

## Introduction. Motifs, textométrie et stylistique des textes littéraires

Dominique Legallois, Dominique Longrée, Iva Novakova and Antoine Silvestre de Sacy

---

**Electronic version**

URL: <https://journals.openedition.org/corpus/10224>

DOI: 10.4000/13woh

ISSN: 1765-3126

**Publisher**

Bases ; corpus et langage - UMR 6039

**Electronic reference**

Dominique Legallois, Dominique Longrée, Iva Novakova and Antoine Silvestre de Sacy, "Introduction. Motifs, textométrie et stylistique des textes littéraires", *Corpus* [Online], 27 | 2025, Online since 02 June 2025, connection on 05 June 2025. URL: <http://journals.openedition.org/corpus/10224> ; DOI: <https://doi.org/10.4000/13woh>

---

This text was automatically generated on June 5, 2025.

The text and other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.

---

# Introduction. Motifs, textométrie et stylistique des textes littéraires

Dominique Legallois, Dominique Longrée, Iva Novakova and Antoine Silvestre de Sacy

---

- 1 Ce numéro de la revue *Corpus* rassemble des analyses mettant en relation la notion de *motif* (et ses différentes conceptions) avec des études stylistiques, stylométriques ou textométriques du texte littéraire.

## Brève histoire du motif

- 2 Bien que la notion de *ngrams* soit antérieure à ses travaux, il revient à A. Salem (1987) d'avoir proposé des méthodes et des outils pour l'analyse statistique d'unités plus grandes que le mot simple ou le lemme simple : des suites de mots ou lemmes se répétant, soit des segments répétés, qui étaient alors identifiés le plus souvent dans le discours syndicaliste et politique. Ces travaux s'inscrivaient en Analyse du Discours sans visée stylistique, mais dans la perspective d'identifier des formes récurrentes possédant notamment une dimension intertextuelle. Par la suite, cette notion de segment a été appliquée sur les annotations en partie du discours (POS) : ce fut en particulier l'objet du travail de Ganascia (2001), qui extrayait automatiquement des séquences récurrentes d'annotations en parties du discours, à partir d'arbres syntaxiques. La méthode fut appliquée à la discrimination des caractéristiques stylistiques de Madame de Lafayette, comparé à des auteurs du 19<sup>e</sup> (Guy de Maupassant, Georges Sand et Marcel Schwob). Quelques patrons saillants de *La princesse de Clèves* et de *La comtesse de Tende* étaient alors identifiés :
- 3 Préposition+Pronom+Infinitif : à le servir ; de le supplier ; de l'éviter ; de l'aimer
- 4 Préposition+en+Infinitif : sans en avoir ; d'en avoir ; d'en attendre ; d'en garantir ; d'en faire
- 5 Adverbe+bien+fait : admirablement bien faits ; parfaitement bien faits ; très bien fait
- 6 GN<sub>pr</sub>+Verbe+DET+N : Le prince de Navarre prit la parole ; La reine de Navarre avait ses favorites ; Monsieur de Nemours prit la reine dauphine ; La comtesse reçut ce billet avec joie.

- 7 Ces patrons sont appelés *motifs* par Ganascia. Le terme est employé dans son acception informatique, plus particulièrement en traitement de données ou en recherche d'informations : un motif peut désigner une séquence ou un ensemble de caractéristiques qui sont recherchées dans un ensemble de données. La recherche est ici non supervisée et est conduite par une méthode inductive (*corpus-driven*) : le motif émerge de l'analyse statistique (ou simplement de l'analyse quantitative) par comparaison avec un corpus de référence. Les quelques exemples donnés ci-dessus illustrent parfaitement certains caractéristiques ou problèmes liés au motif : ainsi, si *Adverbe+bien+fait* est une séquence relativement homogène, car en partie lexicalement déterminée, se pose la question de l'interprétabilité de motifs comme *Préposition+en+Infinitif* (quelles régularités doit-on identifier dans ce segment ?) ou  $GN_{pr}+Verbe+DET+N$  (les réalisations de ce schéma de phrase ont-elles toutes la même fonction, et si oui, laquelle ?). Par ailleurs, on peut légitimement se demander si les motifs identifiés sont propres à l'auteur, ou bien s'ils sont caractéristiques du discours littéraire de l'époque. Cet ensemble de questions assurément persiste dans les analyses actuelles fondées sur la découverte des motifs.

