

Les motifs phraséologiques pour distinguer les genres littéraires. Sur l'exemple des motifs de la communication verbale et non verbale

Iva Novakova

Department of Language Sciences and French as a Foreign Language
Research team LIDILEM: Linguistics and first and second language education
University Grenoble Alpes
CS 40700 – 38058 Grenoble cedex 9
France
E-Mail : iva.novakova@univ-grenoble-alpes.fr

Phraseological motifs for Distinguishing Between Literary Genres. A Case Study on the Motifs of Verbal and Non-Verbal Communication

Abstract. The present paper is based on the assumption that the language of the novel is characterized by a statistically relevant overrepresentation of certain linguistic units (e.g. lexemes, key words, collocations and colligations, Siepmann 2015). First steps towards checking the validity of this hypothesis had been undertaken in pioneering works in the 1990s/2000s (e.g. Stubbs & Barth 2003). These studies were however limited by the small size of their (exclusively English) corpora. The present study explores the role of some patterns (*phraseological motifs*) in distinguishing French literary subgenres. It also proposes a case study of some *motifs* related to the verbal (*dire avec sourire* ‘to say with a smile’) and non-verbal communication (*adresser un sourire* ‘to send a smile’). Unlike traditional corpus-stylistic analyses, which frequently focus on the style of a single author, our *corpus-driven* approach identifies lexico-syntactic constructions in literary genres which are automatically extracted from the corpora.

The main purpose is to show the relevance of the notion of *phraseological motif* (Legallois 2012; Longrée & Mellet 2013; Novakova & Siepmann 2020) for the distinction of literary subgenres. Linking form and meaning, these ‘multidimensional units’ fulfil pragmatic as well as discursive functions.

Submitted: 15/12/2020. Accepted: 10/04/2021

Copyright © 2021 Iva Novakova. Published by Vilnius University Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The data has been extracted from large French corpora of the PhraseoRom research project <https://phraseorom.univ-grenoble-alpes.fr>. They are accessible on <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/index.html> and contain 1000 novels (published from the 1950s to the present), partitioned into six sub-corpora: general literature (GEN), crime fiction (CRIM), romances (ROM), historical novels (HIST), science fiction (SF) and fantasy (FY).

The results of our study reveal some unexpected differences between the literary subgenres: e.g. the motif *dire d'une voix* 'to say in a voice' in HIST compared to GEN. In FY, expressions of verbal communication are related to shouting and screaming. Expressions related to the non-verbal communication (*prendre dans ses bras* 'to take in one's arms') are specific to ROM, where body language is overrepresented. In SF, there is a very limited number of these types of expressions. More generally, the motifs provide the link between the micro level (phraseological recurrences) and the macro level (the fictional script).

Key words: phraseological motifs, extended phraseology, digital humanities, literary genres

1 Introduction

Le principal objectif de cette étude¹ est de montrer la pertinence de la notion de *motif phraséologique* (Legallois 2012 ; Longrée & Mellet 2013 ; Novakova & Siepmann 2020) pour la distinction des genres littéraires. Les motifs en phraséologie sont conçus comme des unités polylexicales présentant des régularités et des variations lexicales et syntaxiques (au niveau paradigmatique et syntagmatique), établies avec des méthodes lexicométriques et associées à des fonctions discursives spécifiques. La principale question sous-jacente à cette recherche est de savoir *si* et *comment* la phraséologie étendue (Legallois & Tutin 2013) permet de distinguer les sous-genres de la littérature contemporaine. L'enjeu sera de les identifier et de les décrire non pas en termes d'univers de référence ou de style individuel des auteurs, comme c'est souvent le cas en stylistique classique, mais à travers des constructions lexico-syntaxiques (CLS) récurrentes, statistiquement spécifiques à tel ou tel sous-genre².

Le point de départ de cette étude repose sur l'hypothèse selon laquelle la langue romanesque, aussi bien dans son ensemble que dans ses genres spécifiques, se caractérise par la surreprésentation statistiquement significative de certains phénomènes linguistiques

¹ Ce travail a été effectué dans le cadre du projet franco-allemand PhraseoRom (2016–2019) <https://phraseorom.univ-grenoble-alpes.fr/>, financé par l'ANR et la DFG.

² Les théories des genres s'appuient essentiellement sur des critères thématiques et de macro-structure (cf. Gymnich, Neumann & Nünning 2007, entre autres). Pour une réflexion théorique plus récente autour des concepts de genre, texte et discours, cf. Monte & Philippe 2014).

(lexèmes, mots clés, collocations, colligations, schémas actanciels, phraséologismes) (Siepmann 2015, 2016). Une amorce de validation de cette hypothèse existe déjà dans l'étude pionnière sur l'anglais de Stubbs & Barth (2003) mais elle n'a été validée que sur des corpus de taille très modestes, excluant les unités « lexicalement riches », immédiatement identifiables comme spécifiques au genre littéraire, comme par exemple *His thoughts were interrupted by, Il en était là de ses réflexions quand* (Siepmann 2015, 382). Par ailleurs, Biber *et al.* (1999) ont identifié certaines propriétés grammaticales spécifiquement littéraires (les constructions démonstratives *that bloody car of mine* ou existentielles en *there*) mais leurs analyses ne prennent pas en compte les mots-clés, ni les segments répétés spécifiques. La présente étude se propose d'aller plus loin en appliquant une méthodologie innovante, fondée sur des approches textométriques et s'appuyant sur de vastes corpus de textes littéraires français, postérieurs à 1950. Cette étude débutera par un aperçu des travaux existants sur la phraséologie et la stylistique des genres littéraires (section 2). Elle sera suivie, dans la section 3, par la définition de la notion de *motif phraséologique*, puis, dans la section 4, par la présentation des corpus et de la méthodologie du travail. La section 5 sera consacrée à la présentation des ALR spécifiques à six sous-genres littéraires. La section 6 proposera une analyse de quelques motifs autour des expressions de communication verbale (p. ex. *dire d'une voix*) et non verbale (p. ex. *adresser un sourire à*), permettant de distinguer ou de rapprocher les sous-genres pour lesquels ces motifs sont spécifiques. Les conclusions sont présentées dans la section 7.

2 Aperçu des travaux existants

2.1 La phraséologie et les approches de l'idiomaticité

Les phraséologismes constitutifs de l'idiomaticité des textes échappent, pour l'instant, à une définition complète et consensuelle : « extended units of meaning » (Sinclair 2004), « constructions » (Goldberg 1995), « collostructions » (Stefanowitsch & Gries 2003), « collocations » (Hausmann 1979 ; Mel'čuk *et al.* 1995 ; Siepmann 2005 ; Tutin 2010), « lexical bundles » (Biber *et al.* 1999), « motifs séquentiels » (Quiniou *et al.* 2012). On constate, par ailleurs, que les différentes approches qui sous-tendent ces notions déplacent leur objet d'études des séquences figées de la phraséologie traditionnelle (formules routinisées, proverbes, collocations binaires (Hausmann 1979 ; Mel'čuk *et al.* 1995 ; Tutin 2010) vers plusieurs types d'unités lexicales « étendues » (Sinclair 2004). Deux notions, particulièrement prometteuses de ce point de vue, sont les « cadres collocationnels » (Renouf & Sinclair 1991), et les « motifs » (Legallois 2006 ; Longrée & Mellet 2013).

