

## A la recherche de *motifs* dans le corpus de romans policiers

Mots clés: textométrie, genre, roman policier, motif, motif sémantique

### Objectifs

Dans cette contribution, nous souhaitons participer au débat qui se déroule autour de la notion de motif. L'objectif de notre proposition est de savoir comment repérer automatiquement ou semi-automatiquement « les motifs » dans le corpus constitué de romans policiers.

La surdéfinition du terme « motif » lors de ces dernières années ne peut entraîner que des malentendus. Cette notion est employée aussi bien dans des analyses littéraires que dans le domaine de la description linguistique surtout dans les travaux exploitant les méthodes de la phraséologie dite étendue qui recourt régulièrement aux outils informatiques.

Les analyses statistiques de données textuelles nous ont permis de dégager les unités linguistiques du roman policier. Nous nous demandons dans quelle mesure les résultats obtenus à l'aide de la méthode textométrique peuvent donner accès aux séquences plus larges caractéristiques du roman policier et faciliter la caractérisation linguistique du genre en question. Dans les études précédentes, nous avons constaté que la méthode statistique ne donne pas toujours accès aux concepts-clés de motifs, par exemple : le lexème <la craie.N> n'est noté que 67 fois, le lexème <le tracé.N> est retenu 81 fois, contrairement au lexème <le corps.N> observé 2559 fois. Le fait de mettre ensemble ces trois lexèmes donne sans aucun doute accès à la scène de crime: *le tracé à la craie du corps de qqn*. Cependant, si les éléments constitutifs de cette structure lexico-syntaxique sont figés, seulement le premier peut permuter avec *le contour*, *la trace* ou *le dessin*. Par conséquent, l'effacement de l'élément pivot : *la craie* ne donne pas accès à la scène de crime: *le dessin du corps*, alors une unité de faible fréquence (*la craie*) devient significative dans l'analyse du roman policier. Prenons un autre exemple : le verbe prédicatif <comparer.V> n'est retenu que 99 fois, mais le concept de *comparaison* devient prototypique au motif Enquête. La réalisation de ce modèle sémantique est la suivante: CAUSE {[comparer (P(x)), (P(y))], [comparaison (P(x), (P(y)))]}, **Picquart compare l'écriture d'Esterhazy avec celle du bordereau qui a fait condamner le capitaine Dreyfus. Elle est identique.** N<HUM : Inspecteur> comparer <N1 : écriture > avec <N2 : écriture >, N1 et N2 être identique, N1=N2.

La question est de savoir comment on pourra dégager des unités pivots inscrites dans les structures lexico-syntaxiques représentatives aux motifs sous-jacents.

### Corpus de travail

Cette étude se base sur un vaste corpus de textes littéraires (325160 sentence delimiters, 11950270 tokens) à savoir des romans policiers écrits ou traduits en français venant d'auteurs reconnus, aussi bien français qu'étrangers : Gaston Leroux, Maurice Leblanc, Conan Doyle, Georges Simenon, Fred Vargas, Agatha Christie, Jean-Christophe Grangé, Arnaldur Indridason, Maxime Chattam, Dorothy L. Sayers, Harlam Coben, Jean-Claude Izzo, Pierre Magnan, Boileau-Narcejac, Exbrayat, et d'autres. Même si la tradition de la linguistique de corpus préfère recourir à des corpus comparables plutôt qu'à des corpus de traductions, jugés peu fiables, d'après nous, il existe l'invariance de certains schémas lexico-syntaxiques appropriés qui discriminent un type de texte, par exemple un roman policier, tel un code ADN. Le traducteur ne peut faire l'impasse sur certains éléments, car ils représentent les marques obligatoires du genre. L'évolution du genre peut modifier certaines caractéristiques, mais le schéma qui permet de reconnaître un roman policier comme tel doit rester stable (ce

type de texte est régi par des principes organisateurs communs à l'ensemble des textes policiers). Les études textométriques comparant deux corpus, un corpus de romans policiers écrits en français et un corpus de romans traduits depuis l'anglais ont prouvé que "la langue traduite ne présente pas de spécificités suffisamment saillantes pour considérer qu'il s'agirait d'un « troisième code » qu'il faudrait bannir de notre étude des sous-genres littéraires" (Kraif, à paraître).

### **La notion de motif**

Du point de vue littéraire, le motif est « une toile de fond, un concept large, désignant soit une certaine attitude — par exemple la révolte —, soit une situation de base impersonnelle dont les acteurs n'ont pas encore été individualisés » (Raymond Trousson).