## Le motif, plus récemment

- 8 Dans la première décennie des années 2000, Longrée et Mellet, travaillant sur les textes latins, cherchent à identifier des configurations présentant de multiples variations lexicales, morphologiques et syntaxiques, au point qu'elles peuvent ne plus avoir de points communs en surface ; ces configurations ou motifs se définissent alors comme des micro-structures récurrentes de  $n$  éléments du texte, munis de leur structure linéaire ; au-delà des variations, l'unité du motif est assurée par sa fonction d'organisateur textuel (mais pas seulement). Ils peuvent caractériser un texte, ou des parties de textes lorsque l'analyse est topologique (Longrée & Mellet, 2013, 2018). En cela, ces travaux voulaient poursuivre par des moyens automatisés le travail de Chausserie-Laprée (1969) sur l'expression narrative chez les historiens latins (l'ouvrage était sous-titré *histoire d'un style*).
- 9 Indépendamment de ces recherches, et dans un certain nombre de collaborations développées depuis 2012 (notamment Cellier et al. 2012), Legallois propose la conception suivante du motif : un motif est un patron lexico-grammatical séquentiel identifié de manière non supervisée par sa surreprésentation statistique dans un texte ou groupe de textes. Ce patron doit être interprétable : il remplit une fonction sémantique et /ou stylistique. C'est une forme généralement plus schématique que le segment répété portant sur des formes de mots ou des lemmes, mais plus spécifique que des *ngrams* de POS, c'est-à-dire des suites d'étiquettes morphosyntaxiques, souvent trop abstraites pour être interprétées (Legallois, Charnois et Poibeu (2016), Legallois, Charnois et Larjavarra (2018) pour des explications plus développées). Le motif est dit, dans cette perspective, séquentiel car il est constitué d'une suite « linéaire » d'éléments.
- 10 Le terme *motif* est également au cœur des travaux menés au laboratoire Lidilem. Ces travaux visent à identifier des patterns de pivots lexicaux spécifiques à des (sous-)genres littéraires. Pour ce faire est utilisée une méthode basée sur des corpus d'arbres de dépendance syntaxique afin d'extraire des données donnant accès à la combinatoire lexico-syntaxique des unités lexicales en s'appuyant sur la notion de

cooccurrence syntaxique (Kraif 2019). L'extraction de ces arbres prend comme point de départ un pivot lexical (verbal ou nominal), sélectionné par des critères de fréquence et de dispersion et produit un ensemble de sous-arbres récurrents (Kraif, 2016). La méthode a été implémentée dans l'outil en ligne, le Lexicoscope<sup>1</sup>.

- 11 Enfin, en parallèle, le laboratoire BCL de l'Université de Nice s'appuie sur des outils d'aide à l'interprétation d'algorithmes de *deep learning* pour explorer la boîte noire des informations cachées dans les couches intermédiaires de l'apprentissage profond de réseaux convolutionnels (Vanni, Mayaffre, Longrée, 2018). Avec cette technique, Hyperdeep, une extension du logiciel de logométrie Hyperbase<sup>2</sup>, est à même non seulement de prédire l'appartenance d'un texte à un style ou à un auteur, mais aussi d'identifier les caractéristiques linguistiques, apprises par le système, et sur lesquelles reposent l'analyse. Certaines de ces caractéristiques sont là encore appelées *motifs*.
- 12 Au regard de ces multiples approches et définitions, à la fois différentes et complémentaires, on pourrait se demander s'il y a... *un motif* émergeant dans ces différentes définitions ou conceptions du motif.