2.2 La phraséologie, la stylistique et la théorie des genres littéraires

À l'heure actuelle, la phraséologie occupe une place centrale au sein de deux domaines de recherche traditionnellement séparés : la linguistique outillée et l'analyse stylistique du

texte littéraire. Un grand nombre de travaux en stylistique (Barthes 1966 ; Greimas 1972 [1982] ; Leech & Short 2007), en stylométrie (Magri-Mourgues 2006), en stylistique de corpus (Stubbs 2005 ; Fischer-Starcke 2010 ; Mahlberg 2013) et en textométrie (Guiraud 1954 ; Brunet 1981) traitent, de manière approfondie, des procédés de style et des spécificités lexicales et grammaticales chez différents auteurs comme Zola, Flaubert, Duras (p. ex. Philippe & Piat 2009 ; Vaudrey-Luigi 2013). D'autres travaux établissent des schémas stéréotypiques récurrents, p. ex. pour le roman policier (Todorov 1980 ; Marion 2009 ; Lits 2011), dans le but de dresser une typologie des genres littéraires. Par ailleurs, les critères retenus dans les théories des genres ont été presque exclusivement basés sur leur contenu (cf. à ce sujet Zymner 2003 ; Frow 2006 ; Gymnich, Neumann & Nünning 2007).

Grâce aux Humanités numériques actuellement en plein essor, les approches traditionnelles des genres s'appuient de plus en plus sur des méthodes innovantes, fondées sur les outils de la linguistique de corpus (cf. entre autres Beauvisage 2001 ; Rastier 2011) et du Traitement automatique du langage (TAL), permettant d'extraire de grandes masses de données langagières pour l'analyse stylistique. Ainsi, récemment, une exploration de grande ampleur sur les constructions lexico-syntaxiques (CLS), spécifiques du roman contemporain, a été réalisée dans le cadre du projet PhraseoRom (<https://phraseorom.univ-grenoble-alpes.fr>). Ses principaux résultats (Novakova & Siepmann 2020, entre autres) ont montré la pertinence de la notion de *motif phraséologique* pour la distinction des sous-genres littéraires.

3 Définition de la notion de *motif phraséologique*

La notion de *motif phraséologique* (ou *textuel*) apparaît comme une nouvelle notion pertinente, permettant d'articuler les analyses du texte littéraire au niveau syntaxico-sémantique avec celles du niveau discursif et textuel. Le motif s'actualise à la fois en langue et en discours (Lala & Siaugues 2020, 110)³. Il ne s'agit donc pas d'éléments fictionnels symboliques ou constitutifs de l'imaginaire d'une œuvre (Bellemin-Noël 1972 ; Richard, 1961, 1979), mais de séquences phraséologiques *observables*, statistiquement saillantes dans un corpus, composées d'unités continues ou discontinues combinant lemmes, catégories morphosyntaxiques, mots-outils, collocations. Ces « unités multidimensionnelles » d'appariement entre forme et sens ont une fonction pragmatique et discursive (Legallois 2012, 45). Sur le plan fonctionnel, le motif est « un « cadre collocationnel » accueillant un ensemble d'éléments fixes et variables, susceptibles d'accompagner la structure textuelle, et simultanément, de caractériser des textes de genres divers » (Longrée & Mellet 2013, 66). Ces regroupements pertinents de patrons lexico-syntaxiques récurrents, dont la spécificité pour tel ou tel genre est établie avec des méthodes lexicométriques, jouent un rôle dans la « cohérence textuelle » (Martin 1983, 16) du discours romanesque. On parle

³ Pour un *aperçu global* de la notion de motif à l'intersection de différentes disciplines comme la folkloristique, la narratologie, la bioinformatique, le TAL et la linguistique (cf. Legallois & Koch 2020, 17-46).

ainsi de la « méthode des motifs », applicable dans la fouille des textes littéraires (Lala & Siaugues 2020, 113 ; Legallois & Koch 2020, 32–41).

En bref, de par son caractère « à la fois structurant et caractérisant » (Mellet & Longrée, 2012, 718), *le motif* s'avère être une séquence opératoire pour la distinction des sous-genres littéraires comme l'ont également montré les travaux du projet PhraseoRom :

The notion of motif lets us fill in the missing link between “macro-level” notions of script or schema that have traditionally been used in cognitive narratology (e.g. study of the plot, isotopies) and the “micro-level” elements that go into making up the script (specific phraseological recurrences). (Novakova & Siepmann 2020, 10)

4 Corpus et méthodologie

Les données sur lesquelles est fondée cette étude sont extraites des corpus numériques, syntaxiquement annotés, constitués pour les besoins du projet PhraseoRom⁴ et regroupés dans la PhraseoBase (<http://phraseotext.univ-grenoble-alpes.fr/phraseobase/index.html>). Les corpus sont interrogeables avec l'outil Lexicoscope (Kraif 2016). Ces corpus réunissent plus de 1000 œuvres contemporaines à partir de 1950, réparties en 6 sous-genres (selon des critères éditoriaux et libraires, prix littéraires, recherches sur Internet, lecture d'extraits) : littérature générale (GEN), romans policiers (POL), sentimentaux (SENT), historiques (HIST), science-fiction (SF) et *fantasy* (FY). Le Tableau 1 récapitule la composition de ces corpus en français :

Sous-genres	Auteurs	Textes	Tokens
FY	43	109	13 966 566
GEN	172	442	34 593 554
HIST	40	114	14 193 642
POL	85	196	18 010 790
SENT	38	110	9 516 633
SF	40	150	13 528 393
Total	418	1 121	103 809 578

Tableau 1. Nombre d'auteurs, de romans et de *tokens*⁵ dans les corpus français PhraseoRom

⁴ Les corpus comparables (français, anglais et allemands) et parallèles (français-anglais et anglais-français) sont consultables en libre accès sur <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/index.html>.

⁵ En traitement automatique du langage (TAL), le *token* (de l'anglais « jeton ») est une unité minimale d'information détectée lors de l'« analyse lexicale » ou « tokenization ». Celle-ci cherche à transformer un texte en une série de *tokens* individuels (éléments atomiques de la chaîne), par exemple dans « J'ai froid. », le modèle de tokenisation doit séparer le « J' » comme étant un premier mot. Par ailleurs, un token n'est pas forcément un mot, il peut être aussi un signe de ponctuation.

Dans la lignée des travaux en linguistique de corpus anglo-saxonne (Sinclair 2004 ; Hoey 2005 ; Biber & Conrad 2009), l'approche appliquée ici est essentiellement inductive (*corpus driven*) : les observations sont guidées par les données, issues des grands corpus littéraires recensés dans le Tableau 1. La méthode consiste à extraire automatiquement des *arbres lexico-syntaxiques récurrents* (ALR) (Kraif 2016) qui regroupent des unités lexicales reliées entre elles par des relations de dépendance syntaxique.

