

Deux motifs stylistiques dans *Les Mystères de Marseille* (1884) d'Émile Zola

Iva Novakova et Julie Sorba*

Université Grenoble Alpes
LIDILEM – EA609

*julie.sorba@univ-grenoble-alpes.fr

Résumé. L'objectif de cet article est de proposer une réflexion épistémologique et méthodologique sur le rôle du motif phraséologique pour caractériser le style singulier d'un auteur (Émile Zola) dans un texte précis (son roman *Les Mystères de Marseille*) par rapport à un corpus de contraste, composé de romans français du 19^e siècle. La question de recherche qui guide ce travail est de déterminer dans quelle mesure la méthodologie innovante appliquée permet d'isoler un ou des motifs spécifiques à un auteur dans un texte, et de là, de tester la pertinence de la notion de motif stylistique d'auteur d'un point de vue linguistique et stylistique. L'étude, qui s'appuie sur l'outil de fouille textuelle Lexicoscope 2.0, s'inscrit entièrement dans les approches corpus driven. Ainsi, en partant du motif phraséologique, deux motifs stylistiques d'auteur, spécifiques au roman *Les Mystères de Marseille* de Zola ont été identifiés et analysés : le motif de la transaction financière et le motif de l'émotion violente. Nous illustrons, enfin, le continuum posé entre les notions de motif phraséologique, de motif stylistique (marqueur de généricité ou d'auteur) et de motif textuel (propre à un texte d'un auteur vs ses autres œuvres), en obtenant ainsi une granularité, nécessaire aux objectifs de cette étude. La « méthode des motifs » apporte ainsi des éléments utiles pour l'analyse linguistique et stylistique des œuvres littéraires.

Abstract. This paper propose an epistemological and methodological reflection about the role of the phraseological motif to characterise the singular style of an author (Émile Zola) in a specific text (his novel *Les Mystères de Marseille*) in contrast with a corpus of 19th-century novels. The main aim of this work is to establish the impact of our innovative methodology for identifying one or more author-specific motifs in a text, and hence to test the relevance of the notion of authorial stylistic motif from a linguistic and stylistic point of view. The study, which is based on the computer tool Lexicoscope 2.0 for text mining, is fully in line with corpus-driven approaches. Thus, starting from the phraseological motif, two stylistic motifs specific to Zola's novel *Les Mystères de Marseille* were identified and analysed: the motif of the financial transaction and the motif of violent emotion. Finally, we propose a continuum between the notions of phraseological motif, stylistic motif (marker of genericity or authorship), and textual motif (specific to a text by an author vs. his other novels), thus obtaining the granularity necessary for the aim of the study. The 'motif method' thus provides useful elements for the linguistic and stylistic analysis of the literary fiction.

1 Introduction

L'objectif de cet article est de proposer une réflexion épistémologique et méthodologique sur le rôle du motif phraséologique pour caractériser le style singulier d'un auteur (Émile Zola) dans un texte précis (son roman *Les Mystères de Marseille*) par rapport à un corpus de romans du 19^e siècle. S'il a été établi, grâce aux méthodes de la linguistique de corpus, que les motifs en tant qu'unités linguistiques constituent des marqueurs de généricité à l'échelle de la phrase ou d'une séquence textuelle, révélant des stéréotypes langagiers généraux qui sous-tendent les sous-genres romanesques contemporains¹, cela n'a pas été le cas pour les motifs stylistiques spécifiques d'un auteur ou d'un texte. Ces derniers sont détectés à l'aide d'un outil, le Lexicoscope 2.0 (Kraif 2019), destiné à la fouille automatique des textes. La question de recherche qui guide ce travail est de déterminer dans quelle mesure la méthodologie innovante appliquée ici permet d'isoler un ou des motifs spécifiques à un auteur dans un texte, et de là, de tester la pertinence de la notion de motif stylistique d'auteur d'un point de vue linguistique et stylistique.

Le roman de Zola, *Les Mystères de Marseille*, « occupe une place particulière » dans l'œuvre de l'écrivain comme l'indique Compère dans son introduction à l'édition de 2018 (p. 7). En effet, la première raison de

cette particularité tiendrait à la place de cette œuvre dans la production romanesque de Zola, sa première version ayant été publiée en 1867, avant le cycle des Rougon-Macquart ; la seconde raison viendrait du fait que « l’auteur lui-même la considère comme mauvaise » (*ibid.*) car il s’agit d’« un travail strictement alimentaire » (Spandonis 2004, p. 97). En effet, il est le fruit d’une commande passée par Léopold Arnaud, le directeur du *Messenger de Provence*, avocat de métier, qui fournit à Zola les documents judiciaires utiles pour décrire le milieu marseillais de l’époque où se déroule l’intrigue de ce roman-feuilleton. Zola définit lui-même *Les Mystères de Marseille* comme « un roman historique contemporain [...] un ouvrage où tout est vrai, où tout a été observé sur nature »². L’intrigue, qui se déroule principalement à Marseille entre 1843 et 1849, raconte les amours malheureuses de Philippe Cayol, un républicain démuné, et de Blanche de Cazalis, la nièce orpheline d’un riche député bien installé, et surtout la tentative de Marius Cayol de sauver son frère du pilori auquel il a été condamné, en plus des cinq ans de prison, pour l’enlèvement de la jeune fille. À la demande de Zola, le roman est réédité en un volume en octobre 1884 chez l’éditeur Georges Charpentier, accompagné d’une préface rédigée pour l’occasion et d’un nombre important de variantes par rapport à la première parution de 1867. C’est sur ce texte remanié que s’appuie notre étude linguistique.

Après une brève présentation de la notion de motif phraséologique (section 2), nous décrirons notre méthodologie, nos corpus et les données sur lesquelles s’appuie ce travail (section 3). La section 4 sera consacrée à l’analyse de deux motifs stylistiques dans le roman de Zola, identifiés grâce nos méthodes lexico-statistiques, à savoir le motif de la transaction financière (section 4.1) et le motif de l’émotion violente (section 4.2). La section 5 propose une synthèse de nos résultats, ainsi que quelques perspectives pouvant compléter utilement ce travail.

