



HAL
open science

L'apport de la méthode des motifs aux analyses phraséologiques en discours

Sascha Diwersy, Dominique Legallois

► **To cite this version:**

Sascha Diwersy, Dominique Legallois. L'apport de la méthode des motifs aux analyses phraséologiques en discours. *Approches théoriques et empiriques en phraséologie*, 2014, Nancy, France. ⟨hal-03546271⟩

HAL Id: hal-03546271

<https://hal.science/hal-03546271v1>

Submitted on 27 Jan 2022

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

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



HAL Authorization

Sascha Diwersy / Dominique Legallois

L'apport de la méthode des motifs aux analyses phraséologiques en discours

0 Introduction

L'analyse présentée dans cet article se fonde sur une méthode encore peu utilisée en linguistique de corpus : la méthode dite des « motifs ». Un motif est un patron lexico-grammatical extrait (automatiquement ou non) à partir de textes étiquetés. On peut le voir comme une généralisation ou une extension de ce que la tradition lexicométrique, à la suite de Salem, a nommé « segment répété ». S'il est identifié par des tests statistiques (par exemple, en comparant deux corpus différents), le motif est alors une unité linguistique caractéristique d'un corpus (par exemple d'un genre, d'un auteur, etc.). Pour certains de ces motifs caractéristiques – ceux qui vont nous retenir dans ce travail –, il est possible de mettre en évidence une dimension phraséologique, à condition de considérer la phraséologie non pas seulement comme un phénomène essentiellement linguistique, mais aussi comme un phénomène discursif et directement lié aux genres textuels.

L'article est construit selon le plan suivant : dans un premier temps, nous précisons la notion de phraséologie sur laquelle notre travail est fondé. Nous adoptons en fait la conception ancienne de la notion de *phrase* que des études récentes ont remotivée. La deuxième partie expose en détail la notion de *motif*, la méthode qui permet l'extraction de ces unités, ainsi que le corpus étudié (le genre « fiction » vs le genre « récit de voyage » au 19^e siècle). La troisième partie présente les résultats de l'extraction automatique des segments répétés et des motifs propres aux deux genres. On y montre, en présentant de nombreux exemples, que les motifs constituent des unités plus pertinentes que les segments répétés pour observer la dimension véritablement phraséologique des genres.

1 Une phraséologie des genres

On considérera que la phraséologie est un phénomène linguistique,¹ mais aussi textuel ou discursif. Phénomène linguistique, la phraséologie est liée aux normes de la langue et se caractérise par deux types de contraintes (Mel'čuk / Milicevic 2014) : premièrement, la sélection non libre sur l'axe paradigmatic des éléments composant l'unité. Il s'ensuit qu'un élément ne peut être remplacé par un élément synonyme. Par exemple, *supprimer un malentendu* pour *dissiper un malentendu*. Deuxièmement, la non-liberté sur l'axe syntagmatique, conséquence de ce que l'on appelle traditionnellement la non-compositionnalité sémantique. Par exemple, *prendre une veste*, qui constitue une locution verbale opaque. Le second type de contrainte suppose le premier : *prendre une veste* n'a pas pour synonyme *prendre un manteau* ou même *attraper / saisir une veste*. En revanche, le premier type est indépendant du second puisque *dissiper un malentendu* est, selon nous, transparent.

Cependant, une autre approche de la phraséologie est possible, tout à fait complémentaire de la première : elle permettrait de saisir des unités pour leurs particularités discursives, contraintes par les normes textuelles et génériques. L'unité phraséologie ne serait plus alors dépendante de contraintes observées dans le système linguistique, mais de contraintes redevables à des styles d'auteurs, ou à des périodes, ou encore – c'est le point qui nous intéresse ici – à des genres. Cette conception n'a rien d'absolument originale : elle ne fait que reconsidérer l'ancienne acception du nom *phrase* sur laquelle s'est morphologiquement créé le terme de *phraséologie*, mais que la linguistique contemporaine a quelque peu perdu de vue. Sans proposer ici une analyse détaillée de la notion de *phrase* (cf. Seguin 1993), rappelons plutôt la définition de Furetière (1690) dans laquelle sont consignés les principaux éléments qui nous intéressent :

« PHRASE : manière d'expression, tour ou construction d'un petit nombre de paroles. Il y a des phrases oratoires, d'autres poétiques. Cette phrase est élégante. Voilà une phrase, une façon de parler italienne, espagnole. C'est là une phrase de Ciceron. On a fait plusieurs Livres & Recueils de phrases, comme Bariot et autres. »

1 Les premiers travaux de Bally (1909) sur la phraséologie linguistique ont suscité une relative indifférence, jusqu'aux années 1970 qui ont vu naître un enthousiasme de plus en plus marqué pour la question. Sur ces questions et sur la place de la phraséologie, cf. Legallois / Tutin (2013).