La spécificité des expressions récurrentes autour d'un mot pivot dans un des sous-genres du corpus est calculée par rapport à l'ensemble du corpus PhraseoRom, selon l'indice statistique *loglikelihood ratio* (LLR, Dunning 1993). La Figure 1 illustre l'expression récurrente, spécifique au sous-genre POL, *allumer une nouvelle cigarette* qui se présente sous forme d'Arbre Lexico-syntaxique Récurrent :

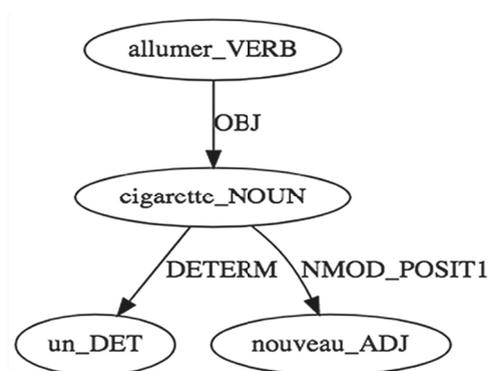


Figure 1. Extraction automatique de l'arbre lexico-syntaxique récurrent (ALR) *allumer une nouvelle cigarette*, spécifique au sous-genre policier (POL)

A la différence de la méthode proposée par Quiniou *et al.* (2012), nos ALR correspondent à des *structures hiérarchiques* (arbres de dépendance). Selon la méthodologie développée dans le cadre de PhraseoRom, les critères retenus pour la sélection des ALR sont les suivants⁶ :

- le seuil de **spécificité statistique** (l'indice LLR) a été fixé à 10,83. C'est le seuil à partir duquel la surreprésentation de l'ALR dans un corpus peut être considérée comme statistiquement significative. Il s'agit d'un rapport de vraisemblance permettant d'évaluer objectivement si la répartition des unités linguistiques au sein d'un corpus est aléatoire ou non ;

⁶Pour plus de détails concernant la méthodologie des ALR, comparée à d'autres méthodologies pour l'extraction de motifs textuels (cf. Legallois et Koch 2020, 35–36).

- le critère de **fréquence absolue** : le nombre d’occurrences d’un ALR doit être supérieur à 10 ;
- le critère de **dispersion** : l’ALR apparaît au moins chez 20 % des auteurs d’un sous-corpus pouvant aller jusqu’à 50 % des auteurs du corpus ;
- le critère **morpho-syntaxique** : l’ALR doit comporter obligatoirement un verbe, ce qui exclut d’emblée des expressions exclusivement référentielles (*A la tombée de la nuit, Monsieur le Procureur, les nains de jardin*) ;
- la **taille** des ALR : ils doivent comporter au moins 3 mots dont un verbe pivot, et ce jusqu’à 7 mots.

Une fois extraites sous forme d’ALR, les données brutes ont été classées et analysées sur quatre niveaux : lexical, syntaxique, sémantique et discursif, dans la lignée des modèles fonctionnels et contextualistes (Sinclair 2004 ; Hoey 2005). Une grille a été conçue (Goossens *et al.* 2020) pour l’annotation sémantique des 8415 items retenus pour le français. Cette grille regroupe des dimensions (catégories) sémantiques comme *action, communication, cognition*⁷, etc. (Tableau 2) :

Dimensions sémantiques	Valeurs	Exemples
Action	Mouvement	<i>faire un bond, lever la main, clapped his hands</i>
	Déplacement	<i>traverser le hall, walk along the beach, follow into the kitchen</i>
Setting	Lieu	<i>la pièce était plongée dans le silence, l’endroit est désert</i>
	Temps	<i>le temps était gris, dehors il pleuvait</i>
Communication	Verbale	<i>raconter une histoire, crier de joie, dire d’une voix, have a conversation</i>
	Physique	<i>lever les yeux vers lui, acknowledge with a nod</i>
Cognition		<i>je ne me souviens pas, think for a moment, going through his mind</i>
Qualia	Sensation	<i>avoir froid, my heart is beating</i>
	Affect	<i>se mettre en colère</i>

Tableau 2. Extrait de la grille sémantique du projet PhraseoRom

⁷ Cette grille a été créée dans la continuité des ontologies WordNet (Fellbaum 1998) ou des *functional groups clusters* de Mahlberg (2007). Le manuel d’utilisation de la grille sémantique (en français et en anglais) est consultable sur <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/ressources-fr.html>.

Cette étape importante d'annotation sémantique des ALR a permis ensuite de les regrouper automatiquement⁸ et de faire émerger des *motifs textuels*. Dans notre méthodologie, les ALR constituent des « guides » pour l'identification des motifs. Enfin, un système d'annotations fines des fonctions discursives (FD) des motifs a été également élaboré par les stylisticiens du projet. La liste des principales fonctions discursives – *narratives* et *descriptives* (Adam 2011, 267) – a été enrichie, suite à l'analyse d'au moins une centaine d'occurrences des motifs extraits des différents sous-corpus. Ainsi, d'autres FD – *affektive*, *pragmatique*, *cognitive* (avec deux sous-classes, *commentative* et *mémorielle*) – ont été rajoutées⁹. Ces FD ont permis de prendre en compte la dimension textuelle des expressions lexico-syntaxiques (extraites sous forme d'ALR), c'est-à-dire leur rôle dans la structuration des textes littéraires.

La section 5 présentera les expressions statistiquement spécifiques aux six sous-genres comparés ici. Dans la section 6 seront analysés plus en détails quelques motifs formés autour des ALR classés dans la dimension sémantique *communication verbale* et *non verbale* (cf. Tableau 2) au sein des différents sous-genres ; nous montrerons enfin comment ces motifs permettent de les différencier ou de les rapprocher.

5 Les expressions lexico-syntaxiques statistiquement spécifiques aux six sous-genres littéraires étudiés

Une série d'expressions ont été retenues ici parmi les plus fréquentes extraites dans chacun des six sous-genres comparés. L'approche *corpus driven* révèle des résultats aussi bien attendus, présentant des récurrences thématiques mais aussi inattendus, considérés plutôt comme étant des récurrences stylistiques (cf. Kraif & Novakova & Sorba 2016).

5.1 Expressions spécifiques au sous-genre littérature générale (GEN)

Les romans classés dans le sous-genre GEN présentent une grande variété d'expressions appartenant aux différentes dimensions sémantiques (cf. Tableau 2). Ce nombre important de résultats pourrait être dû, entre autres, à la taille importante du sous-corpus GEN (34 M. de mots, cf. Tableau 1). L'observation des données a montré que les ALR spécifiques à GEN appartiennent surtout à la catégorie des actions. Le Tableau 3 donne un aperçu de ces résultats : il présente une sélection d'ALR, classés par ordre décroissant de leur indice de spécificité LLR (cf. section 4) :

⁸ Le regroupement automatique des 8415 ALR pour le français a été effectué à l'aide du logiciel Word2vec.

⁹ La liste complète et les définitions des fonctions discursives (FD) des motifs, établies dans le cadre du projet PhraseoRom sont consultables sur <http://phraseotext.univ-grenoble-alpes.fr/phraseobase/ressources-fr.html>.

Dimension sémantique	Expressions	Indice LLR
Actions thématiques	écrire un roman	91,65
	lire un livre	52,80
Actions du quotidien	fumer une cigarette	240,86
	prendre le train	132,05
	prendre un verre	72,53
	ouvrir la fenêtre	67,42
	acheter du pain	19,85
	faire la vaisselle	18,54
	allumer une lampe	12,74
Actions : mouvements/déplacements	marcher dans la rue	68,88
	aller à la messe	43,39
	monter les escaliers	26,35
	sortir dans la rue	20,69
Actions contemplatives	regarder la mer	90,40
	regarder le ciel	71,80
	se regarder dans la glace	29,39

Tableau 3. Expressions spécifiques à GEN, classées selon leur dimension sémantique et l'indice statistique LLR¹⁰

Comme le montre le Tableau 3, dans le sous-corpus GEN émergent, sans surprise, des actions « thématiques », qui renvoient à des activités d'écriture <écrire une lettre| un roman (d'amour, d'espionnage)| un poème| une ligne| un article| une page| un scénario> ou de lecture <lire un livre| un (nouveau) roman, une lettre d'amour, de son fiancé>¹¹. Ce résultat n'est pas étonnant, car les activités scripturales témoignent de l'intérêt de l'auteur/narrateur envers le processus d'écriture, fort légitime pour ce sous-genre (Gonon *et al.* 2018, 9).