## Éléments constitutifs du motif

- 13 On peut d'abord constater que les approches du motif identifient un invariant lexical (un mot pivot) ou formel (un patron grammatical) ou encore fonctionnel (un ensemble de formes dédiées à une même fonction) ; « autour » de cet invariant gravitent des éléments plus hétérogènes, acteurs de la variation du motif, de ses métamorphoses. On constate également que le motif est une unité discriminante ; c'est un élément distinctif qui permet de différencier un texte d'autres textes, et qui, en cela participe d'un style. Ce pouvoir discriminant ne va pas sans une certaine récurrence : un hapax ne pourrait constituer un motif. Par ailleurs, la phraséologie a bien souvent offert un cadre accueillant aux approches du motif pour réfléchir à la nature de cette unité : cœur invariant lié au figement, mais aussi variabilité qui ne gomme pas malgré tout l'identification d'une même récurrence. Ainsi, l'approche du motif entreprise par le laboratoire Lidilem s'inscrit particulièrement en phraséologie du texte littéraire (à la suite notamment au projet PhraseoRom<sup>3</sup>, cf. Novakova & Siepmann, 2020). Cependant, la phraséologie peut également constituer un lieu trop étroit par son indifférence (relative) à la textualité (et par sa subordination à la lexicologie) ; c'est pourquoi la stylistique constitue également un domaine privilégié du motif à condition que soit dépassée une certaine conception du style fondée principalement sur les tropes ou figures : le motif, bien souvent, est assez peu esthétiquement remarquable. Par ailleurs, cette stylistique, peut compter sur l'analyse des motifs pour prendre comme objet des études comparatives : comparaison entre langues (cf. ici même l'article de Kiatphotha), comparaison entre genres discursifs (cf. ici même l'article de Magri, Mayaffre et Vanni), comparaison entre genres et langues (cf. l'article de Gymnich et Novakova), ou encore comparaison entre écrit et oral (article ici même de Wang et Tutin).
- 14 Les articles de ce numéro de Corpus consistent en des réflexions et des analyses de cette unité encore émergente, de motif. Ils recouvrent deux sections ; une première, articulée sur les périmètres des motifs et les méthodes de détection. Une seconde, consistant en études de cas.

## Périmètres des motifs et méthodes de détection

- 15 Olivier Kraif, Iva Novakova et Julie Sorba développent dans leur article une réflexion sur le motif phraséologique comme marqueur générique et explorent la possibilité d'identifier des motifs propres à des auteurs ou à des textes spécifiques ; à cette fin, trois ensembles de données sont analysés : d'abord les constructions lexico-syntaxiques de deux sous-genres romanesques (la fantaisie et la littérature générale), puis les romans d'Anne Robillard comparés au sous-corpus de fantaisie, et enfin les romans de Romain Gary (Émile Ajar inclus) comparés entre eux et avec le sous-corpus de littérature générale. Il s'agit pour les auteurs de proposer une perspective holistique à travers la notion de *motifs phraséologiques* sur le phénomène de style, allant du style d'un genre à celui d'un auteur, puis à celui du texte.
- 16 L'article, écrit par Dominique Longrée et Laurent Vanni, propose de comparer à travers l'exemple du *De Bello Gallico* de César, les capacités du logiciel Hyperbase Web Edition (Vanni et al. 2020) en matière de détection de motifs avec celles d'autres logiciels accessibles. L'analyse se penche également sur différentes méthodes de fouille de données et d'analyses statistiques, tout en considérant l'impact potentiel de l'intelligence artificielle dans ce domaine. Selon les auteurs, l'avenir de la recherche sur les motifs textuels repose sur l'intégration de l'intelligence artificielle et des techniques de *deep learning*, qui offrent des possibilités prometteuses pour explorer les relations entre motifs textuels et intertextualité
- 17 Véronique Magri, Damon Mayaffre et Laurent Vanni utilisent un corpus contrastif composé de textes narratifs à la fois factuels (récits de voyage) et fictifs, pour évaluer des approches complémentaires d'identification de formes : l'analyse quantitative, et les modèles convolutionnels et transformers. Ainsi sont identifiés des indicateurs du registre didactique dans le genre récits de voyage (par exemple, *c'est-à-dire*), mais aussi des structures plus abstraites : grâce aux techniques de convolution et d'attention automatique liées au *deep learning*, le réseau de neurones met en évidence les différentes variations lexicales de la colligation morpho-syntaxique ADV+ADJ, ainsi que ADV+ADJ pluriel, sous des formes telles que *plus libres, bien rondes, très grands, fort belles* ou *plus petites*.
- 18 La note rédigée par Antoine Silvestre de Sacy et Dominique Legallois a pour objectif de présenter à la fois la méthode des motifs séquentiels développée depuis une dizaine d'années, et les outils récents (MotiveR et Pymotifs) qui permettent de réaliser des extractions de motifs dans les textes. Quelques résultats à partir de l'analyse des nouvelles des *Soirées de Médan* y sont exposés.