La *phrase*, c'est-à-dire l'unité phraséologie, est définie par Furetière selon un critère assez peu formel, qui la rend soupçonnable aux yeux de la linguistique moderne : la formule, la manière de dire, l'expression, le tour que les locuteurs reconnaissent comme étant typique :

- d'un genre (l'oratoire, le poétique, etc.) ;
- d'un idiome (les gallicismes, par exemple) ;
- d'un style (d'auteur), ou d'un idiolecte.

Pour illustrer par un exemple ce qui relève d'une phraséologie stylistique, nous proposons ces énoncés tirés des Rougon-Macquart :

- (1) Maintenant, Mme Marty avait la face animée et nerveuse d'une enfant qui a bu du vin pur (Au Bonheur des Dames).
- (2) Elle arrondissait le dos, avait la mine confite et gourmande d'une dévote qui va à la messe (L'Assommoir).
- (3) Et, malgré elle, dès que le vieux arrivait avec son odeur de cimetière, elle retombait à ses réflexions, et prenait l'air allumé et craintif d'une épouse qui rêve de donner des coups de canif dans le contrat (L'Assommoir).
- (4) Il se pelotonnait, se cachait sous le drap, avec le souffle fort et continu d'un animal qui souffre (L'Assommoir).

On voit dans ces énoncés ce qui fait « style » : l'expression de la manière par la référence à une représentation générique permettant la compréhension d'un trait physique. Ce stylème est spécifique² de Zola, si on compare ses écrits à ceux de ses contemporains. On considérera qu'il est exprimé par une *phrase* lexicalement et grammaticalement marquée, sans pourtant qu'il y ait contrainte linguistique.

La phraséologie stylistique partage bien des aspects avec la phraséologie générique ; la différence principale étant que les unités phraséologiques génériques ne doivent pas être définies, autant que faire se peut, à partir de l'idiolecte des locuteurs, mais par rapport à ce qui, dans un ensemble de textes appartenant au même genre, est commun à ces textes (ou en tous cas, à une partie des textes).

L'aspect a priori non formel de la phraséologie générique – les contraintes ne se laissent pas définir aussi clairement que celles de la phraséologie

2 Insistons sur le fait que « spécifique » ne veut pas dire « exclusif », mais sur-employé par rapport à un corpus de référence.

linguistique³ – doit être compensé par une observation minutieuse en corpus. Pour ce faire, un dispositif outillé est nécessaire, qui permet de mettre en évidence les types de patrons tels que celui observé chez Zola, soit ce que nous nommons *motifs* ; la section suivante est consacrée à la présentation de la méthode.

2 Méthode pour l'identification de la phraséologie des genres

2.1 La notion de motif

Les motifs ne sont pas de simples configurations syntaxiques ; ce sont plutôt des patrons lexico-grammaticaux (Cellier et al. 2012a et 2012b) qui se définissent par :

- leur récurrence ou leur spécificité ;
- une régularité d'ordre lexical (un même paradigme lexical est employé dans le patron) ou / et fonctionnel : un motif possède une fonction sémantique, pragmatique, rhétorique, voire une fonction d'organisation textuelle.

Longrée / Mellet proposent une définition assez proche :

« la notion de motif est conçue comme un moyen de conceptualiser la multidimensionnalité (ou le caractère multiniveaux) de certaines formes récurrentes qui sollicitent à la fois le lexique, les catégories grammaticales et la syntaxe, éventuellement la prosodie, la métrique » (Longrée / Mellet, 2013 : 42).

Les auteurs ajoutent une propriété supplémentaire : la possible non-séquentialité des éléments qui composent le motif.⁴

2.2 Le corpus

Le corpus choisi est directement inspiré de l'étude de Magri (2009) portant sur le récit de voyage au 19^e siècle. Magri a ainsi étudié les spécificités lexicales, morfo-syntaxiques et énonciatives du genre, en le comparant à la fiction. Un des intérêts de cette étude est que les textes sélectionnés pour

3 Nous ne prétendons pas, évidemment, que la phraséologie linguistique se laisse appréhender sans difficultés.

4 Ce paramètre n'est pas pris en compte dans notre étude.

illustrer les deux genres (12 textes par genres), ont été écrits par les mêmes auteurs. Ainsi, le corpus « récits de voyage » est composé de :