On trouve aussi, dans la littérature générale, une surabondance d'expressions renvoyant à des (menues) actions, liées au quotidien des personnages : ils fument des cigarettes, font la vaisselle, allument des lampes, achètent du pain, prennent (boivent) un verre, ouvrent des fenêtres. D'autres actions renvoient à leurs mouvements ou déplacements : les protagonistes prennent le train, marchent dans la rue, vont à la messe, montent des escaliers. Enfin, d'autres actions, plus contemplatives, émergent aussi comme spéci-

¹⁰ Les mises à jour régulières du Lexicoscope 2.0 peuvent entraîner de légères différences entre les chiffres indiqués ici et les données statistiques du logiciel sans que cela n'impacte la validité de nos résultats. Pour consulter le Lexicoscope 2.0. http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/

¹¹ Les ALR sont indiqués en italique et entre chevrons, p. ex. <écrire un roman>. Les variations paradigmatiques du cœur du motif sont séparées par |, tandis que les extensions syntagmatiques sont indiquées entre parenthèses.

fiques du GEN : les personnages se regardent dans la glace, regardent la mer, le ciel. En revanche, peu d'expressions verbo-nominales appartenant à la dimension *communication verbale* et *non verbale* apparaissent dans ce sous-genre (cf. 6.1).

5.2 Expressions spécifiques au sous-genre roman sentimental (SENT)

Les données du sous-corpus SENT révèlent une surabondance d'expressions liées au langage du corps, qui contribuent à la communication non verbale (physique) entre les personnages. Le tableau 4 illustre les ALR les plus spécifiques de SENT dans ces catégories :

Dimension sémantique	Expressions	LLR
Actions thématiques	faire l'amour	244,39
	déposer un baiser	97,13
Actions : mouvement du corps ou des parties du visage	prendre dans ses bras	241,22
	poser sa main	165,30
	prendre par la taille	113,72
Communication non verbale	froncer les sourcils	141,33
	lever les yeux	129,68
	adresser un sourire à	106,42
	esquisser un sourire	81,30
	baisser les yeux	48,77

Tableau 4. Expressions spécifiques à SENT

Des ALR correspondant à des actions thématiques (<*faire l'amour*>, <*déposer un baiser*>) apparaissent sans surprise dans SENT. On relève également une surabondance d'expressions renvoyant aux mouvements du corps (<*prendre dans ses bras*>, <*poser la main*>) ou du visage (<*baisser les yeux*>, <*froncer les sourcils*>, <*esquisser un sourire*>). Ces ALR de communication non verbale génèrent des motifs phraséologiques abondants qui traduisent aussi, de manière directe ou implicite, les affects des personnages : par ex. <*froncer*| hausser| lever (légèrement) *les* (beaux) *sourcils*| cils (roux) (d'un air sévère). Dans la section 6.2 sera présentée une analyse des motifs identifiés autour de l'ALR <*adresser un sourire*>, statistiquement spécifique à SENT (cf. Tableau 4).

5.3 Expressions spécifiques au sous-genre roman policier (POL)

A la différence des sous-genres GEN et SENT, dans le POL émergent des actions thématiques plus ciblées comme par ex. <*découvrir le corps*| le cadavre>, <*prévenir la police*>, <*ouvrir le coffre*>. Le tableau 5 regroupe une partie des expressions présentant un indice de spécificité élevé dans POL :

Dimension sémantique	Expressions	LLR
Actions thématiques	découvrir le corps	159,71
	prévenir la police	154,85
	ouvrir le coffre	132,15
Actions : mouvements du corps	secouer la tête	216,56
	hausser les épaules	104,87
	hocher la tête	103,89

Tableau 5. Expressions spécifiques à POL

Les résultats montrent qu'à la différence de SENT, les ALR liés au langage du corps et donc, à la communication non verbale, sont beaucoup moins nombreux et apparaissent surtout autour des deux pivots *tête* (<*secouer*| *hocher la tête*>) ou *épaules* (<*hausser les épaules*>), ce qui permet de distinguer, entre autres, ces deux sous-genres (cf. section 6.2).

5.4 Expressions spécifiques au sous-genre historique (HIST)

A la différence des sous-genres examinés ci-dessus (GEN, SENT, POL), dans le sous-genre historique (HIST), ce sont les ALR de la catégorie *communication verbale* qui sont les plus représentés, ce qui est un résultat surprenant :

Dimension sémantique	Expressions	LLR
Communication verbale	dire avec un sourire	343,72
	dire d'une voix	370,34
	donner l'ordre	283,08
Actions thématiques	faire la guerre	119,60
	mettre le siège	118,28
	quitter la cour	107,10
	rétablir l'ordre	67,84

Tableau 6. Expressions spécifiques à HIST

Les ALR <*dire avec sourire*> ou <*dire d'une voix*> agrègent des motifs variés d'interaction verbale (cf. 6.1, *infra*). En revanche, des ALR exprimant des actions thématiques comme <*faire la guerre*>, <*mettre le siège*> (devant Paris, la cité, son château), <*quitter la cour*> sont naturellement spécifiques au roman historique.

5.5 Expressions spécifiques au sous-genre science-fiction (SF)

Dans la SF, ce sont surtout les expressions autour des actions liées à des déplacements virtuels dans le temps et l'espace (<*voyager dans le temps*>, <*traverser l'espace*>), ou à des actions accomplies par des entités inanimées (<*l'écran s'allume, s'éteint*>, <*la sonnerie retentit*>) qui émergent des données. Elles contribuent, naturellement, à l'univers plus « scientifique » de la science-fiction (Klein 1997). En revanche, d'autres ALR beaucoup moins attendus apparaissent comme étant des marqueurs génériques : il s'agit,

par exemple, d'ALR relevant de la catégorie sémantique de la *cognition* (<prendre conscience>, <faire des hypothèses>, <mettre au point>) (Goossens *et al.* 2020, 211). Ces expressions soulignent la mobilisation de la réflexion, la construction du raisonnement logique qui correspond à la plus grande « scientificité » de ce sous-genre, comparé à la FY. Elles entraînent une posture de lecture « plus vigilante », liée à la recherche d'indices (Saint-Gelais 1999). Le Tableau 7 récapitule ces résultats :

Dimension sémantique	Expressions	LLR
Actions thématiques	voyager dans le temps	127,13
	l'écran montre	118,44
	apparaître sur l'écran	110,66
	l'écran s'éteint	81,11
	l'écran s'allume	52,08
	traverser l'espace	45,26
Cognition	mettre au point	233,02
	prendre conscience	184,57
	avoir l'impression	146,99
	arriver à la conclusion	31,55
	trouver la solution	25,44

Tableau 7. Expressions spécifiques à SF

Les expressions relevant de la communication verbale ou non verbale y sont très peu représentées, sauf l'ALR <lancer un coup d'œil> qui s'est avéré, de manière inattendue, spécifique à la SF (cf. section 6.2).