2 La notion de motif phraséologique

Les motifs en « phraséologie étendue » (Legallois & Tutin 2013) sont conçus comme des unités multidimensionnelles d’appariement entre forme et sens qui ont une fonction pragmatique et discursive (Legallois 2012, p. 45). Dans ce cadre, le motif n’est pas un élément fictionnel symbolique ou constitutif de l’imaginaire d’une œuvre (Bellemin-Noël 1972 ; Richard 1979), mais un observable phraséologique, composé d’unités continues ou discontinues combinant lemmes, catégories morphosyntaxiques, mots-outils, etc. Il s’agit ainsi d’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, p. 66). Les motifs phraséologiques, détectés grâce à des méthodes statistiques (cf. section 3.1) correspondent à des « patrons lexico-syntaxiques récurrents, qui présentent des régularités et des variations lexicales et syntaxiques aux niveaux paradigmatique et syntagmatique, associées à des fonctions discursives spécifiques » (Sorba & Novakova 2022, § 9). Dans la fiction romanesque, ces fonctions discursives (FD) assurent « la cohérence textuelle » (Martin 1983, p. 100). À côté des deux FD *narrative* et *descriptive* théorisées par Adam (2011, p. 267), nous nous appuyons sur une liste enrichie de FD, établie dans le cadre du projet PhraseoRom (Diwersy et al. 2021), à savoir :

- *infranarrative* : renvoie à des actions minimales d’arrière-plan, sans rapport direct avec l’intrigue principale ;
- *infradescriptive* : s’applique à une action ou à un geste répété, caractérisant le personnage ;
- *affective* : décrit les affects, les sentiments, les émotions des personnages ;
- *pragmatique* : se rapporte aux actes de langage entre les personnages ;
- *cognitive* : implique des processus cognitifs comme les hypothèses, l’appréhension des événements, les réflexions.

Conformément à l’objectif de ce travail, en partant du motif phraséologique, nous testerons la notion de motif stylistique d’auteur (Kraif, Novakova & Sorba 2023b) tout en l’inscrivant dans une dimension collective du style d’une époque³. En effet, à l’instar de Philippe (2021), nous déclinons la définition standard du style⁴ comme « signature singulière d’une personnalité singulière » (p. 6) pour faire valoir « le “régime d’historicité stylistique” des textes, c’est-à-dire la façon dont ils se situent dans le mouvement évolutif. » (p. 15). Il s’agit ainsi de s’écarter de la conception « auteuriste » du style⁵ pour « penser les

œuvres comme des productions ne se limitant pas à un complexe formel et mettant en jeu bien plus qu'un geste individuel » (p. 23) et considérer que tout texte littéraire est « polychronique »⁶ (p. 215). Notre méthodologie permet de détecter non pas un fait langagier isolé perçu comme anormal mais une anomalie statistique récurrente comme nous allons le voir dans la section suivante.

3 Méthodologie, corpus et données

Dans cette section, nous présenterons tout d'abord la méthodologie mise en œuvre pour la fouille de corpus textuels (section 3.1), puis les deux corpus de notre étude (section 3.2) et les données extraites automatiquement (section 3.3) à partir desquelles nous avons construit nos résultats.

3.1 Méthodologie : la fouille outillée de corpus textuels

Les données de notre étude sont extraites d'un corpus textuel au moyen de l'outil de fouille Lexicoscope 2.0. Celui-ci a été initialement conçu pour l'analyse de corpus syntaxiquement arborés en dépendances syntaxiques⁷ afin d'extraire des données sur la combinatoire lexico-syntaxique des unités lexicales (Kraif 2019). Ces informations se présentent sous la forme d'un lexicogramme⁸ regroupant les collocatifs statistiquement représentatifs du mot ou de l'expression pivot faisant l'objet de la requête. La saillance statistique d'un collocatif est repérée sur la base de mesures d'associations qui permettent de distinguer les cooccurrences fortuites des cooccurrences récurrentes. C'est le rapport de vraisemblance (*Log Likelihood Ratio* ou LLR, Dunning 1993) qui est ici utilisé. La figure 1 permet de visualiser sous la forme d'un lexicogramme les cooccurents statistiquement significatifs⁹ de la lexie *franc_NOUN* dans *Les Mystères de Marseille* :

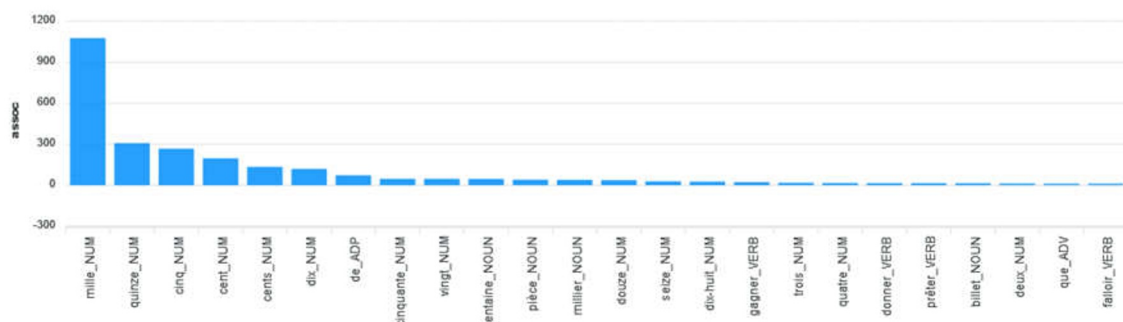


Figure 1. Lexicogramme de la lexie *franc* dans *Les Mystères de Marseille*.

Dans la figure 1, on observe que le collocatif le plus fréquent est le numéral *mille* (LLR 1076,2), suivi par d'autres numéraux comme *quinze* (LLR 608,51), *cinq* (LLR 266,73), *cent* (LLR 196,98), etc.