Chateaubriand : *Itinéraire de Paris à Jérusalem*
 Du Camp : *Le Nil, Égypte et Nubie*
 Flaubert : *Par les champs et par les grèves*
 Fromentin : *Un été dans la Sahara*
 Gautier : *Voyage en Espagne*
 Hugo : *Le Rhin*
 Lamartine : *Souvenirs, impressions, pensées et paysages pendant un voyage en Orient (1832-1833) ou Notes d'un voyageur*
 Loti : *Japoneries d'automne*
 Maupassant : *La vie errante*
 Nerval : *Voyage en Orient*
 Sand : *Un hiver à Majorque*
 Stendhal : *Mémoires d'un touriste*

Le corpus « fiction » est composé de :

Chateaubriand : *Vie de Rancé*
 Du Camp : *Mémoires d'un suicidé*
 Flaubert : *Madame Bovary*
 Fromentin : *Dominique*
 Gautier : *Le capitaine Fracasse*
 Hugo : *Notre Dame de Paris*
 Lamartine : *Le Tailleur de pierre de Saint-Point*
 Loti : *Madame Chrysanthème*
 Maupassant : *Une vie*
 Nerval : *Aurélia*
 Sand : *La petite Fadette*
 Stendhal : *Le Rouge et le Noir*

2.3 La méthode

Techniquement, plusieurs méthodes sont possibles. Le site SDMC⁵ conçu au laboratoire Greyc de l'Université de Caen propose à tout utilisateur

5 <https://sdmc.greyc.fr/>

d'identifier des suites de combinaisons d'items morpheo-syntaxiques et lexicaux dans des corpus préalablement annotés. Pour ce travail, nous avons opéré ainsi : nous avons transformé chacun des textes annotés (lemme et catégorie morpheo-syntaxique)⁶ en autant de textes réduits en une suite d'éléments correspondant soit à des catégories générales (nom, adjectif, verbes, etc.), soit à des lemmes pour les mots les plus courants (*avoir, être, faire*, etc.), soit à des formes invariables pour les mots grammaticaux. Les temps verbaux sont également considérés. Les processus d'annotation et de transformation d'une phrase peuvent être illustrés par cet exemple (Maupassant, *La vie errante*) :

- (5) J'ai quitté Paris et même la France, parce que la Tour Eiffel finissait par m'ennuyer trop.

J' je PPER1S ai avoir VINDP1S quitté quitter VPARPMS
 Paris Paris NPHSIG et et COO même même ADV la
 le DETDFS France France NP , , PCTFAIB parce que
 parce que SUB la le DETDFS tour tour NCFS Eiffel
 eiffel NP finissait finir VINDI3S par par PREP
 m' me PPER1S ennuyer ennuyer VINF trop trop ADV . .
 PCTFORTE

Le texte annoté est transformé ainsi :

Je avoir PASS NP et même le NP, parce que le N NP finir par me INF trop.

Sont alors identifiées, à la manière des segments répétés, les suites récurrentes. Seules ont été retenues les suites qui apparaissent deux fois chez au moins cinq auteurs. Cette contrainte permet de préserver l'analyse des effets d'auteur, c'est-à-dire des formes qui seraient abondantes chez un auteur et peu présentes chez les autres. Nous opérons ensuite un calcul des spécificités (en termes de la loi hypergéométrique selon la méthode proposée par Lafon 1980) pour faire émerger statistiquement les séquences caractéristiques. Un retour au

6 L'annotation a été effectuée au moyen du logiciel Cordial Analyseur (version 14.0).

texte permet alors d'identifier les suites interprétables comme unités, c'est-à-dire les motifs.

3. Résultats et commentaires

Deux analyses ont en fait été menées : l'extraction des segments répétés spécifiques et l'extraction des patrons lexico-grammaticaux spécifiques. Nous donnons d'abord les résultats concernant les segments répétés (d'une longueur de 4 à 6 unités).

3.1 Les segments répétés

Nous rappelons qu'un segment répété (ou n-gramme) est une suite récurrente de x mots (ici représentés par des lemmes).

<i>Segment Répété</i>	<i>Score</i>	<i>Fréquence dans « fiction »</i>	<i>Fréquence totale</i>
ce ne être pas	Inf	268	499
il ne avoir pas	Inf	152	253
ne être ce pas	Inf	131	195
que il ne avoir	Inf	135	197
que est ce que	Inf	123	158
je ne savoir pas	Inf	82	111
je ne vouloir pas	Inf	76	98
elle se mettre à	Inf	62	62
se écrier t il	Inf	61	61
il lui sembler que	Inf	49	49
est ce que je	Inf	43	43
de le hôtel de	Inf	41	41
vous ne avoir pas	Inf	39	39
est ce que vous	Inf	35	35
tu ne avoir pas	Inf	35	35
lui prendre le main	15.95	33	33
si il avoir être	15.95	33	33