## Études de cas

- 19 Par le biais d'une approche stylistique qualitative, enrichie d'une modélisation computationnelle des textes par la méthode des motifs séquentiels, l'étude d'Antoine Silvestre de Sacy explore, les phénomènes de segmentation et d'hypersegmentation dans les textes de Louis-Ferdinand Céline. L'importance d'appréhender la phrase non seulement comme une entité délimitée par la ponctuation, mais également à travers ses éléments satellites, est mise en lumière. Dans ce contexte, les transformations syntaxiques résultant de l'hypersegmentation sont variées et reposent sur des procédés

comme les dislocations, ruptures, ellipses et hyperbates. Elles créent des effets expressifs variés : effets d'insistance, de mise en lumière, de descriptions impressionnistes, ainsi que des sensations d'attente, de surprise, d'humour et d'oralité.

- 20 L'article de Marion Gymnich et d'Iva Novakova présente une méthode de comparaison phraséologique visant à identifier et caractériser deux sous-genres littéraires : les romans de *fantasy* et les romans historiques, en français et en anglais. La méthode s'appuie sur l'extraction automatique de constructions lexico-syntaxiques récurrentes à partir de grands corpus de romans contemporains publiés depuis 1950. L'analyse se concentre sur deux motifs phraséologiques (<draw sword> / <dégainer l'épée> dans les romans de *fantasy*, et <raised his voice> / <dire d'une voix> dans les romans historiques), qui se révèlent être des indicateurs génériques fiables et mettent en lumière des différences stylistiques entre les romans français et anglo-saxons appartenant à ces deux sous-genres.
- 21 L'article de Jirawan Kiatphotha examine les motifs phraséologiques liés à l'expression adresser un sourire ainsi qu'à son équivalent thaï *sòj jím* dans la littérature sentimentale contemporaine française et thaïlandaise. L'étude met en évidence d'importantes variations structurelles : le français privilégie des constructions élaborées avec une caractérisation adjectivale riche, tandis que le thaï adopte une approche plus succincte, souvent intégrée dans des structures à verbes sériels. Sur le plan discursif, l'expression remplit une fonction infra-narrative commune aux deux langues. Cependant, elle est davantage liée aux scènes émotionnelles en français, alors qu'en thaï, elle joue principalement un rôle dans le développement narratif. Ces différences illustrent des disparités culturelles dans l'expression de la communication non verbale.
- 22 Enfin, l'article rédigé par Yinjie Wang et Agnès Tutin s'intéresse aux formules exprimant la surprise dans les dialogues des romans contemporains français, telles que *Tu plaisantes ?*. L'analyse des occurrences extraites du corpus CRFC met en lumière des régularités sur les plans syntaxique, lexical et sémantique. Par ailleurs, l'étude du corpus PhraseoRom (à partir des années 1980) révèle que l'emploi de ces expressions est étroitement corrélé au genre littéraire. En comparaison avec des interactions orales authentiques (ESLO2 et CEFC-ORFEO), il apparaît que certaines formules orales représentées dans les romans, comme *Tu rigoles ?* et *Ma parole*, sont massivement présentes

