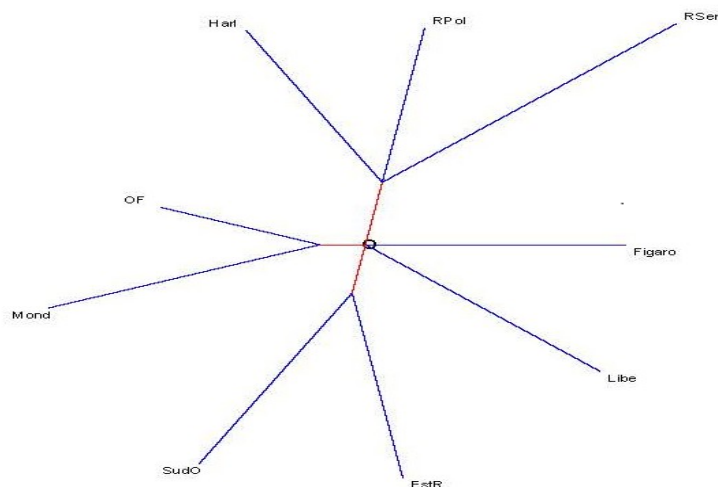


Figure 8- La distribution des structures les plus fréquentes

Cette analyse arborée permet d'obtenir une classification des sous-corpus selon l'emploi qu'ils font des 39 structures les plus fréquentes, celles dont la fréquence absolue dépasse 7. Ce graphique a l'avantage de permettre le regroupement des trois corpus de romans mais les traits bleus sont longs et la distance du nœud à la racine est courte, ce qui veut dire que les proximités des individus regroupés sont assez faibles.

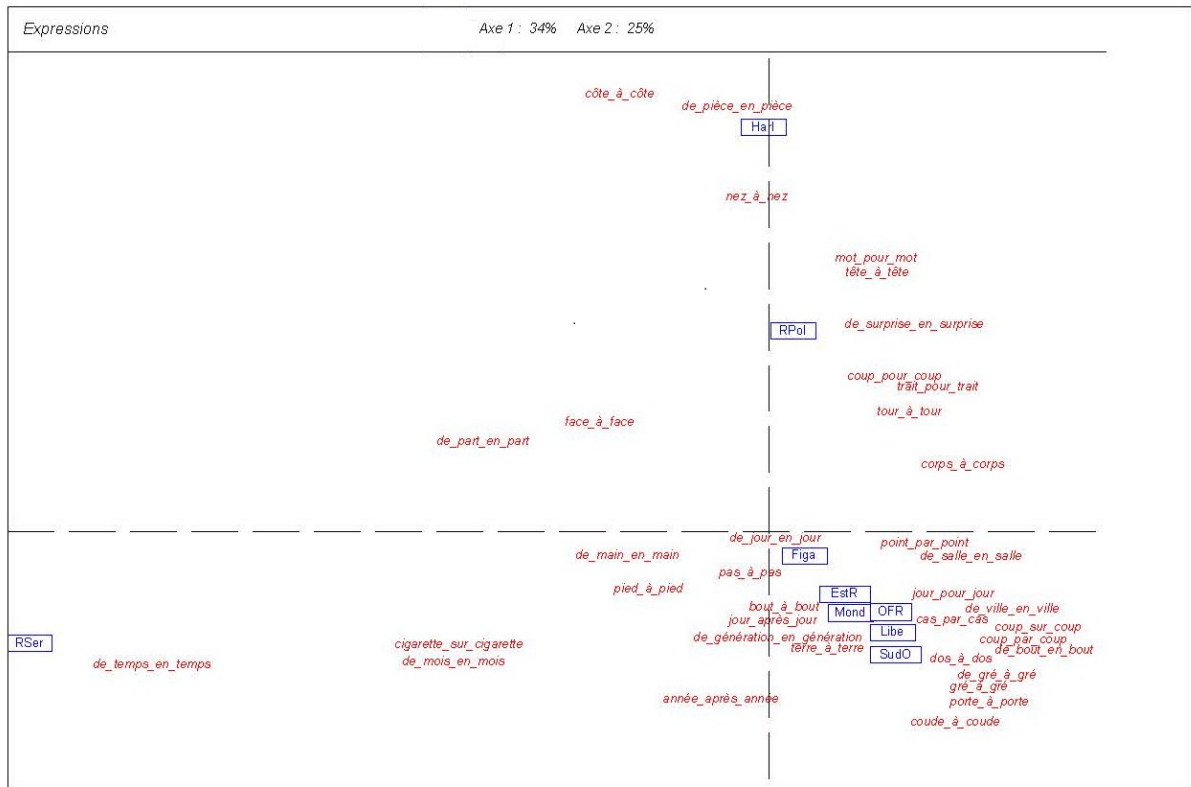


Figure 8- La distribution des structures les plus fréquentes explicitées

Cette AFC propose la visualisation des 39 structures les plus fréquentes et leur proximité avec les sous-corpus. La distribution est assez claire : le *Roman sérieux* se trouve isolé et proche d'expressions comme « de temps en temps » ou « cigarette sur cigarette », tandis que la collection *Harlequin* est projetée aux antipodes, dans le quadrant nord-est, proche de « de pièce en pièce ». Le *Roman policier* assure, quant à lui, la transition entre ce type de roman sentimental et l'ensemble médiatique, concentré dans le quadrant sud-est. Ces corpus de presse partagent donc bien des points communs, tandis que les romans divergent et se liguent contre la presse. Ce sont des distinctions qui seraient à affiner qui concernent le contenu lexical des expressions.