être sur le point de	15.95	33	33
ce être moi qui	15.35	32	32
ne pouvoir se empêcher de	15.35	32	32
est ce que tu	14.88	31	31
comme si il avoir	14.78	80	115
se écrier t elle	14.39	30	30
je ne être pas	14.26	109	176
ajouter t il en	13.43	28	28
on avoir dire que	12.94	58	78
il falloir que je	12.31	49	63
elle ne être pas	12.19	57	78
dire il en se	11.99	25	25

Tableau 1 : trente premiers segments spécifiques à « fiction »

<i>Segment Répété</i>	<i>Score</i>	<i>Fréquence dans « voyage »</i>	<i>Fréquence totale</i>
de le montagne de	Inf	106	106
le bord de le mer	Inf	100	100
le tombeau de le	13.12	75	75
le ruine de le	12.77	73	73
de le porte de	11.90	68	68
le palais de le	11.20	64	64
le ville de le	11.20	64	64
de le temple de	10.85	62	62
à le bord de le mer	10.85	62	62
le vallée de le	10.67	61	61
à le haut de	10.50	60	60
nous entrer dans le	10.32	59	59
de le ville et	9.97	57	57
de le ville de	9.62	55	55

le pays de le	9.62	55	55
de le palais de	9.27	53	53
le époque de le	9.27	53	53
de le mer de	9.10	52	52
à quelque pas de	9.10	52	52
saint jean d acre	8.75	50	50
de le mer le	8.57	49	49
le église de le	8.57	49	49
de le montagne de le	8.40	48	48
de le plus grand	8.40	48	48
le un sur le	8.40	48	48
ça et là un	8.40	48	48
ça et là de	8.22	47	47
le montagne de le	8.13	105	117
le pied de le montagne	8.05	46	46

Tableau 2 : trente premiers segments spécifiques à « récits de voyage »

La comparaison des deux tableaux montre que la fiction est plus affectée par les segments répétés phraséologiques (discursifs) exprimant :

- la modalité (au sens général) : *je ne savoir pas, je ne pouvoir pas, il falloir que je, ne pouvoir se empêcher de*, etc.
- la modalité « représentative » : *comme si il avoir, on avoir dire que, il lui sembler que*, etc.
- l'aspect : *elle se mettre à, être sur le point de*, etc.

Les récits de voyage – en ce qui concerne les trente premières formes – se montrent peu affectés : les segments répétés sont principalement des syntagmes dans lesquels des noms à valeur géographique sont employés : *montagne, ville, pays, mer, porte, vallée*. Rien n'apparaît, du tableau 2, qui relève véritablement de la formule, de l'expression. On pourrait conclure à une sensibilité moindre de ce corpus à la phraséologie, mais on peut, au préalable, se poser la question de la méthode employée : les segments répétés sont-ils les formes les plus à même de nous nous faire apprécier la phraséologie

spécifique de ce genre ? La prochaine section montre que les motifs donnent des résultats plus intéressants en ce qui concerne cette phraséologie.

3.2 Les motifs

La méthode exposée en 2.3 génère un nombre de formes particulièrement élevé, dont certaines sont ininterprétables. Il s'agit alors pour l'analyste d'observer quelles sont celles qui constituent véritablement des motifs, c'est-à-dire des formes manifestant des régularités d'ordre lexico-grammatical et fonctionnel. Seul un examen du comportement des patrons dans les textes permet d'identifier les motifs véritablement phraséologiques. C'est là la limite de l'automatisation.

Dans ce qui suit, nous donnons pour chaque genre une dizaine d'exemples de motifs phraséologiques. Nous ne donnons pas, pour alléger l'article, de mesure statistique, mais pour chaque motif, nous donnons entre parenthèses sa fréquence dans le sous-corpus par rapport à la fréquence totale.

Pour le genre Fiction :

Il est à remarquer que le lexique lié à l'expression des parties du corps est en emploi surnuméraire dans le genre Fiction, comparé au genre Voyage. Mais ce lexique fait l'objet de régularités constructionnelles ; il est employé dans des séquences facilement identifiables. En voici deux :⁷

- il VIMP à le N DET N (114/169)

le motif exprime le fait qu'un personnage porte un objet caractéristique :

- (6) Mais il avait aux pieds ces semelles de plomb dont on prétend qu'usent certains coureurs en leurs exercices pour être plus légers ensuite (Gautier, *le Capitaine Fracasse*).
- (7) Il avait au col et aux poignets des touffes de doreloterie (Hugo, *Notre Dame de Paris*)
- (8) Au moment où je franchissais la porte, un homme vêtu de blanc, dont je distinguais mal la figure, me menaçait d'une arme qu'il tenait à la main (Nerval, *Aurélia*)

⁷ La valeur des codes est indiquée en annexe.