Dans sa version la plus récente, l'outil Lexicoscope 2.0 permet de comparer deux corpus afin d'en extraire les spécificités cooccurentielles de manière contrastive. En effet, l'architecture du Lexicoscope 2.0 est conçue pour faire émerger des contrastes entre différents sous-corpus. Ceux-ci peuvent notamment apparaître sous la forme d'un déséquilibre dans la répartition des fréquences, calculable grâce à une mesure de spécificité telle que la *keyness* issue du rapport de vraisemblance (Bertels & Speelman 2013). Cette approche contrastive, consistant à faire émerger des anomalies statistiques pour ensuite, dans un second temps, dégager des pistes interprétatives s'inscrit typiquement dans une démarche *corpus driven*.

3.2 Les deux corpus de l'étude

Les deux corpus qui ont été comparés dans le cadre de notre étude sont (i) *Les Mystères de Marseille* (1884) et (ii) PhraseoRom_19^e, un corpus constitué de 67 romans français parus entre 1830 et 1899¹⁰. Le tableau 1 ci-dessous récapitule les caractéristiques de ces deux corpus :

Tableau 1. Présentation des deux corpus de l'étude.

	Nombre de tokens	Nombre de textes	Nombre d'auteur
<i>Les Mystères de Marseille</i> (1884)	157 768	1	1 (Émile Zola)
PhraseoRom_19 ^e (1830-1899)	8 380 969	67	43 (Allais, Mérimée, Sand, Balzac, etc.)

Notons que si le déséquilibre de taille entre les deux corpus n'entrave en rien l'extraction des spécificités cooccurentielles des *Mystères de Marseille* comparé à PhraseoRom_19^e, il est impossible en revanche d'obtenir des données statistiquement significatives pour la comparaison inverse.

3.3 Présentation des données

La comparaison de ces deux corpus a permis d'extraire 279 cooccurrences statistiquement significatives dans *Les Mystères de Marseille* comparé à PhraseoRom_19^e. Nous avons choisi comme échantillon de travail pour cette étude les 150 premières cooccurrences (LLR > 15,74). Sans surprise, nous trouvons parmi les 15 premières (LLR > 100), les noms des personnages, aussi bien sous la forme de noms propres (*M. Cazalis* rang 1, *M. Martelly* rang 2, *abbé Chastanier* rang 3, *M. Girousse* rang 4, etc.) que de substantifs désignant leur fonction (*maître portefaix* rang 6, *garde national* rang 7, *ancien député* rang 15). Sur les 150 cooccurrences, les noms des personnages en occupent 36 (soit 24 %). Les toponymes occupent aussi une bonne place dans la liste avec 24 occurrences (soit 16 % ; *cours Bonaparte*, *Saint Barnabé*, *Place aux Œufs*, *rue de Rome*, *Saint Ferréol*, *Canebière*, etc.), ce qui permet à Zola d'ancrer son intrigue dans le décor singulier du Marseille des années 1840. La spécificité élevée des anthroponymes et des toponymes s'explique aisément par l'univers référentiel propre au roman de Zola et par la technique de construction de son univers romanesque¹¹.

En écartant ces cooccurrences matérialisant l'univers référentiel propre au roman de Zola, nous avons constaté deux saillances particulières de cooccurrences autour de l'argent et de l'expression des affects (voir tableau 2).

Tableau 2. Les cooccurrences saillantes autour de l'argent et de l'expression des affects

Thématique de l'argent	Thématique des affects
<i>mille franc</i> (LLR 132,42), <i>quinze franc</i> (LLR 83,15), <i>cinq franc</i> (LLR 42,67), <i>pièce or</i> (LLR 41,83), <i>faux billet</i> (LLR 36,65), <i>centaine franc</i> (LLR 23,03), <i>dix franc</i> (LLR 22,76), <i>acquérir immeuble</i> (LLR 20,61), <i>gérer fortune</i> (LLR 20,61), <i>empocher argent</i> (LLR 19,09), <i>emprunter argent</i> (LLR 17,01), <i>prêter argent</i> (LLR 17,01), <i>somme nécessaire</i> (LLR 16,52), <i>millier franc</i> (LLR 16,05)	<i>émotion poignant</i> (LLR 29,02), <i>peur atroce</i> (LLR 26,54), <i>âpre désir</i> (LLR 25,77), <i>serrer gorge</i> (LLR 22,78), <i>éclater sanglot</i> (LLR 22,58), <i>goûter joie</i> (LLR 17,8)

Comme le montrent les données du tableau 2, on comptabilise 14 expressions liées à la thématique de l'argent (sur les 150 expressions retenues pour cette étude), soit 9,3 % contre seulement 6 expressions

renvoyant à la thématique des affects (soit 4 %). Ce sont ces deux pistes que nous allons suivre à la recherche de motifs stylistiques dans le roman *Les Mystères de Marseille*.

4 Deux motifs stylistiques saillants

L'analyse de ces données statistiques a permis d'identifier deux motifs stylistiques d'auteur, spécifiques au roman de Zola : le motif de la transaction financière et le motif de l'émotion violente, que nous allons analyser respectivement dans les sections 4.1 et 4.2.

4.1 Autour de l'argent : le motif de la transaction financière

La folie de l'argent est un thème particulièrement présent dans *Les Mystères de Marseille* (Compère 2018, p. 33). Pour Spandonis (2004, p. 101), le roman se transforme progressivement en « dévoilement des misères et bassesse de la société contemporaine [...] : prostitutions, vol, usure, escroqueries en tous genres, fatale passion du jeu, concupiscence, etc. ». Zola lui-même intervient directement en tant que narrateur, au chapitre XIII, pour appuyer ce point :

Il est à remarquer que la passion du jeu désole surtout les grands centres de commerce. Lorsqu'une population entière est livrée à une spéculation effrénée, lorsque toute les classes d'une ville trafiquent du matin au soir, il est presque impossible que ce peuple de négociants ne se jette pas dans les émotions poignantes du jeu. Le jeu devient alors une spéculation qui s'ajoute aux autres ; on spéculé sur le hasard, on continue la nuit la besogne du jour ; pendant le jour on a tâché d'augmenter sa fortune en vendant de n'importe quoi, et, pendant la nuit, on tâche d'augmenter le gain en le hasardant sur le tapis vert. (p. 227).