## Conclusion

Des conclusions peuvent être tirées de l'observation des résultats statistiques. Des tendances en langue se dégagent assez nettement sur l'emploi des prépositions dans les structures NpN.

Les observations portent d'abord sur la fréquence des prépositions dans ces structures : une hiérarchie se dessine qui donne la primauté à la préposition « à » et l'emploi le plus faible à la préposition « contre », en échelonnant entre ces deux pôles l'emploi de « en, pour, sur, par, après ». Lors de la précédente étude, c'est le même constat du privilège quantitatif attribué à la préposition « à » qui avait été constaté de même qu'exactement la même échelle ordonnée des prépositions qui avait été observée.

C'est aussi la productivité des structures qui peut être commentée : cette productivité est inverse de la fréquence de la préposition dans ces structures. Les résultats obtenus à partir de l'extrait de *Frantext* concordent avec les résultats de la présente étude et en confortent par conséquent les conclusions linguistiques. « Après » et « en » dans l'expression « de...en » sont deux prépositions qui présentent un degré de lexicalisation moindre appuyé par un éventail large de substantifs de part et d'autre de la préposition. En revanche, « contre » s'était révélée une préposition également productive dans le corpus littéraire extrait de *Frantext* ; cela n'est confirmé que par la collection *Harlequin* dans le corpus actuel. La tendance à la lexicalisation est manifeste pour les prépositions « pour » (avec l'exception du *Sud-Ouest*), « sur » (sauf *Ouest-France* et *Le Figaro*), « contre » (sauf *Le Figaro* et *Harlequin*).

Les tendances génériques qu'on aurait aimé voir se dessiner avec plus de netteté ne sont pas clairement marquées sur le plan linguistique. On constate en particulier un sur-emploi dans le corpus médiatique de « de...à » (surtout dans *L'Est Républicain*, *Libération* et *Le Monde*), « par », « sur » et dans l'ensemble fiction de « de...en », surtout dans le Roman dit sérieux.

C'est dans le choix des substantifs de part et d'autre de la préposition que se marquent davantage les particularités génériques, comme l'a montré le dernier graphique avec une distribution claire dans les quadrants de l'AFC. Les particularités génériques sont davantage à chercher du côté du lexical et du sémantique que du côté syntaxique.

Cette étude qui avait pour enjeu de faire émerger le statut de marqueur générique pour les structures NpN n'a pas clairement démontré le rôle de ces dernières dans la différenciation de genres aussi éloignés *a priori* que le corpus médiatique et le corpus littéraire. En revanche, elle a fait apparaître une opposition forte entre les sous-genres littéraires, entre le *Roman sérieux* et la collection *Harlequin* qui pourrait être affinée par l'examen précis des différentes structures et en particulier des hapax, réputés créations et non pas figements.

## Références

- Jackendoff R. (2008). « Construction after construction and its theoretical challenges ». *Language*, vol. 84 (1) : 8-28
- Legallois D, Charnois Th. & Poibeau Th. (à par.). « Sans un mot, Nicole se lova contre Franck et ne bougea plus : la méthode des « motifs » pour repérer les clichés dans les romans sentimentaux. ». *Lidil*
- Longrée D & Mellet S. (2013). « Le motif : une unité phraséologique englobante ? Etendre le champ de la phraséologie de la langue au discours ». *Langages* (189) : 68-80.
- Longrée D. Luong X et Mellet (2008). « Les motifs : un outil pour la caractérisation topologique des textes », *JADT*, Lyon : 733-744.
- Magri V. & Purnelle G. (2012). « Mot à mot, brin par brin : les suites [Nom préposition Nom] comme motifs », *JADT*, Liège.
- Quiniou S., Cellier P., Charnois Th. & Legallois D. (2012). What About Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics ? » in *Lecture Notes in Computer Science*, Springer Vol. 7181 : 166-177.

## Annexe - Le corpus

Sous-corpus	Nombre de formes	Sous-corpus	Nombre de formes
<i>Le Figaro</i>	1552842	<i>L'Est Républicain</i>	1743122
<i>Libération</i>	1695772	<i>Roman Policier</i>	3449740
<i>Le Monde</i>	1620852	<i>Roman Sérieux</i>	3125014
<i>Ouest-France</i>	1581703	<i>Harlequin</i>	3523252
<i>Sud-Ouest</i>	1705856		

LES SUITES [NOM PREPOSITION NOM] COMME INDICES DE LITTERARITE ?