- DETPOSS N ADJ se VIMP (28/28)

Une partie du corps du personnage s'anime, ou devient saillante :

- (9) Ses lèvres pâles s'agitaient sans parler, son visage était marbré de taches rouges (Du Camp, *Mémoires d'un suicidé*).
- (10) Au seuil du chenil, un chien unique, flottant dans sa peau trop large où ses muscles détendus se dessinaient en lignes flasques, sommeillait (Gautier, *le Capitaine Fracasse*).
- (11) Quelle secrète pensée faisait sourire sa bouche avec tant d'amertume au même moment où ses sourcils froncés se rapprochaient comme deux taureaux qui vont lutter ? (Hugo, *Notre Dame de Paris*).

Par cet autre motif, il s'agit d'exprimer la production d'un geste, ou d'un son :

- il V-PS un N ADJ (47/57)

- (12) Elle fit une inclination rapide et referma la fenêtre (Flaubert, *Madame Bovary*).
- (13) Il poussa une sorte de soupir excédé (Fromentin, *Dominique*).
- (14) Elle eut un mouvement nerveux bien marqué quand elle reconnut l'écriture de la maréchale (Stendhal, *Le Rouge et le Noir*).

La phraséologie en discours peut également se définir par des aspects rythmiques. La fiction, par rapport aux récits de voyage, favorise l'accumulation d'actions, exprimée par des verbes ou groupes verbaux juxtaposés :

- , VIMP, VIMP, (19/19)

- (15) Les truands hurlaient, haletaient, juraient, montaient (Hugo, *Notre Dame de Paris*).
- (16) Et toute la surface de la baie, piquée de pluie, tourmentée par de les tourbillons qui arrivaient de partout, clapotait, gémissait, se démenait dans une agitation extrême (Loti, *Madame Chrysanthème*).

Le motif « rythmique » peut reprendre le sujet :

- il VIMP, il VIMP (19/19)

- (17) Il était fou de rage, il écumait, il trépignait (Maxime Du Camp, *Mémoires d'un suicidé*).

- (18) Elle allait, elle venait, elle n'avait jamais été si attentive pour couper le pain de mon pauvre frère et pour amuser Annette (Lamartine, *Le tailleur de pierre de Saint-Point*).

ou bien l'objet :

- le VIMP, le VIMP (25/25)

- (19) Elle lui apportait des linges pour ses cataplasmes, et le consolait, l'encourageait (Flaubert, *Madame Bovary*).

Parmi les constructions typiques de la Fiction, on identifie les phrases intensives consécutives :⁸

- un N si ADJ que (39/52)

- (20) Et puis je suis tiraillé par un regret si singulier qu'il en est ridicule (Lamartine, *Le tailleur de pierre de Saint-Point*).

- (21) Cette fois Rosalie fit un effort si brusque qu'elle arracha ses mains de celles de sa maîtresse, et se sauva comme une folle (Maupassant, *Une vie*).

On peut se demander toutefois si les deux exemples illustrent le même patron puisque le pronom de la corrélatrice en (19) reprend le sentiment qui affecte le personnage, tandis qu'en (20), c'est le personnage lui-même qui est repris.

On considérera encore des comparatives de supériorité avec deux adjectifs coordonnés :

- plus ADJ et plus ADJ que (21/21)

- (22) Elle me parut plus grande et plus belle qu'une créature humaine (Du Camp, *Mémoire d'un suicidé*).

- (23) Le sucre en poudre même lui parut plus blanc et plus fin qu'ailleurs (Flaubert, *Madame Bovary*).

La proposition complexe – ou un certain type et mode de construction de cette proposition – peut être caractéristique du genre grâce à une spécificité expressive : ici l'effet de l'environnement sur le corps du personnage :

- qui lui VIMP le N (36/36)

8 Que Adam (2011) analyse en détail dans une perspective de linguistique des genres.

- (24) Les ardoises laissaient tomber d'aplomb une chaleur lourde, qui lui serrait les tempes et l'étouffait (Flaubert, *Madame Bovary*).
- (25) Elle courut à la fenêtre, et elle se mit à pleurer affreusement avec des cris involontaires qui lui déchiraient la gorge (Maupassant, *Une vie*).

Considérons à présent quelques motifs spécifiques du récit de voyage.