5.6 Expressions spécifiques au sous-genre *fantasy* (FY)

Enfin, dans la FY, hormis les expressions d'actions thématiques, des ALR renvoyant à la *communication verbale* sont apparus comme spécifiques :

Dimensions sémantique	Expressions	LLR	
Actions thématiques	dégainer l'épée	361,14	
	encocher une flèche	229,05	
	briser la nuque	97,01	
	se jeter dans la mêlée	69,15	
	arracher le cœur	66,71	
	serrer le poing	63,65	
	briser les os	47,95	
	frapper dans le dos	35,28	
	Communication verbale	pousser un cri	235,91
		arracher un cri	62,68

Tableau 8. Expressions spécifiques à FY¹²

¹² Pour une analyse détaillée des différentes catégories sémantiques d'ALR spécifiques à la *fantasy*, comparée à la SF, cf. Goossens *et al.* (2020).

Le vocabulaire médiéval, caractéristique de ce sous-genre, apparaît ici à travers la surreprésentation d'ALR exprimant des actions guerrières violentes ou agressives (<*dégainer l'épée*>, <*encocher une flèche*>, <*arracher le cœur*>). S'en détachent aussi, de manière plus inattendue, des ALR relevant de la communication entre les personnages comme <*pousser des cris*> ou <*arracher des cris*>, qui seront analysés dans la section 6.1.

5.7 Synthèse sur les dimensions sémantiques des ALR spécifiques aux six sous-genres

Suite à l'observation des données, quelques tendances se dégagent : les ALR relevant de la catégorie des actions sont parmi les plus représentés mais les types d'actions varient selon les sous-genres : dans GEN, ce sont les actions quotidiennes qui sont surreprésentées, dans SENT, les actions liées aux mouvements du corps, dans POL, les actions thématiques policières, dans HIST, la communication verbale et les actions guerrières. Dans le sous-genre FY, les marqueurs génériques relèvent du vocabulaire guerrier médiéval et violent et de la communication par des cris, en revanche, dans SF ce sont les actions thématiques réalisées par des objets ou des instruments inanimés (vocabulaire plus « scientifique ») et la catégorie de la cognition qui s'imposent statistiquement. Plus généralement, la méthode des ALR, leur classement sémantique, ainsi que leurs regroupements en motifs phraséologiques permettent de distinguer les sous-genres analysés. Afin d'illustrer plus concrètement ces résultats, une étude de cas est proposée dans la section 6, portant sur les ALR de la dimension sémantique de la *communication*.

6 Les motifs de la communication (verbale et non verbale) pour distinguer les six sous-genres

Cette catégorie se subdivise en deux valeurs (cf. Tableau 2) : la communication verbale (<*raconter une histoire*>, <*crier de joie*>, <*dire d'une voix*>) et la communication physique (<*esquisser un sourire*>, <*lever les yeux vers*>). Ces ALR, qui génèrent des motifs, sont globalement bien représentés dans les six sous-genres étudiés mais des tendances, parfois surprenantes, se dégagent, révélées grâce à notre approche *corpus driven*. La section (6.1), analyse les motifs autour des expressions de la communication verbale dans HIST, comparé à GEN et à FY, et dans la section (6.2) sont comparées les expressions de la communication non verbale dans SF, SENT et POL.

6.1 Les motifs de la communication verbale dans HIST, GEN et FY

Un résultat inattendu a émergé des données, notamment la haute spécificité des ALR <*dire d'une voix*> (LLR 370, 34) et <*dire avec sourire*> (LLR 343,72) dans les romans historiques (cf. Tableau 6). Ces ALR agrègent autour d'eux des motifs phraséologiques riches¹³ : des variations paradigmatiques sont observées aussi bien sur le verbe dire|

¹³ Cf. aussi à ce sujet Gonon & Sorba 2019.

ordonner| répliquer| répondre| expliquer| lancer| prononcer| ajouter| crier| reprendre que sur le nom (voix| ton), mais aussi sur l'extension adjectivale : voix forte| sourde| claire| étouffée| ferme| calme| grave| tremblante| sèche| entrecoupée). En voici quelques exemples, issus du sous-corpus HIST :

- (1) *Et comme je le regardais d'un air interrogatif, il **dit d'une voix aussi douce et basse** que s'il me parlait à travers la grille d'un confessionnal : – Monsieur le Comte, vous connaissez, je crois, le chanoine Fogacer ?* (Merle R., *Fortune de France 9. Les Roses de la vie*, 1995)
- (2) *Comme il se dirigeait vers la porte, le capitaine **dit d'une voix douce qui tremblait un peu** : – Tout de même, Lambert, je crois que nous avons vraiment là de braves gens.* (Clavel B., *Le Soleil des Morts*, 1998)
- (3) *Le silence se prolongea quelques instants, lourd et angoissant puis, se levant brusquement en claquant des mains, la baronne **dit d'une voix joyeuse** : – Avec votre aide, marquis, nous viendrons bien à bout de toutes ces méchantes gens !* (Fajardie F. *Le conseil des troubles*, 2007)

La distribution syntaxique du motif dans les exemples de (1) à (3) est identique : il est présent dans les séquences qui précèdent le discours direct. Le motif fait ici partie d'un script (Baroni 2007)¹⁴ n'ayant pas d'impact important sur l'enchaînement narratif (FD *infra-narrative* ou *faiblement narrative*). Comme le montrent les exemples, l'acte de parole correspond à une action minimale. Par ailleurs, l'extension adjectivale dans la syntaxe du cadre collocationnel du motif, qui correspond à l'ALR *dire d'une voix*, peut provoquer un déplacement du focus des qualités propres de la voix (intensité, diction) vers la traduction des affects des personnages à travers l'acte de langage (*dire d'une voix douce qui tremblait* en 2) ou *d'une voix joyeuse* en 3). Ici, la fonction discursive du motif est mixte : *infra-narrative* et *affective*.

En revanche, dans GEN, on observe une forte présence de verbes de parole : <N + *dire, lancer, proposer, annoncer, répéter, expliquer, avouer, mentionner, prier, citer, avouer*>. Ces ALR à pivot verbal ne donnent cependant pas lieu à une grande variation sur le plan syntagmatique. Souvent, ces verbes de parole apparaissent à la 1^{ère} personne : <*je lui dis*> (LLR 726,16), <*je n'ose pas demander*> (33,42), <*je n'ose pas dire*> (26,36), <*je lui cite*> (22,27), <*je vous écris*> (21,83), <*j'ai failli lui demander*> (20,73), <*je lui écris*> (19,15), <*je le mentionne*> (18,39), <*je peux dire*> (18,25), <*j'ai envie de lui dire*> (17,01), etc. Les ALR cités correspondent en effet « à une interlocution tournée

¹⁴ Selon Baroni (2007, 175), les « scripts » correspondent à des actions routinières qui n'ont pas besoin d'être développées dans les récits car le lecteur connaît « leurs scénarios » communs, comme c'est le cas d'allumer une cigarette, mettre ses chaussures, ouvrir la fenêtre ou prendre le café.

vers l'intériorité ou une discussion tournée vers soi-même » (Gonon & Sorba 2019, 4). Plus généralement, comme l'indiquent les auteures, ces résultats pourraient s'expliquer par le fait que : « Les romans de littérature générale se caractérisent plutôt comme des récits psychologiques, avec une réflexivité bien plus marquée du narrateur-personnage ou des autres personnages » (*idem*, 14). On trouve aussi dans GEN des ALR de type verbo-nominal comme <*raconter une histoire*> (LLR 64,00), <*adresser la parole*> (LLR 25,90), <*faire des discours*> (LLR 22,31) qui sont davantage sujets à des variations mais leur fréquence est beaucoup plus restreinte ce sous-genre, surtout comparée au grand nombre de verbes de parole qui y sont présents.