C'est d'ailleurs, selon Guedj (1968), la caractéristique du monde dépeint par Zola : « le milieu le plus typique n'est pas l'usine ou le comptoir du négociant, mais le tripot. » (1968, p. 18-19) et c'est dans ce cadre que Marius Cayol cherche désespérément à se procurer les quinze mille francs nécessaires pour corrompre le geôlier et faire évader son frère.

Parmi les cooccurrences les plus fréquentes dans le roman, le substantif *franc* occupe les premières places et revient souvent comme le montre le tableau 2. L'analyse de la combinatoire de cette lexie dans le roman révèle que son collocatif le plus fréquent est, de manière attendue, un numéral (16/24 collocatifs soit 66 %). En outre, la particularité de ces collocatifs numéraux est d'exprimer un grand nombre : en plus d'être le collocatif le plus spécifique de *franc*, *mille* est le collocatif privilégié des autres cooccurrences comprenant un numéral : *quinze mille francs*, *cinq mille francs*, *dix mille francs*, *centaine de mille francs*. On observe ainsi une variation paradigmatique des déterminants numéraux et une prédilection pour l'expression des grands nombres. Néanmoins, l'abondance de lexies dénotant le comptage de l'argent n'est pas spécifique à ce roman de Zola. En effet, dans les romans du 19^e siècle, on mesure plutôt de l'argent que des distances : par exemple, les unités monétaires (*franc*, *sou*, *livre*, *pièce*, *billet*) constituent la majorité des occurrences des lexies désignant une unité de mesure (Kraif, Novakova & Sorba 2023a). Ainsi, le « régime d'historicité stylistique » de ce roman de Zola l'inscrirait dans une tendance linguistique de fond de l'époque.

Si l'on observe maintenant, dans le roman, le profil combinatoire (ou *wordsketch*) de *franc* avec les verbes, on note la présence de 10 collocatifs verbaux attestés au moins dans 2 occurrences :

Tableau 3. Wordsketch de *franc* : les verbes dont le pivot est objet.

	fréquence	LLR
donner VERB	8	30.61
voler VERB	2	30.39
rendre VERB	2	30.39
emprunter VERB	2	30.39
contenir VERB	2	30.39
posséder VERB	2	30.39
dépenser VERB	2	30.39
payer VERB	2	30.39
gagner VERB	6	27.91
prêter VERB	4	24.65

Dans leur grande majorité, ces verbes impliquent une transaction financière, que celle-ci soit honnête ou non (*donner, voler, rendre, emprunter, dépenser, payer, gagner, prêter*). En analysant les occurrences des collocations verbo-nominales dont la fréquence est supérieure à 2 (*donner, gagner, prêter*), on s'aperçoit que les variations syntagmatiques sont bien présentes en personne et en temps :

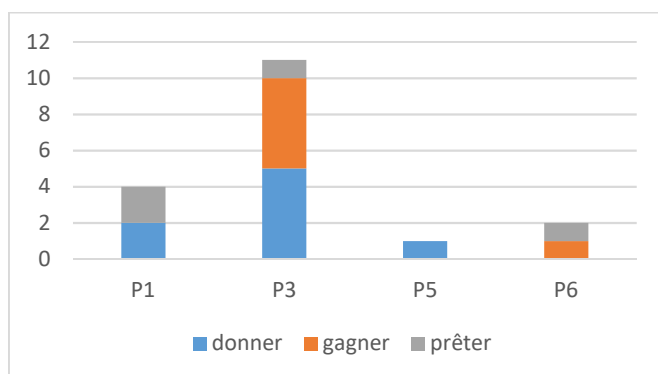


Figure 2. Variations syntagmatiques en personne pour les 3 collocations verbo-nominales.

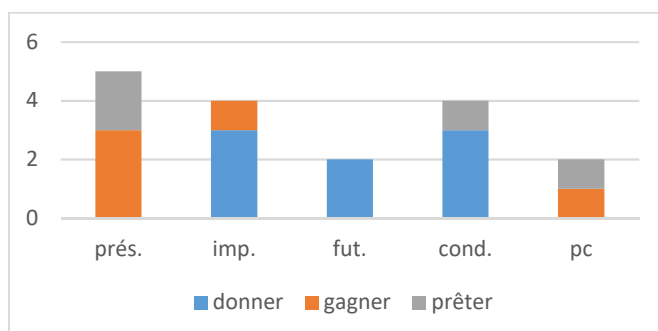


Figure 3. Variations syntagmatiques en temps pour les 3 collocations verbo-nominales.

En observant particulièrement ces 18 occurrences, on constate que les valeurs modales sont très présentes. En effet, celles-ci peuvent être marquées par les tiroirs verbaux (futur, conditionnel, pour le potentiel et l'irréel, ex. 1, 3, 4), par l'emploi de l'auxiliaire modal (ex. 2) ou encore par des tournures particulières (*il*

avait souvent entendu parler de..., ex. 5) et font finalement peser une grande incertitude sur l'issue de la transaction financière :

- 1) Vous viendrez plus tard à mon secours, **vous** me **donnerez** les quinze mille francs sou à sou, quand vous pourrez.
- 2) – Ah ! dit-il d'une voix étrange, **on peut gagner** cinq mille francs dans une nuit ?
- 3) Dans ce moment-là, **il aurait** volontiers **prêté** mille francs au jeune homme.
- 4) S'il les avait eus, **il aurait donné** volontiers les mille francs que demandait l'usurière.
- 5) Il savait bien qu'il y avait dans Marseille, comme dans tous les grands centres d'industrie, des employés qui volent leurs patrons pour satisfaire leurs vices et leur amour du luxe ; il avait souvent entendu parler de **ces commis qui gagnent** cent ou cent cinquante francs par mois, et qui trouvent moyen de perdre dans les cercles des sommes énormes, de jeter des pièces de vingt francs aux filles, de vivre dans les restaurants et les cafés.