Pour « récits de voyage » :

Dans les descriptions, on est surpris de la fréquence d'une construction attributive dont l'attribut exprime une caractéristique intrinsèque au référent :

- le N être de un N ADJ (27/32)

- (26) Le ciel à l'orient était d'un rouge vif, qui pâlisait à mesure que la lumière croissait (Chateaubriand, *Itinéraire de Paris à Jérusalem*).
- (27) La place du Marché, qui entoure deux côtés de la cathédrale, est d'un ensemble copieux, fleuri et divertissant (Hugo, *Le Rhin*).
- (28) Le haïk est d'une étoffe de coton cassante et légère (Fromentin, *Un été dans la Sahara*).

Toujours dans les descriptions de paysages ou de bâtiments, l'auteur s'appuie sur des expansions par une relative typique en *dont*, dans laquelle figure un verbe à forme pronominale et à valeur existentielle (*se dresser, se perdre*, etc.) :

- (et) dont le N se VP (53/53)

- (29) Quand on est arrivé à la hauteur de la Pigue, on est retourné vers la France que l'on aperçoit dans les nuages et dont les plaines se dressent au loin comme des immenses tableaux suspendus (Flaubert, *Par les champs et par les grèves*).
- (30) Après un kilomètre ou deux de ce piétinement épuisant, j'entrevois dans la brume, au loin, devant moi, une tour mince et pointue (...) et dont le sommet se perd dans la nuée (Maupassant, *La vie errante*).

Parmi les motifs spécifiques, on identifie également cette expansion par une relative explicative :

- le N ADJ qui VP le N(62/62)

(31) Dans la haie rose, qui couvre la muraille, les chrysanthèmes sont grands comme des arbres, et leurs fleurs sont larges comme des soleils (Loti, *Japoneries d'automne*).

(32) Là, c'est-à-dire à la fontaine Tritzella, nous nous trouvons derrière Misitra et presque au pied du château ruiné qui commande la ville (Chateaubriand, *Itinéraire de Paris à Jérusalem*).

On considère également phraséologique et typique du récit de voyage, la forme suivante, dédiée au compte rendu de perception :

- Loc, on VP un N (72/72)

(33) Tout au fond, du côté de l'Ethiopie, on distingue un rideau de verdure à moitié disparue, derrière une brume de sable soulevée par le vent (Du Camp, *Le Nil, Egypte et Nubie*).

(34) En suivant la côte européenne du Bosphore, on voit une longue file de maisons de campagne habitées généralement par des employés du sultan (Nerval, *Voyage en orient*).

À des fins d'accréditation du témoignage (Legallois et al. 2011), il s'agit parfois de frapper le lecteur par une sorte d'exubérance du paysage – les descriptions empruntent alors des formes exprimant une saturation :

- DET N être PASS de N // PASS de N, de N (319/376)

(35) Toute la colline est remplie de sépultures antiques et de pagodes cachées sous les arbres (Loti, *Japoneries d'automne*).

(36) La place était encombrée de ballots, de marchandises, de chevaux, de chiens sans maître, et de Turcs accroupis qui fumaient à l'ombre (Lamartine, *Voyage en Orient*).

Signalons encore ces deux motifs, formellement très proches, qui sont exclusifs du récit de voyage :

- rien de plus ADJ que ce N (54/54)

(37) Rien de plus délicieux que ces voyages en caravane quand le pays est beau (Lamartine, *Notes d'un voyageur*).

(38) Rien de plus sauvage que cette manière de se réjouir en se brisant le tympan avec le claquement du bois (Sand, *Un hiver à Majorque*).

- rien ne être plus ADJ que (22/22)

- (39) Rien n'est plus charmant que de voir aller et venir par petits groupes les jeunes femmes et les jeunes filles en mantille (Gautier, *Voyage en Espagne*).
- (40) Rien n'est plus intéressant que l'étude de l'organisation de cet immense domaine où les intérêts des indigènes ont été sauvegardés avec autant de soin que ceux des Européens (Maupassant, *Une vie errante*).
- (41) Rien n'était plus pittoresque que cette immense procession de femmes et de jeunes filles dans les costumes les plus étranges et les plus splendides (Lamartine, *Notes d'un voyageur*).

Enfin, donnons ce dernier motif, qui par hyperbole, participe également à l'emphase et à l'enthousiasme de l'écrivain voyageur :

- de le plus ADJ N de le N (32/45)

- (42) Après Strabon, nous ne savons plus rien de Carthage, sinon qu'elle était devenue une des plus grandes et des plus belles villes du monde (Chateaubriand, *Itinéraire de Paris à Jérusalem*).
- (43) Mais il en reste assez pour signaler au loin le promontoire à l'œil du navigateur, et à l'admiration de l'homme épris des plus belles scènes de la nature (Lamartine, *Voyage en Orient*).