Enfin, en FY, les personnages (monstres humanoïdes, soldats, hérauts, chevaux, faucons, loups) communiquent beaucoup en poussant des cris, des rugissements, des grognements, des hurlements, des hennissements. On y trouve des motifs bien fournis et variés autour de l'ALR <*pousser un cri*> (LLR 235, 91) et sa version causative <*arracher un cri*> (LLR 62,68). En schéma : X <*pousse*| étouffe| arrache (alors, aussitôt, soudain) un (petit, grand| long) cri (hennissement, rugissement, grognement) + (ADJ strident| terrible| déchirant| de rage| de douleur| de guerre| de joie). Employées en phrase indépendante, ces expressions renvoient à des actes expressifs de langage :

- (4) *Kroll poussa un hurlement bestial. Subitement, comme en réponse à un appel, un feu vorace jaillit, crevant le cuir de ses gants.* (Thomas J., *Lunardente. La cité noire*, 2012)
- (5) *Dans son box, Murmure poussa un hennissement. Il avait entendu la voix de sa maîtresse et, songeant qu'elle le rejoignait, lui témoignait sa joie de la voir.* (Bottero P., *Le pacte des March'Ombre. Tome3 : Ellana : La prophétie*, 2008)

En (4), Kroll pousse un hurlement bestial ayant comme conséquence le jaillissement d'un feu ; en (5) Murmure pousse un hennissement pour attirer l'attention et la bienveillance de sa maîtresse. L'acte de langage a ici des répercussions sur les autres actions. A l'enchaînement narratif (FD essentiellement *narrative* du motif), se superpose parfois une fonction discursive *pragmatique* (acte expressif de langage). Une autre distribution très fréquente de ce motif dans FY est la construction gérondive ou participiale (p. présent). Elle apporte une précision qui enrichit la description du personnage et de ses émotions :

- (6) *Une petite créature entra alors en rampant sur le plancher brillant du temple en poussant des cris de plaisir.* (A. Robillard, *Les Chevaliers d'émeraude 6. Le Journal d'Onyx*, 2004)
- (7) *Il s'arc-bouta pour libérer ses jambes, qu'il lança de toutes ses forces dans la partie la plus molle de l'abdomen du Scorpion, lui arrachant un cri de rage.* (Debien C., *Le cycle de Lahm I L'éveil du roi*, 2010)

Les motifs acquièrent ainsi, en (6) et en (7), une double fonction : *infra-narrative* et *indirectement descriptive*. Ces FD peuvent se voir enrichies d'une dimension affective et ce, grâce à la présence de l'extension du pivot nominal *cri* par un complément du nom (cri de plaisir (ex. 6), de rage, ex. (7)). Il s'agit d'affects qui traduisent des états euphoriques (*cris de victoire, de triomphe*) ou des émotions positives ou négatives (*cris de plaisir, de joie, de rage*).

En bref, les résultats montrent que les motifs de communication verbale permettent de distinguer les trois sous-genres : HIST, GEN et FY. La section (6.2) propose une analyse des motifs de la communication non verbale qui se sont révélés comme étant spécifiques à trois autres sous-genres : SF, SENT, POL.

6.2 Les motifs de la communication non verbale dans SF, SENT et POL

Même si le nombre d'ALR de la catégorie de la communication non verbale est très restreint dans SF (comme d'ailleurs ceux de la catégorie de la communication verbale), on y trouve l'ALR <*lancer un coup d'œil*> (LLR 80,12), un résultat assez inattendu pour ce sous-genre. Cet ALR présente des variations syntagmatiques sur le verbe (*lancer, jeter, glisser*) et sur le N (*coup d'œil, regard*) qui est suivi d'une extension adjectivale (*inquisiteur, désespéré, halluciné, discret, exaspéré, maussade*) :

- (8) *Il prend des poses en ne cessant de répéter « Quoi d'neuf, docteur ? » (Il m'a lancé un coup d'œil inquisiteur.) Vous avez une idée de ce qui peut se passer ?* (Wagner R., *Les futurs mystères de Paris*, 1998)
- (9) *Sylvaine lança un coup d'œil désespéré à Edmé. « J'pensais trouver le rat dans l'coin, reprit le colosse. Pas tomber sur vous deux. Comment vous avez fait pour... »* (Bordage P., *Porteurs d'âmes*, 2007)

Les motifs au sein du discours direct (ex. 8 et 9), relevés dans SF, contribuent à rythmer l'échange (FD *infra-narrative*). Lorsqu'ils apparaissent dans des phrases indépendantes ou coordonnées, le motif du regard lancé introduit une brève séquence descriptive : il s'agit le plus souvent de la description d'un autre personnage :

- (10) *Elle me lança un regard maussade. Je la détaillai plutôt mignonne quoiqu'un poil gironde, genre fleur en bouton. Prometteuse, en tout cas. D'appétissantes lèvres boudeuses, des cheveux noirs incroyablement fins, des yeux pers ennuyés, aux cernes mauves.* (Ligny J.-M., *Aqua*, 2006)
- (11) *Il écrasa les rigoles de sueur qui lui dégouлинаient dans les yeux et lança un coup d'œil exaspéré au donneur. Vrai qu'il ressemblait à sa mère. La même finesse de traits, les mêmes yeux clairs, la même chevelure folle, la même allure à la fois gauche et gracile.* (Bordage P., *Les derniers hommes 2 Le cinquième ange*, 2000)

Comme on le voit en (10) et en (11), la séquence descriptive peut être assortie d'un commentaire et d'une réflexion du personnage « observateur » au sein d'une communication non verbale. Dans ce cas, les motifs véhiculent également une composante cognitive qui apparaît dans les commentaires « Prometteuse, en tout cas. » (en 10) et « Vrai qu'il ressemblait à sa mère. » (en 11). Le motif se voit attribuer ici une FD mixte : *descriptive* et *cognitive*.

A la différence des autres sous-genres, le langage du corps est très présent dans SENT (cf. Tableau 4). Les personnages y communiquent avec le sourire (communication non verbale). Le cœur du motif <*adresser*| lancer| faire un sourire> connaît des extensions adjectivales variées (sourire *gêné*, *résigné*, *complice*, *enfantin*, *épanoui*), qui véhiculent une dimension affective. Ces motifs que l'on trouve souvent dans le dialogue ont ainsi une double fonction discursive : *infra-narrative* (contribuant à meubler la conversation) et *affective* : sourire gêné (ex. 12) ou épanoui (ex. 13) :

(12) – Marie-Claude ! La jeune fille **lui adressa un sourire gêné**. – Tiens ! Un revenant ! s'exclama-t-elle, d'une voix qui lui parut contrainte. (Simart H., *Trahison*, 1980).