Ces exemples illustrent également des variations paradigmatiques autour du « cœur du motif » avec la présence ou non du complément indirect du verbe (ex. 1, 3, 4), d'un déterminant devant le montant en francs ou d'un circonstant (*volontiers, par mois, dans une nuit, sou à sou*). On peut ainsi modéliser le patron suivant : X (*donner, gagner, prêter*, etc.) (DET) numéral *franc* (à Y) (circonstant). Ce patron permet de structurer la progression narrative de la quête de Marius Cayol (ex. 1, 2, 3, 4) ou d'insérer des séquences descriptives sur le milieu des tripots dans lequel cette quête va le mener (ex. 5). Les deux FD occupées sont ainsi celles classiquement repérées dans le roman : narrative et descriptive. Ce motif phraséologique de la transaction financière apparaît ainsi comme un motif stylistique d'auteur utilisé par Zola dans *Les Mystères de Marseille*.

4.2 Autour des affects : le motif de l'émotion violente

Le deuxième motif émergent est lié à l'expression des affects des personnages et ce, à travers 3 collocations nominales qui suivent le patron N + ADJ (*émotion poignante, peur atroce, âpre désir*) et 3 collocations verbo-nominales (*serrer à la gorge, éclater en sanglot, goûter la joie*) qui en constituent le « cadre collocationnel » (cf. section 2). Comme indiqué *supra*, ces collocations ne constituent que 4 % de la totalité des 150 cooccurrences de l'étude. Elles sont donc plus de 2 fois moins nombreuses que celles autour de l'argent (9,3 %). Ce résultat est un indice de la place accordée dans ce roman à la vénalité, comparé aux sentiments des personnages.

D'un point de vue linguistique, dans les collocations du type N + ADJ étudiées, l'adjectif¹² modifieur renvoie à des émotions intenses, de polarité négative. C'est tout à fait conforme à l'observation de Plantin (2012, p. 632) qui pose la polarité et l'intensité comme deux composantes essentielles de la séquence émotionnelle. C'est la raison pour laquelle nous avons choisi d'appeler le motif qu'elles constituent « le motif de l'émotion violente ».

L'analyse des données montre que ces affects sont surtout provoqués par l'addiction aux jeux de hasard provoquant une *émotion poignante* (ex. 6), le danger de devoir rendre l'argent extorqué entraînant une *peur atroce* (ex. 7) ou des malversations sordides guidées par un *âpre désir* de s'enrichir (ex. 8) :

- 6) Quand il ouvrit les yeux, il aperçut sur sa commode l'argent qu'il avait gagné. Les reflets fauves qui couraient sur les pièces d'or l'effrayèrent presque ; tout d'un coup, il se rappela avec une netteté étrange la nuit qu'il avait passée ; **et une émotion poignante le prit à la gorge**. Il eut peur d'être devenu joueur, car sa première pensée, au réveil, avait été qu'il retournerait le soir au tripot et qu'il gagnerait encore.
- 7) Lors de la fuite de sa nièce avec Philippe, **il eut une peur atroce d'être obligé de rendre ses comptes de tutelle**, car il serait tombé dans une véritable misère, si on lui avait retiré cet argent des mains. Depuis plusieurs mois, il ne vivait plus que sur le bien de sa nièce.
- 8) Ce sont de petits employés qui ont des appétits de luxe et de débauche [...]. Ils voient autour d'eux les gens riches se vautrer dans les jouissances, avoir des maîtresses s'étaler dans des voitures, épuiser les joies bruyantes de la vie, une jalousie les prend, **ils ont l'âpre désir de mener une pareille existence de fêtes et de plaisirs**. Alors, pour se procurer de l'argent, ils jouent ; ils jouent d'abord leurs appointments ; puis, quand la chance leur est contraire, ils volent leurs patrons, ils entrent dans le crime.

Par ailleurs, les relations amoureuses sont source d'*émotion poignante*, lorsque Blanche, amoureuse de Philippe, s'enfuit avec celui-ci de la maison où elle habite avec son oncle (ex. 9) :

9) Alors, fermant les yeux, s'abandonnant, Blanche descendit le coteau à grands pas, au bras de Philippe. Comme elle s'éloignait, elle regarda une dernière fois la maison qu'elle quittait, et une émotion poignante lui mit de grosses larmes dans les yeux.

Enfin, l'*émotion poignante* peut être provoquée par un souvenir d'enfance (ex. 10) ou des actes violents dans la maison assiégée par les soldats de la garde républicaine (ex. 11) :

10) Lorsque Marius passa devant la maison de campagne où sa mère l'avait bercé, une émotion poignante lui mit de grosses larmes dans les yeux. Au milieu du silence de ce désert morne et brûlé, il croyait entendre la voix aimée de la sainte femme dont le souvenir le soutenait dans la tâche de délivrance qui l'accablait.

11) Ils montèrent au premier étage, et virent avec angoisse, par la fenêtre du palier, que la maison assiégée était celle d'en face. Pendant quelques minutes, ils restèrent immobiles et haletants : chaque coup de hache trouvait un écho dans leur poitrine. Jamais une émotion si poignante ne les avait serrés à la gorge. Ils suivaient les diverses phases du siège avec une anxiété douloureuse.

Si on regarde maintenant de plus près les distributions dans lesquelles se trouve le motif de l'émotion violente que génèrent ces collocations, deux principales tendances se dégagent :

(i) Le SN *émotion poignante* (ex. 6, 9, 10, 11) est en tête de phrase en fonction de sujet syntaxique. Une variation paradigmatique est observée sur le verbe dans : *une émotion poignante le prit à la gorge ; les avait serrés à la gorge, lui mit des larmes dans les yeux.* On retrouve aussi en position de sujet syntaxique la collocation *âpre désir* (ex. 12, 13) :

12) **Un désir âpre** avait dû emporter le gracieux abbé et lui faire oublier la prudence sournoise qui le guidait dans tous ses actes.