3.3 Discussion

L'étude proposée ici s'inscrit dans une analyse des genres, mais ne s'attache pas, contrairement à la grande majorité des travaux entrepris dans ce domaine, à considérer les « simples » traits morpho-syntaxiques et leurs corrélations. Les motifs, par leur nature syntagmatique, permettent véritablement d'appréhender les genres par des « entrées discursives » que sont – en l'occurrence – les phrases spécifiques. Une analyse factorielle⁹ établie sur les motifs de chacun des textes – fictions et récits de voyage – montre bien que le genre est parfaitement discriminable par ces séquences. Les récits de voyage (notés auteur_V) se regroupent dans la partie gauche de la carte, les fictions (notées auteur_R) dans la partie droite. Seul, le roman de Loti vient se mêler aux récits de voyage (la position du roman est difficilement

9 Pour le principe de l'analyse factorielle, on peut lire l'article de Viprey (2000).

repérable sur la carte), ce qui était assez prévisible, puisque Loti était un écrivain voyageur.

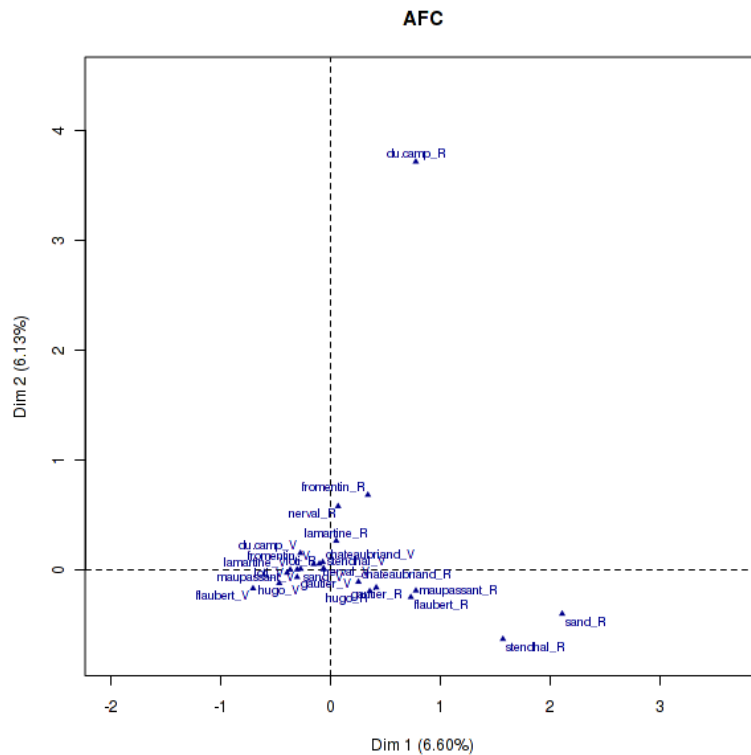


Figure 1 : Analyse factorielle sur les Fictions et Récits de voyage¹⁰

La méthode permet également d'extraire des patrons qui échappent à notre intuition du genre. Par exemple, pour la Fiction, la spécificité du patron V-PS il en PRES nous apparaît immédiatement (*ajouta-t-il en riant, murmura-t-elle en pleurant*, etc.). Mais beaucoup d'autres nous apparaissent moins saillants, comme ADJNUM ou ADJNUM ADJNUM N pour le corpus Voyage :

10 Le graphique a été créé au moyen du package R FactoMineR (Lê, Josse & Husson 2008).

- (44) À côté, sept ou huit grands vautours bruns, à tête chenue, s'étaient réunis contre un gypaète blanc qu'ils accablaient de coups de bec (Du Camp, *Le Nil, Egypte et Nubie*).
- (45) Le cirque de Malaga est d'une grandeur vraiment antique et peut contenir douze ou quinze mille spectateurs dans son vaste entonnoir (Gautier, *Voyage en Espagne*).

Ce motif est pourtant extrêmement fréquent et n'est pas, lui non plus, sans jouer un rôle dans l'accréditation : une évaluation précise et donc exacte est toujours suspecte.

La phraséologie discursive, qu'elle soit générique ou stylistique, exploite autant les formes offertes par la phraséologie linguistique (par exemple, les tours en « rien n'être plus ADJ que », « rien de plus ADJ que », qui peuvent être considérés comme des patrons mémorisés), que les phénomènes relevant du cliché ou du stéréotype, exprimé par des formes régulières (par exemple, « DET N être PASS de N » permettant de décrire des scènes « saturées » – *toute la colline est remplie de sépultures antiques*).