En linguistique la notion de motif est aussi ambiguë. Aujourd'hui, elle est étroitement liée à la phraséologie dit étendue : « motifs » (Longrée & Mellet 2013), « motifs séquentiels » (Quiniou *et al.* 2012), « routines discursives » (Née, Sitri, Veniard 2014), « segments répétés » (Salem 1986), « formulations stéréotypées » (Grossmann 2014), « motif sémantique » (Muryn, Niziołek 2016) et d'autres.

En linguistique de corpus, les motifs lexicaux complexes sont construits à partir des propriétés des mots (eg. : "un mot de lemme 'aimer' suivi à au plus de 2 mots d'un mot commençant par 'pouv'") (TXM, <http://textometrie.ens-lyon.fr/spip.php?rubrique96&lang=fr>).

Dans notre approche le motif est conçu comme combinaison de concepts clés et le genre littéraire est envisagé comme combinaison de motifs organisés hiérarchiquement. La question qui se pose est de savoir avec quels moyens linguistiques et informatiques, on peut dégager les réalisations de ces motifs. Dans nos analyses, nous ferons recours aux outils du traitement automatique comme Nooj, ou TXM.

## Bibliographie

- Hajok, A. (à paraître) : « Le prédicat *comparer* et le concept de comparaison » in Uniwersytet Pedagogiczny, Moskwa.
- Hajok, A. (à paraître) : « Le prédicat *comparer* et le concept de comparaison » in Uniwersytet Pedagogiczny, Moskwa.
- Grossmann, F. (2014). Les verbes de constat dans l'écrit scientifique. In Tutin, A. et Grossmann, F., L'écrit scientifique, Du lexique au discours, Rennes, Presses Universitaires de Rennes, pp.85-100.
- Hajok, A., Muryn, T., Niziołek, M., Gabrysiak, K., (2017). « La structure linéaire de la matrice lexico-syntaxique du roman policier », In : Studii de lingvistică. , Vol. 7, s. 169-183.
- Kraif, O. (2017): „Traduire le polar : une étude textométrique comparée de la phraséologie du roman policier en français source et cible”. In: *Synergies Pologne*, (à paraître)
- Legallois D., Tutin A. (2013). « Vers une extension du domaine de la phraséologie ». *Langages* 189.
- Legallois, D. (2006). « Des phrases entre elles à l'unité réticulaire du texte ». *Langages*, 40e année, n° 163. *Unité(s) du texte*, Dominique Legallois (éds). p. 56-70.
- Longrée D., Mellet S. (2013). « Le motif : une unité englobante ? Étendre le champ de la phraséologie de la langue au discours ». *Langages* 189, p. 65-80.
- Muryn, T., Niziołek, M., (2017),. « L'intertext dans le roman policier », In: *Synergies Pologne* nr 14, 2017 (à paraître).
- Muryn, T., Niziołek, M. (2016) : «Pour une analyse phraséologique du roman policier», In : Mogorron Huetra, Pedro, Cuadrado Rey, Analia, Martinez Blasco, Iván, Navarro Brotons, Lucia, (éds), *Fraseologia, variacion y traducción*, Peter Lang <https://www.peterlang.com/view/9783631694879/xhtml/fm001.xhtml>
- Muryn T., Niziołek M., Hajok A., Prażuch W., Gabrysiak K. (2016). « La matrice lexico-syntaxique du roman policier ». 5<sup>ième</sup> Congrès Mondial de Linguistique Française, le 4-8 juillet 2016, Institut de Linguistique Française, Université de Rabelais de Tours, France. <http://dx.doi.org/10.1051/shsconf/20162706007>
- Muryn T., Niziołek M., Prażuch W., Hajok A., (2015) "*Ce café me réveille* ou l'intensité inférée ». In *L'intensification et ses différents aspects* (Krystyna Wróblewska-Pawlak & Anna Kieliszcyk éds.), Université de Varsovie, p. 85-100.
- Quiniou S., Cellier P., Charnois T., Legallois D. (2012). « Fouille de données pour la stylistique : cas des motifs séquentiels émergents », in Proceedings of the 11th International Conference on the Statistical Analysis of Textual Data, Liege, p. 821-833.
- Rastier F. (1996). « Pour une sémantique des textes. Questions d'épistémologie». [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_PourSdT.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_PourSdT.html)
- Salem, A., (1986) « Segments répétés et analyse statistique des données textuelles », *Histoire & Mesure*, p. 5-28
- Tutin A., Grossmann, F. (2002). « Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif », *Revue française de linguistique appliquée* 2002/1 (Vol. VII), p. 7-25.
- Née, E., Sitri, F., Veniard, M. (2014). « Pour une approche des routines discursives dans les écrits professionnels ». CMLF, Berlin, Allemagne.