13) Au fond, **son plus âpre désir** était de retrouver l'enfant de sa nièce et de s'en emparer. Alors seulement, il pourrait disposer de la fortune qui dormait entre ses mains.

Cette distribution est présente dans la moitié des cas (6/12 occurrences). L'effet obtenu est celui d'un affect mis « sous projecteur » (Leeman 1987, p. 246), d'une mise en relief de l'émotion. Se trouvant le plus souvent inséré dans des séquences descriptives, ce motif, qui renvoie au ressenti des personnages, assume une fonction discursive affective.

(ii) La seconde distribution saillante repérée au sein de ce motif est l'extension syntagmatique de la collocation « cœur du motif » par des compléments du nom, sous forme de construction infinitivale (Vinf). C'est surtout le cas de *peur atroce* ou *âpre désir*, où l'extension explique la cause ayant provoqué l'affect (ex. 7, 8) : *une peur atroce d'être obligé de rendre ses comptes de tutelle ; l'âpre désir de mener une pareille existence de fêtes et de plaisirs.* Cette complémentation sert de justification de l'émotion provoquée et permet de lancer le « travail textuel d'étayage » (Jacquin & Micheli 2012, p. 603). Ce motif de l'émotion violente qui émaille les séquences descriptives assume une fois de plus une FD affective, en support de la description.

Par ailleurs, l'analyse des occurrences des 2 collocations verbo-nominales statistiquement significatives *serrer à la gorge* et *goûter la joie*¹³ révèle des configurations assez similaires à celles décrites précédemment pour les collocations nominales. De manière récurrente, celles-ci attirent des noms d'affect (*émotion, angoisse, amertume*) comme sujet de la phrase. Généralement, le motif apparaît en contexte amoureux (ex. 14) ou empathique (ex. 15), de jeux de hasard (ex. 16), ou de différentes péripéties (ex. 11) :

14) Une amertume suprême la serrait à la gorge, lorsqu'elle songeait que quelques mois avaient suffi pour la mener, des espoirs de bonheur qui chantaient dans le cœur de toutes les jeunes filles, aux sombres pensées de remords qui emplissent l'âme des pénitentes.

15) La jeune sœur avait joint les mains. Elle regardait Marius d'une façon humble et suppliante, sans pouvoir répondre, tant l'angoisse la serrait à la gorge.

16) Toute sa fièvre tomba, l'émotion ne le serra plus à la gorge. L'argent le brûlait, lorsqu'il le touchait ; il aurait voulu achever de perdre cet argent et se retirer les poches vides.

L'émotion exprimée par le motif est violente, provoquant une réaction physique chez le personnage. Enfin, la collocation *goûter une joie* présente systématiquement une extension adjectivale (une joie *cruelle, pure, sombre, amère*), ainsi qu'un complément du N, sous forme de Vinf, spécifiant la cause de l'affect. Une variation paradigmatique sur le déterminant est également observée (*une/la joie*). En voici deux exemples :

17) Alors il étouffait Blanche d'une caresse plus rude. Il avait fini par **goûter une joie amère** à la faire courir dans les pierres des chemins.

18) La colère aveugle de M. de Cazalis était tombée. Il tenait sa vengeance. Il allait pouvoir prouver sa toute-puissance en écrasant un de ces républicains qu'il détestait. Maintenant, il ne désirait plus que **goûter la joie cruelle** de jouer avec sa proie.

Comme le montrent les exemples 17 et 18, la joie est souvent de polarité négative (*amère*) ou malsaine (*cruelle*). La fonction affective du motif vient, une fois de plus, s'imbriquer dans la FD descriptive des séquences. Dans son commentaire du passage de l'exemple 17, Guedj (1968, p. 16) souligne que « cet érotisme tragique, teinté de sadisme », qui est celui de Philippe Cayol, se retrouve dans d'autres œuvres de Zola comme *La Terre*.

En bref, les données statistiques ont permis d'identifier, à partir des collocations binaires extraites comme hautement spécifiques, des unités polylexicales formant le motif de l'émotion violente. Celui-ci présente des fonctions discursives analogues (descriptive et affective), ainsi que deux distributions saillantes : la position sujet et la complémentation sous forme de Vinf. Ces deux distributions permettent de mettre en relief (« effet de projecteur ») et de justifier les affects des personnages (« travail textuel d'étayage »). On observe aussi des variations paradigmatiques (sur le verbe ou le déterminant) au sein du motif. Il s'agit donc d'un motif stylistique propre à Zola (comparé au corpus de contraste de romans du 19^e s.).

5 Conclusion

En partant du motif phraséologique, nous avons identifié dans ce travail deux motifs stylistiques d'auteur, spécifiques au roman *les Mystères de Marseille* de Zola : le motif de la transaction financière et le motif de l'émotion violente. Les particularités de ces motifs ont été mises au jour grâce au Lexicoscope 2.0, un outil de fouille automatique des textes mettant en œuvre des méthodes statistiques, propres à la linguistique de corpus. Les approches inductives appliquées ici viennent appuyer et enrichir les observations faites par les analyses des stylisticiens.

Dans un travail récent (Kraif, Novakova & Sorba 2023b), un continuum a été établi entre le motif phraséologique, le motif stylistique, que celui-ci soit marqueur de généricité ou d'auteur, et le motif textuel (propre à un texte d'un auteur vs ses autres œuvres). Pour compléter notre réflexion épistémologique sur la notion de motif, et notamment sur celle de motif textuel, il serait pertinent de comparer le « rendement stylistique » (Philippe 2021, p. 123) du motif de la transaction financière dans les autres romans de Zola où la thématique de l'argent occupe une place centrale comme *La Curée*, *L'Argent*, *Au bonheur des dames*.