(13) – Êtes-vous libre tout de suite ? ... C'est-à-dire dès aujourd'hui... même maintenant ! – Oui, dis-je fermement. La secrétaire très, très enceinte **me lança un sourire épanoui**. (De Buron N., *C'est fou ce qu'on voit de choses dans la vie !* 2006)

Par ailleurs, le sourire offre parfois une voie d'accès aux pensées et à l'état d'esprit du personnage :

(14) Steven la fixa avec attention. Elle **lui adressa un immense sourire** : qu'il n'aille surtout pas imaginer qu'elle était jalouse ! (Boissard J., *Croisière*, 1988)

(15) Il **lui adresse un pâle sourire qui signifie** : ça va mieux, j'assume ; elle a un immense mouvement de tendresse envers lui et manque de lui proposer de l'accompagner chez Clara. (Pancol K., *Encore une danse*, 1998)

Ainsi, en (14), le motif est suivi d'un commentaire : *qu'il n'aille surtout pas imaginer qu'elle était jalouse*, en (15) la signification du sourire est décryptée : *ça va mieux, j'assume*. À la fonction *infra-narrative* du motif s'ajoute, dans ces cas, une dimension *cognitive commentative*.

Enfin, dans POL, où dominent les ALR d'actions policières (<*découvrir le cadavre*>, <*ouvrir le coffre*>, <*prévenir la police*>, cf. Tableau 5), les ALR de communication non verbale se font rares, surtout si on les compare au sous-genre SENT. Deux ALR statistiquement spécifiques dans cette catégorie sont, toutefois, présents dans POL autour des

pivots nominaux *tête* et *épaules* : <hausser les épaules> (LLR 104,87), <hocher la tête> (LLR 103,89). Les deux motifs fonctionnent de manière identique sur le plan discursif. Ils ouvrent (ex. 16) ou ferment le dialogue (ex. 17) et ont, de ce fait, une FD infra-narrative :

- (16) *Je les lui tendis. Il les parcourut et **hocha la tête**. – Hum... Détective privé... hum... Demeurez loin ?* (Malet L., *Corrida aux Champs-Élysées*, 1956)
- (17) – *Et les cartons d'allumettes. Qu'est-ce que ça signifie ? demanda le sergent-détective... Le Gnome **haussa les épaules**.* (Michaud M., *Je me souviens*, 2012)

Les ALR en question peuvent s'insérer aussi dans une série d'actions et contribuent à la progression narrative (FD narrative) :

- (18) *Il **hocha la tête** et ils s'élançèrent du même pas vers le sentier. Les uns après les autres, les bungalows s'envolaient en libérant leur boule de lave.* (Chattam M., *Le cycle de l'homme 3 La théorie Gaia*, 2008)
- (19) *Il **haussa les épaules** et se dirigea vers son bureau : même si on allait sûrement le retrouver bientôt, il n'y avait pas de risque à prendre. Il appela donc la police de Montréal pour lui faire part des exigences de Hamel.* (*idem*)

En bref, les motifs qui expriment des actes de communication non verbale (<lancer un coup d'œil>, <adresser un sourire>, <hocher la tête>, <hausser les épaules>) fonctionnent différemment sur le plan discursif dans les sous-genres comparés : le motif du sourire dans SENT remplit des fonctions discursives subtiles et variées : infra-narrative et affective dans le dialogue, infra-narrative et cognitive en parataxe juxtaposée. Les données ont révélé aussi, de manière inattendue, la présence statistiquement significative du motif <lancer un coup d'œil> dans SF, remplissant une fonction infra-narrative (lorsqu'il est placé dans le dialogue) ou descriptive au début d'une séquence contenant un commentaire sur un autre personnage vers lequel le regard est lancé. Enfin, dans POL, les ALR autour des pivots nominaux *tête* ou *épaules* ont une FD essentiellement infra-narrative. L'analyse stylistique fine des FD des motifs montre que ces derniers se distinguent sur le plan discursif selon leurs distributions et environnements contextuels et qu'ils sont inégalement représentés dans les sous-genres contrastés.

7 Conclusion

Les motifs de la communication verbale et non verbale, identifiés grâce à notre approche *corpus driven*, ont permis de révéler certaines différences entre les sous-genres littéraires étudiés. Cette approche a fait aussi émerger des expressions que l'on ne soupçonne pas

intuitivement d'être spécifiques à tel ou tel sous-genre, comme par ex. le motif <dire d'une voix> dans le roman historique, comparé à la littérature générale où l'on trouve une sur-représentation statistique de verbes *dicendi* ne formant pas ou peu de motifs. En FY, on découvre les expressions de communication verbale renvoyant à des cris, à des hurlements. Par ailleurs, les expressions de la catégorie communication non verbale (<prendre dans ses bras>, <poser la main>, <froncer le sourcil>, <adresser un sourire>) sont spécifiques surtout à SENT, où le langage du corps est surreprésenté. Elles sont donc distinctives pour SENT, surtout comparé à POL, où seulement <hocher la tête> et <hausser les épaules> apparaissent comme spécifiques dans cette catégorie. Enfin, en SF, on observe un nombre très restreint d'expression de communication verbale et non verbale, excepté le motif inattendu généré par l'ALR <lancer un coup d'œil>. Ainsi, l'étude a montré que les motifs, identifiés à l'aide d'outils lexicométriques sur de grands corpus littéraires, constituent des schémas pouvant nous renseigner sur la structuration et l'interprétation des textes (Gonon & Sorba 2019, 19). Plus généralement, les motifs assurent le lien entre le micro-niveau (les variations récurrentes sur le plan syntagmatique et paradigmatique) et le macro-niveau discursif (celui du script narratif ou fictionnel), et ce à travers l'étude fine de leurs fonctions discursives (Novakova & Siepmann 2020), comme proposé aussi dans cette étude. Ces séquences phraséologiques récurrentes permettent d'analyser et de distinguer les sous-genres littéraires. La phraséologie étendue peut donc contribuer à l'élaboration « d'une théorie opérationnelle des genres » (Rastier 2011, 72).

Références

- Adam, Jean-Michel. 1992 [2011]. *Les textes : types et prototypes*. Paris : Armand Colin.
- Baroni, Raphaël. 2007. *La Tension narrative : suspense, curiosité et surprise*. Paris : Seuil.
- Barthes, Roland. 1966. Introduction à l'analyse structurale des récits. *Communications* 8, 1–27.
- Beauvisage, Thomas. 2001. Exploiter des données morphosyntaxiques pour l'étude statistique des genres : application au roman policier. *TAL* 43. Available at: <http://www.revue-texto.net/Inedits/Beauvisage/index.html>.
- Bellemin-Noël, Jean. 1972. Le motif des orangers dans La Chartreuse de Parme. *Littérature* 5, 26–33.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London : Longman.
- Biber, Douglas & Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge : Cambridge University Press.
- Brunet, Etienne. 1981. *Le Vocabulaire français de 1789 à nos jours*. Geneva/Paris : Slatkine et Champion.
- Dunning, Ted. 1993. Accurate Methods for the Statistics of Surprise and Coincidence: Computational Linguistics. *Computational Linguistics* 19 (1), 61–74.