4 Conclusion

L'exemple pris dans cet article pour décrire la méthode des motifs phraséologiques génériques est littéraire ; en cela, même si ce travail se situe en amont d'une analyse interprétative des textes, l'exemple illustre la façon dont les spécialistes de ce discours pourraient exploiter avec profit la méthode. Elle permet, au-delà des particularités morpho-syntaxiques des genres, mais aussi des segments répétés, de mettre en évidence les régularités des patrons syntaxiques, et aussi – c'est une possibilité qu'offre la méthode – leur évolution en diachronie : on pourrait ainsi évaluer les changements qui affectent, par exemple, le récit de voyage depuis le début du 19^e siècle,¹¹ jusqu'au 20^e. Mais la méthode s'applique, évidemment, à d'autres genres et à d'autres discours :¹² les motifs sont des entrées privilégiées pour observer et analyser les routines discursives et les tours. Reste encore à perfectionner la méthode, notamment la technique d'annotation : la découverte automatique des motifs est dépendante de l'annotation préalable. Nous avons exposé celle qui a été exposée ici, mais il est toujours envisageable d'enrichir l'annotation :

11 Berty (2001).

12 Legallois (2012) pour d'autres exemples.

des dimensions sémantiques (par exemple, des types de verbes, des types de noms), des dimensions pragmatiques (par exemple, les effets argumentatifs de tel et tel adverbe ou connecteur), des dimensions interactionnelles, etc. devront être prises en compte pour apprécier avec plus de précision les particularités non plus *phrastiques* mais *phraséologiques* des discours.

Annexe

ADJ adjectif
ADJNUM adjectif numéral
DET déterminant
DETPOSS déterminant possessif
LOC expression à valeur locative
N nom commun
PASS participe passé
PRES participe présent
VIMP verbe conjugué à l'indicatif imparfait
VP verbe conjugué à l'indicatif présent
V-PS verbe conjugué au passé simple

Bibliographie

- Adam, Jean-Michel, 2011. « Les consécutives intensives : un schéma syntaxique commun à plusieurs genres de discours ». In : *Linx* 64-65, 115-131.
- Bally, Charles, 1909. *Traité de stylistique française*. Paris : Klincksieck.
- Berty, Valérie, 2001. *Un essai de typologie narrative des récits de voyage français au 19ème siècle*. Paris : L'Harmattan.
- Cellier, Péguy / Quiniou, Solenn / Charnois, Thierry / Legallois, Dominique, 2012a. « Fouille de données pour la stylistique : l'exemple des motifs émergents ». In : Anne Dister/ Dominique Longrée/ Gérald Purnelle (éds) *Actes des Journées Internationales d'analyse statistique des données textuelles (JADT'12)*. Liège, Belgique, 821-833, publié sur <http://lexicometrica.univ-paris3.fr/jadt/> (consulté le 17 juillet 2015)
- Cellier, Péguy / Quiniou, Solenn / Charnois, Thierry / Legallois, Dominique, 2012b. « What About Sequential Data Mining Techniques to Identify

- Linguistic Patterns for Stylistics ? ». In : *Lecture Notes in Computer Science*, Springer Vol. 7181, 166-177.
- Lafon, Pierre, 1980. « Sur la variabilité de la fréquence des formes dans un corpus ». In : *Mots* 1, 127-165.
- Lê, Sébastien / Josse, Julie / Husson, François, 2008. « FactoMineR: An R Package for Multivariate Analysis ». In : *Journal of Statistical Software* 25 (1), 1-18.
- Legallois, Dominique, 2012. « La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ». In : *Corpus* 11, 31-54.
- Legallois, Dominique / Malgouze, Yannick / Vigier, Luc, 2011. *L'accréditation des discours testimoniaux*, Toulouse : Editions Universitaires du Sud.
- Legallois, Dominique / Tutin, Agnès, 2013. « Présentation : Vers une extension du domaine de la phraséologie ». In : *Langages* 189, 3-25.
- Longrée, Dominique / Mellet, Sylvie, 2013. « Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours ». In : *Langages* 189, 68-80.
- Magri Véronique, 2009. *Le Voyage à pas comptés. Le Récit de voyage : mesures, langue et genre de discours*, Paris : H. Champion.
- Salem, André, 1987. *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
- Mel'čuk Igor / Milicevic, Jasmina, 2014. *Introduction à la linguistique* (en trois volumes). Paris : Hermann.
- Seguin, Jean-Pierre, 1993. *L'Invention de la phrase au XVIIIe siècle*. Louvain / Paris : Peeters / BIG.
- Viprey, Jean-Marie, 2000. « La Statistique de norme endogène dans l'analyse de textes ». In : *Linx* 43, URL : <http://linx.revues.org/1099> (mis en ligne le 11 juillet 2012, consulté le 17 juillet 2015).