Plus généralement, la « méthode des motifs », applicable dans la fouille des textes (Lala & Siaugues 2020, p. 113 ; Legallois & Koch 2020, p. 32-41), apporte des éléments utiles à l'analyse linguistique et stylistique des œuvres littéraires. Le motif phraséologique, qui peut se réaliser en motif stylistique ou textuel, assure ainsi le lien entre le micro-niveau, celui des récurrences phraséologiques spécifiques, et le macro-niveau, celui du script narratif ou fictionnel (Novakova et Siepmann 2020, p. 10).

Références bibliographiques

- Adam, J.-M. (2011). *Les textes : types et prototypes*. Paris : Armand Colin.
- Bartsch, S. & Evert, S. (2014). Towards a Firthian notion of collocation. In A. Abel & L. Lemnitzer (dir.), *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern*, 2, 48-61.
- Bellemin-Noël, J. (1972). Le motif des orangers dans *La Chartreuse de Parme*. *Littérature*, 5, 26-33.
- Bertels, A. & Speelmann, D. (2013). 'Keywords Method' versus 'Calcul des Spécificités'. A comparison of tools and methods. *International Journal of Corpus Linguistics*, 18(4), 536-560.
- Brunet, É. (2009). On a compté trois millions de mots chez Zola. Et alors ? In D. Mayaffre (dir.). *Comptes d'auteurs. Tome I. Études statistiques de Rabelais à Gracq*. Paris : Honoré Champion, 151-176.
- Compère, D. (2018) : voir Zola.

- Diwersy, S., Gonon, L., Goossens, V., Kraif, O., Novakova, I., Sorba, J. & Vidotto, I. (2021). La phraséologie du roman contemporain dans les corpus et les applications de la PhraseoBase. *Corpus*, 22. <https://doi.org/10.4000/corpus.6101>
- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence, *Computational Linguistics*, 19(1), 61-74.
- Gonon, L., Goossens, V. & Novakova, I. (2020). Les phraséologismes spécifiques à deux sous-genres de la paralittérature : le roman sentimental et le roman policier. In S. Mejri, L. Meneses-Lerin & B. Buffard-Moret (dir.). *La phraséologie française en questions ?* Paris : Hermann, 212-225.
- Gonon, L. & Sorba, J. (2019). Phraséologismes spécifiques dans les romans historiques et dans les romans de littérature blanche. *Journal of French Language Studies*, 30(1), 1-20.
- Guedj, A. (1968). *Les Mystères de Marseille* ou l'acte de naissance du naturalisme. *Les cahiers naturalistes*, 35, 1-19.
- Hamon, P. (1983). *Le Personnel de roman. Le système des personnages dans les Rougon-Macquart d'Émile Zola*. Genève : Droz.
- Heiden, S. & Tournier, M. (2001). Lexicométrie textuelle, sens et stratégie discursive. In J. J. de Bustos Tovar (dir.). *Simposio internacional de análisis del discurso*. Visor, 2287-2300.
- Jacquin, J. & Micheli, R. (2012). Entre texte et interaction : propositions méthodologiques pour une approche discursive de l'argumentation en sciences du langage. *Actes du CMLF 2012 - 3^e Congrès Mondial de Linguistique Française*, EDP Sciences, 599-610.
- Kraif, O. (2019). Explorer la combinatoire lexico-syntaxique des mots et expressions avec le Lexicoscope. *Langue française*, 203, 67-83.
- Kraif, O. & Diwersy, S. (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques. *Actes de la conférence conjointe JAP-TALN-RECITAL 2012*, volume 2 : TALN ATALA & AFCEP, 399-406.
- Kraif, O., Novakova, I. & Sorba, J. (2023a). Littérature d'anticipation et science-fiction : quelle filiation ? Un éclairage phraséologique sur corpus. Communication présentée au Colloque international *Littérature d'anticipation et humanités numériques*. ENS-Lyon, 16 et 17 mars 2023.
- Kraif, O., Novakova, I. & Sorba, J. (2023b). Motif phraséologique vs motif textuel : réflexion conceptuelle et méthodologique sur corpus littéraire contemporain. Communication présentée aux Journées d'étude *Identifier les régularités stylistiques dans le discours littéraire*. Paris, Université Sorbonne nouvelle, 30 novembre et 1^{er} décembre 2023.
- Lala, M.-C. & Siaugues F. (2020). Les questions du motif à la textualité littéraire au regard de nouvelles approches. In L. Fesenmeier & I. Novakova (dir.). *Phraséologie et stylistique de la langue littéraire. Phraseology and stylistics of Literary Language. Approches interdisciplinaires. Interdisciplinary Approaches*. Berne : Peter Lang, 109-126.
- Leeman, D. (1987). À ma grande surprise... *Revue québécoise de linguistique*, 16(2), 225-265.
- Legallois, D. (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus*, 11, 31-54.
- Legallois, D. & Koch, S. (2020). The Notion of Motif Where Disciplines Intersect: Folkloristics, Narrativity, Bioinformatics, Automatic Text Processing and Linguistics. In I. Novakova & D. Siepmann (dir.). *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Londres : Palgrave Macmillan, 17-46.
- Legallois, D. & Tutin, A. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, 189, 3-25.
- Lexicoscope 2.0 http://phraseotext.univ-grenoble-alpes.fr/lexicoscope_2.0/
- Longrée, D. & Mellet, S. (2013). Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours. *Langages*, 189, 68-80.
- Martin, R. (1983). *Pour une logique du sens*. Paris : PUF.
- Novakova, I. & Siepmann, D. (dir.) (2020). *Phraseology and Style in Subgenres of the Novel: A Synthesis of Corpus and Literary Perspectives*. Londres : Palgrave Macmillan.
- Novakova, I. (2023). Comment identifier et définir les genres littéraires à travers la phraséologie étendue ? In S. Bédouret-Larraburu, C. Copy & R. Nita (dir.). *Lexique et frontières des genres*. Pau : PUPPA, 379-394.
- Philippe, G. (2021). *Pourquoi le style change-t-il ?* Bruxelles : Les Impressions Nouvelles.