- Fellbaum, Christiane. 1998. A Semantic Network of English: The Mother of All Word-Nets. *Computers and the Humanities* 32, 209–222.
- Fischer-Starcke, Bettina. 2010. *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. London : Continuum.
- Frow, John. 2005. *Genre*. London/New York : Routledge.
- Gonon, Laetitia, Vannina Goossens, Olivier Kraif, Iva Novakova, Julie Sorba. 2018. Motifs textuels spécifiques au genre policier et à la littérature « blanche ». In *Actes du VIème Congrès Mondial de Linguistique Française, 9–13 juillet 2018*. Mons : Institut de Linguistique Française. Available at : <https://doi.org/10.1051/shsconf/20184606007>
- Gonon, Laetitia & Julie Sorba. 2019. Phraséologismes spécifiques dans les romans historiques et dans les romans de littérature blanche. *Journal of French Language Studies* 30 (1), 1–20.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Constructions*. Chicago : The University of Chicago Press.
- Goossens, Vannina, Clémence Jacquot & Susanne Dyka. 2020. Science Fiction versus Fantasy: A Semantic Categorization and its Contribution to Distinguishing Two Literary Genres. In *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Iva Novakova & Dirk Siepmann, eds. Cham/London : Palgrave Macmillan. 189–221.
- Greimas, Algirdas Julien. 1982 [1972]. *Essais de sémiotique poétique*. Paris : Larousse.
- Guiraud, Pierre. 1954. *Les caractères statistiques du vocabulaire*. Paris : P.U.F.
- Gymnich, Marion, Birgit Neumann & Ansgar Nünning, eds. 2007. *Gattungstheorie und Gattungsgeschichte*. Trier : WVT.
- Hausmann, Franz Joseph. 1979. Un dictionnaire des collocations est-il possible ? *Travaux de littérature et de linguistique de l'université de Strasbourg* 17 (1), 187–195.
- Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. London/New York : Routledge.
- Kraif, Olivier. 2016. Le Lexicoscope : un outil d'extraction des séquences phraséologiques basé sur des corpus arborés. *Cahiers de lexicologie* 108, 91–106.
- Lala, Marie-Christine & Fanny Siaugues. 2020. Les questions du motif à la textualité littéraire au regard de nouvelles approches. In *Phraséologie et stylistique de la langue littéraire. Phraseology and stylistics of Literary Language. Approches interdisciplinaires. Interdisciplinary Approaches*. Ludwig Fesenmeier & Iva Novakova, eds. Bern : Peter Lang. 109–126.
- Leech, Geoffrey & Mick Short. 2007. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed. London : Pearson.
- Legallois, Dominique. 2006. Quand le texte signale sa structure : la fonction textuelle des noms sous-spécifiés. *Corela* 5. Available at : <https://journals.openedition.org/corela/pdf/1465>

- Legallois, Dominique. 2012. La Colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus* 11. Available at : <http://corpus.revues.org/2202>
- Legallois, Dominique & Agnès Tutin. 2013. Présentation : vers une extension du domaine de la phraséologie. *Langages* 189, 3–25.
- Legallois, Dominique & Stefan Koch. 2020. The Notion of Motif Where Disciplines Intersect: Folkloristics, Narrativity, Bioinformatics, Automatic Text Processing and Linguistics. In *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Iva Novakova & Dirk Siepmann, eds. Cham/London : Palgrave Macmillan. 17–46.
- Lits, Marc. 2011. *Le Roman policier dans tous ses états : d'Arsène Lupin à Navarro*. Limoges : Pulim.
- Longrée, Dominique & Sylvie Mellet. 2013. Le Motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue aux discours. *Langages* 189, 68–80.
- Magri-Mourgues, Véronique. 2006. Stylistique générique et statistique. *Les Cahiers de la MSH Ledoux* 8, 655–666.
- Mahlberg, Michaela. 2007. Clusters, Key Clusters and Local Textual Functions in Dickens. *Corpora* 2 (1), 1–31.
- Mahlberg, Michaela. 2013. *Corpus Stylistics and Dickens's Fiction*. London/New York : Routledge.
- Mellet, Sylvie & Dominique Longrée. 2012. Légitimité d'une unité textométrique : le motif. In *Actes des 11^e Journées internationales d'Analyse statistiques des Données Textuelles – JADT 2012*. Anne Dister, Dominique Longrée & Gerald Purnelle, eds. Liège, Belgique. 715–728.
- Marion, François. 2009. Le stéréotype dans le roman policier. *Cahiers de narratologie* 17. Available at : <http://narratologie.revues.org/1095>
- Martin, Robert. 1983. *Pour une logique du sens*. Paris : PUF.
- Mel'čuk, Igor, André Clas & Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Monte, Michèle & Gilles Philippe, eds. 2014. *Genres et textes : Déterminations, évolutions, confrontations*. Lyon : Presses universitaires de Lyon.
- Novakova, Iva & Dirk Siepmann, eds. 2020. *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Cham/London : Palgrave Macmillan.
- Novakova, Iva & Dirk Siepmann. 2020. Literary Style, Corpus Stylistic, and Lexico-Grammatical Narrative Patterns: Toward the Concept of Literary Motifs. In *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Iva Novakova & Dirk Siepmann, eds. Cham/London : Palgrave Macmillan. 1–15.

- Quiniou, Solen, Peggy Cellier, Thierry Charnois & Dominique Legallois. 2012. Fouille de données pour la stylistique : cas des motifs séquentiels émergents. In *Actes des 11es journées internationales d'analyse statistique des données textuelles*. Liège, 13–15 June 2012. 821–833.
- Philippe, Gilles & Julien Piat. 2009. *La Langue littéraire : une histoire de la prose en France de Gustave Flaubert à Claude Simon*. Paris : Fayard.
- Rastier, François. 2011. *La Mesure et le grain : sémantique de corpus*. Paris : Honoré Champion.
- Renouf, Antoinette & John Sinclair. 1991. Collocational Frameworks in English. In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. Karin Aijmer & Bengt Altenberg, eds. London : Longman. 128–144.
- Richard, Jean-Pierre. 1961. *L'Univers imaginaire de Mallarmé*. Paris : Seuil.
- Richard, Jean-Pierre. 1979. *Microlectures*. Paris : Seuil.
- Saint-Gelais, Richard. 1999. *L'Empire du pseudo. Modernités de la science-fiction*. Québec : Nota bene.
- Siepmann, Dirk. 2005. Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects. *International Journal of Lexicography* 18 (4), 409–444.
- Siepmann, Dirk. 2015. A Corpus-Based Investigation into Key Words and Key Patterns in Post-War Fiction. *Functions of Language* 22 (3), 362–399.
- Siepmann, Dirk. 2016. Lexicologie et phraséologie du roman contemporain : quelques pistes pour le français et l'anglais. *Cahiers de lexicologie* 108 (1), 21–41.
- Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. London : Routledge.
- Stefanowitsch, Anatol & Stefan Gries. 2003. Collocations: Investigating the Interaction between Words and Constructions. *International Journal of Corpus Linguistics* 8 (2), 209–243.
- Stubbs, Michael. 2005. Conrad in the Computer: Examples of Quantitative Stylistic Methods. *Language and Literature* 14 (1), 5–24.
- Stubbs, Michael & Isabel Barth. 2003. Using Recurrent Phrases as Text-Type Discriminators: A Quantitative Method and Some Findings. *Functions of Language* 10 (1), 61–104.
- Todorov, Tzvetan. 1980. *Poétique de la prose : choix, suivi de nouvelles recherches sur le récit*. Paris : Éditions du Seuil.
- Tutin, Agnès. 2010. *Sens et combinatoire lexicale: de la langue au discours*. Habilitation thesis, University of Grenoble. Available at: http://w3.u-grenoble3.fr/lidilem/labo/file/HDR_Tutin.pdf.
- Vaudrey-Luigi, Sandrine. 2013. *La langue romanesque de Marguerite Duras : Une liberté souvenante*. Paris : Garnier.
- Zymner, Rüdiger. 2003. *Gattungstheorie: Probleme und Positionen der Literaturwissenschaft*. Paderborn : Mentis.