## Conclusion

- 23 Ainsi les articles du présent numéro s'inscrivent dans trois « traditions » différentes du motif : celle des motifs dits « lexicaux » du laboratoire Lidilem de Grenoble, celle des motifs séquentiels initiée par Legallois, et celle des motifs détectés par les techniques d'intelligence artificielle du laboratoire BCL de Nice. Ces trois perspectives, insistons sur ce point, travaillent depuis longtemps en étroite collaboration, en essayant chacune d'intégrer au mieux les apports des autres. Mais dans une certaine mesure, on peut déclarer cette collaboration encore incomplète : les auteurs des travaux publiés ici comptent également sur les expertises des lecteurs de ce numéro, notamment celles des stylisticiens et des spécialistes du texte littéraire. Ceux-ci peuvent prendre connaissance d'un ensemble d'outils présentés ici (le Lexicoscope, MotiveR et PyMotif, Hyperdeep) et apporter à leurs auteurs des éléments pouvant enrichir à la fois la

description de cette unité qu'est le motif, et la réflexion sur une stylistique outillée et fondée sur ce type d'unité.

---

## BIBLIOGRAPHY

Chausserie-Laprée (1969), *L'expression narrative chez les historiens latins. Histoire d'un style*. Paris, De Boccard.

Cellier, P., Quiniou, S., 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.

Ganascia, J. G., (2001), « Extraction automatique de motifs syntaxiques » In *Actes de la 8ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN'2001)*. Tours (France), edited by Jean Véronis, Laurence Danlos, Pierre Zweigenbaum, Nathalie Gasiglia, and Pascal Amsili. Accessed January 28, 2019. <http://talnarchives.atala.org/TALN/TALN-2001/taln-2001-long-017.pdf>.

Kraif, O. (2016), "Le lexicoscope: un outil d'extraction des séquences phraséologiques basé sur des corpus arborés." *Cahiers de Lexicologie*, 108, 91–106

Kraif O. (2019), « Explorer la combinatoire lexico-syntaxique des mots et expressions avec le Lexicoscope », *Langue française*, 203, 67-82.

Legallois, D., Charnois, Th., Poibeau, Th., Thierry, (2016), « Repérer les clichés dans les romans sentimentaux grâce à la méthode des 'motifs' », *Lidil*, 53, 95-117.

Legallois, D., Charnois, Th., Larjavaara, M., (2018), The Grammar of Genres and Styles, "The balance between quantitative and qualitative literary stylistics : how the method of 'motifs' can help", Berlin/Boston, de Gruyter, 164-194.

Longrée, D. et Mellet S., (2013). « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours », *Langages*, 189, 68-80.

Longrée, D., et S. Mellet., (2018), "Towards a topological grammar of genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach". In *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, edited by D. Legallois, Th. Charnois, and M. Larjavaara, Berlin/Boston: de Gruyter, 140–163.

Novakova, I., & Siepman, D. (Eds.), (2020). *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Palgrave Macmillan

Salem A. (1987), *Pratique des segments répétés. Essai de statistique textuelle*, Paris, Klincksieck.

Vanni L., Mayaffre D., Longrée D., (2018), « ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables ». In : *14èmes Journées internationales d'Analyse statistique des Données Textuelles. JADT 2018*, Rome, 459-466.

Vanni L., Corneli C., Longrée D., Mayaffre D., Precioso F., (2020), « « Hyperdeep : deep learning descriptif pour l'analyse de données textuelles ». In : *15èmes Journées Internationales d'Analyse statistique des Données Textuelles. JADT 2020*, Toulouse, 230-243.

## NOTES

1. [http://phraseotext.univ-grenoble-alpes.fr/lexicoscope\\_2.0/](http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/)
  2. <https://hyperbase.unice.fr/>
  3. <https://lidilem.univ-grenoble-alpes.fr/node/16/axes-recherche/axe-1-descriptions-linguistiques-tal-corpus/projets-axe-1/phraseorom-phraseologie-roman-contemporain-francais-anglais-et-allemand>
- 

## AUTHORS

### **DOMINIQUE LEGALLOIS**

Université Sorbonne Nouvelle - Lattice

### **DOMINIQUE LONGRÉE**

Université de Liège - Lasla

### **IVA NOVAKOVA**

Université Grenoble Alpes - Lidilem

### **ANTOINE SILVESTRE DE SACY**

Université Sorbonne Nouvelle - Thalim & Lattice