- Plantin, C. (2012). Les séquences discursives émotionnées : définition et application à des données tirées de la base CLAPI. *Actes du CMLF 2012 - 3^e Congrès Mondial de Linguistique Française*, EDP Sciences, 629-642. <https://doi.org/10.1051/shsconf/20120100218>
- Richard, J.-P. (1979). *Microlectures*. Paris : Seuil.
- Sorba, J. (2022). *Phraséologie et genres textuels. Perspectives synchroniques et diachroniques*. Mémoire de synthèse présenté pour l'habilitation à diriger des recherches. Université Grenoble Alpes. <https://hal.archives-ouvertes.fr/tel-03891983v1>
- Sorba, J. & Novakova, I. (2022). Le dragon dans le roman contemporain : un motif pour distinguer les sous-genres romanesques ? *Iris*, 42 DOI : [10.35562/iris.2922](https://doi.org/10.35562/iris.2922)
- Spandonis, S. (2004). Roman feuilleton et poétique naturaliste : une étude des *Mystères de Marseille*. In J.-P. Leduc-Adine & H. Mitterand (dir.). *Lire / Dé-lire Zola*. Nouveau Monde édition, 97-115.
- TLFi <http://atilf.atilf.fr/tlf.htm>
- Zola, É. (2018). *Les Mystères de Marseille*. Édition de Daniel Compère. Paris : Classiques Garnier.

¹ Cf. Novakova & Siepmann (2020) et les travaux réalisés dans le cadre et dans la continuité du projet franco-allemand ANR-DFG PhraseoRom (2016-2020) sur la phraséologie spécifique du roman contemporain français, anglais et allemand (par ex. Gonon & Sorba 2019 ; Gonon, Goossens & Novakova 2020 ; Novakova 2023). <https://phraseorom.univ-grenoble-alpes.fr/fr>

² Lettre à Léopold Arnaud le 27 février 1867, cité par Compère dans son introduction (2018, p. 17). À la suite de Guedj (1968), Spandonis (2004) voit surtout dans ce roman « un laboratoire pour l'apprenti naturaliste, un moyen d'élaborer sa méthode : la documentation » (p. 98).

³ « Sur son versant historique, la stylistique, nourrie par l'analyse du discours, raconte une autre histoire : celle de l'avènement, de l'évolution et du déclin de formes et de pratiques langagières partagées par les textes littéraires. Dans cette perspective, les traits de style sont perçus comme des façons de dire et de faire qui émergent, se diffusent voire s'imposent jusqu'à devenir caractéristiques de la littérature d'une époque, puis se banalisent et s'usent » (R. Mahrer, communication personnelle, 2023).

⁴ On trouve cette définition standard dans le TLFi : « l'ensemble des moyens d'expression (vocabulaire, images, tours de phrase, rythme) qui traduisent de façon originale les pensées, les sentiments, toute la personnalité d'un auteur ».

⁵ Philippe (2021) évoque Viktor Vynogradov qui acceptait « presque à contrecœur » de « confronter les pratiques d'un auteur avec celles de ses confrères ou, selon une méthode rétrospective et projective », de rapprocher telle pratique de telles autres dans la succession des temps, afin d'établir des ressemblances susceptibles d'interprétation. » (p. 6).

⁶ Philippe (2021) définit l'écriture polychronique comme « la coprésence de formes stylistiques plus ou moins usées et de formes stylistiques plus ou moins innovantes » (p. 215).

⁷ Notre méthodologie privilégie la notion de cooccurrence syntaxique plus riche que la cooccurrence de surface. Comme l'ont montré Bartsch & Evert (2014), les cooccurrents syntaxiques (par ex., entre un nom et le verbe dont il est l'objet ou entre un nom et l'adjectif qui le modifie) permettent d'identifier plus finement les collocations intéressantes que les cooccurrences de surface, considérés à travers un empan fixe centré autour des mots.

⁸ Sur le concept de *lexicogramme* (Heiden & Tournier 2001) : « il s'agit d'établir, pour un pivot donné, la liste de ses cooccurrents les plus fréquents, à gauche et à droite, en faisant l'extraction des fréquences de cooccurrence et en calculant des mesures d'association statistiques (telles que le rapport de vraisemblance ou le t-score). » (Kraif & Diwersy 2012, p. 400).

⁹ Le seuil de représentativité statistique est fixé à un LLR > 10,83.

¹⁰ Ce corpus a été constitué lors d'un précédent projet de recherche : PhraseoRom_19^e (ANR-15-IDEX-02) et les principes de sa constitution sont expliqués dans Sorba (2022, p. 126-129).

¹¹ Dans son introduction au roman, Compère rappelle « l'importance du facteur topographique dans la construction de l'univers romanesque de Zola » à la suite de Hamon (1983, cité par Compère 2018, p. 31). Par ailleurs Brunet (2009, p. 172) a relevé par des méthodes statistiques dans les Rougon-Macquart que « les mots familiers à Zola qui apparaissent dans la partie haute sont ceux de ses personnages (*abbé, dame, docteur, vendeur, mari*), du cadre qui les entoure (*chaussée, trottoir, préfecture, coron, boutique, comptoir, rue, quartier, fenêtre, magasin, cuisine, chambre*) [...] ».

¹² À la suite de Brunet (2009, p. 162) qui relevait déjà par des méthodes statistiques que l'adjectif trouvait « son apogée chez Zola », Philippe (2021, p. 227) indique : « [...] Zola recourt d'ailleurs plus volontiers à l'adjectif que les autres

romanciers de sa génération, alors que la méfiance envers cette catégorie de mots ne cessera de monter jusqu'à atteindre un sommet dans les années 1920 ».

¹³ La collocation *éclater en sanglot* a été écartée des analyses à cause de son caractère plus figé et donc de l'absence de variabilité en son sein sur les plans syntagmatique et paradigmatique.